Proceedings of the 2016 Federated Conference on Computer Science and Information Systems

September 11–14, 2016. Gdańsk, Poland

Maria Ganzha, Leszek Maciaszek, Marcin Paprzycki (eds.)

PTI

IEEE

Annals of Computer Science and Information Systems, Volume 8

Proceedings of the 2016 Federated Conference on Computer Science and Information Systems

Maria Ganzha, Leszek Maciaszek, Marcin Paprzycki (eds.)

2016, Warszawa,
Polskie Towarzystwo
Informatyczne

2016, New York City,
Institute of Electrical and
Electronics Engineers

Annals of Computer Science and Information Systems, Volume 8

Proceedings of the 2016 Federated Conference on Computer Science and Information Systems

ART: ISBN 978-83-60910-92-7, IEEE Catalog Number CFP1685N-ART
USB: ISBN 978-83-60810-91-0, IEEE Catalog Number CFP1685N-USB
WEB: ISBN 978-83-60810-90-3

ISSN 2300-5963
DOI 10.15439/978-83-60810-90-3

Ul. Solec 38/103
00-394 Warsaw
Poland

10662 Los Vaqueros Circle
Los Alamitos, CA 90720
USA

Contact: secretariat@fedcsis.org
http://annals-csis.org/

Cover:
Jana Waleria Denisiuk,
Elbląg, Poland

Also in this series:

Volume 9: Position Papers of the 2016 Federated Conference on Computer Science and Information Systems, **ISBN WEB: 978-83-60810-93-4, ISBN USB: 978-83-60810-94-1**
Volume 7: Proceedings of the LQMR Workshop, **ISBN WEB: 978-83-60810-78-1, ISBN USB: 978-83-60810-79-8**
Volume 6: Position Papers of the 2015 Federated Conference on Computer Science and Information Systems, **ISBN WEB: 978-83-60810-76-7, ISBN USB: 978-83-60810-77-4**
Volume 5: Proceedings of the 2015 Federated Conference on Computer Science and Information Systems, **ISBN WEB: 978-83-60810-66-8, ISBN USB: 978-83-60810-67-5**
Volume 4: Proceedings of the E2LP Workshop, **ISBN WEB: 978-83-60810-64-4, ISBN USB: 978-83-60810-63-7**
Volume 3: Position Papers of the 2014 Federated Conference on Computer Science and Information Systems, **ISBN WEB: 978-83-60810-60-6, ISBN USB: 978-83-60810-59-0**
Volume 2: Proceedings of the 2014 Federated Conference on Computer Science and Information Systems, **WEB: ISBN 978-83-60810-58-3, USB: ISBN 978-83-60810-57-6, ART: ISBN 978-83-60810-61-3**
Volume 1: Position Papers of the 2013 Federated Conference on Computer Science and Information Systems (FedCSIS), **ISBN WEB: 978-83-60810-55-2, ISBN USB: 978-83-60810-56-9**

DEAR Reader, it is our pleasure to present to you Proceedings of the 2016 Federated Conference on Computer Science and Information Systems (FedCSIS), which took place in Gdańsk, Poland, on September 11–14, 2016.

FedCSIS 2016 was Chaired by prof. Krzysztof Goczyła, while Zenon Filipiak acted as the Chair of the Organizing Committee. This year, FedCSIS was organized by the Polish Information Processing Society (Mazovia Chapter), Systems Research Institute Polish Academy of Sciences, Warsaw Univesity of Technology, Wrocław University of Economics, and Gdańsk University of Technology. It was organized in technical cooperation with: IEEE Region 8, IEEE SMC Technical Committee on Computational Collective Intelligence, IEEE Computer Society Technical Committee on Intelligent Informatics, IEEE Poland Section, Computer Society Chapter Poland, Gdańsk Computer Society Chapter Poland, Polish Chapter of the IEEE Computational Intelligence Society, ACM Special Interest Group on Applied Computing, Łódź ACM Chapter, European Alliance for Innovation (EAI), Committee of Computer Science of the Polish Academy of Sciences, Polish Operational and Systems Research Society, Mazovia Cluster ICT Poland and Eastern Cluster ICT Poland. Furthermore, the 11th International Symposium Advances in Artificial Intelligence and Applications (AAIA'16) was organized in technical cooperation with: International Fuzzy Systems Association, European Society for Fuzzy Logic and Technology, International Rough Set Society and Polish Neural Networks Society.

FedCSIS 2016 consisted of the following events (conferences, symposia, workshops, special sessions). These events were grouped into FedCSIS conference areas**,** of various degree of integration. Specifically, those listed in italics and without indication of the year 2016 signify "abstract areas" with no direct paper submissions (i.e. paper submissions only within enclosed events).

- **AAIA'16 – 11th International Symposium Advances in Artificial Intelligence and Applications**
  - AIMaVIG'16 – 2nd International Workshop on Artificial Intelligence in Machine Vision and Graphics
  - AIMA'16 – 6th International Workshop on Artificial Intelligence in Medical Applications
  - AIRIM'16 – 1st International Workshop on AI aspects of Reasoning, Information, and Memory
  - ASIR'16 – 6th International Workshop on Advances in Semantic Information Retrieval
  - DSTUI'16 – 6th International Workshop on Dealing with Spatial and Temporal Uncertainty and Imprecision
  - LTA'16 – 1st International Workshop on Language Technologies and Applications
  - WCO'16 – 9th International Workshop on Computational Optimization
- ***CSS - Computer Science & Systems***
  - AIPC'16 – 1st International Workshop on Advances in Image Processing and Colorization
  - CANA'16 – 9th Computer Aspects of Numerical Algorithms
  - CPORA'16 – 1st Workshop on Constraint Programming and Operation Research Applications
  - IWCPS'16 – 3rd International Workshop on Cyber-Physical Systems
  - MMAP'16 – 9th International Symposium on Multimedia Applications and Processing
  - WSC'16 – 8th Workshop on Scalable Computing
- **ECRM – Education, Curricula & Research Methods**
  - IEES'16 – 1st International E-education Symposium - Education of the Future
  - DS-RAIT'16 – 3rd Doctoral Symposium on Recent Advances in Information Technology
- **iNetSApp'16 – 4th International Conference on Innovative Network Systems and Applications**
  - EAIS'16 – 3rd Workshop on Emerging Aspects in Information Security
  - SoFAST-WS'16 – 5th International Symposium on Frontiers in Network Applications, Network Systems and Web Services
  - WSN'16 – 5th International Conference on Wireless Sensor Networks
- ***IT4MBS – Information Technology for Management, Business & Society***
  - ABICT'16 – 7th International Workshop on Advances in Business ICT
  - AITM'16 – 14th Conference on Advanced Information Technologies for Management
  - ISM'16 – 11th Conference on Information Systems Management
  - KAM'16 – 22nd Conference on Knowledge Acquisition and Management
  - UHH'16 – 2nd International Workshop on Ubiquitous Home Healthcare
- ***JAWS – Joint Agent-oriented Workshops in Synergy***
  - MAS&S'16 – 10th International Workshop on Multi-Agent Systems and Simulations
  - SEN-MAS'16 – 4th International Workshop on Smart Energy Networks & Multi-Agent Systems
- **SSD&A – Software Systems Development & Applications**
  - BTMSPA'16 – 1st Symposium on Balancing Traditional and Modern Software Process Approaches
  - MDASD'16 – 4th Workshop on Model Driven Approaches in System Development
  - MIDI'16 – 4th Conference on Multimedia, Interaction, Design and Innovation
  - SEW-36 – The 36th IEEE Software Engineering Workshop

This year (2016) is the year of the 90th anniversary of the birth and the 10th anniversary of the death of Professor Zdzisław Pawlak. Therefore, a special plenary panel devoted to the “Legacy of Professor Zdzisław Pawlak” has been or-

ganized. During the panel, friends, students and collaborators of Prof. Pawlak shared their memories and reflections.

Furthermore, an AAIA'16 Data Mining Competition, focused on "Predicting Dangerous Seismic Events in Active Coal Mines" has been organized. Its results constitute a separate section in these proceedings. Awards for the winners of the contest were sponsored by: Research and Development Center EMAG and the Mazovia Chapter of the Polish Information Processing Society.

Each paper, found in this volume, was refereed by at least two referees and the acceptance rate of full (regular) papers was 25.39% (130 papers out of 512 submissions). Here, let us note that this year the number of submissions has been the largest in the history of the FedCSIS conference series.

Each event constituting FedCSIS had its own Organizing and Program Committee. We would like to express our warmest gratitude to the members of all of them for their hard work attracting and later refereeing 512 submissions.

FedCSIS 2016 was organized under the auspices of dr Jarosław Gowin, Minister of Science and Higher Education, Anna Streżyńska, Minister of Digital Affairs, Paweł Adamowicz, Mayor of the City of Gdańsk, Wojciech Szczurek Mayor of the City of Gdynia, and prof. Henryk Krawczyk, Rector of the Gdańsk University of Technology.

Finally, FedCSIS 2016 was sponsored by the Ministry of Science and Higher Eduction and Intel.

Maria Ganzha, *Co-Chair of the FedCSIS Conference Series, Systems Research Institute Polish Academy of Sciences, Warsaw, Poland, and Warsaw University of Technology, Poland*
Leszek Maciaszek, *Co-Chair of the FedCSIS Conference Series, Wrocław University of Economics, Wrocław, Poland and Macquarie University, Sydney, Australia*
Marcin Paprzycki, *Co-Chair of the FedCSIS Conference Series, Systems Research Institute Polish Academy of Sciences, Warsaw and Management Academy, Warsaw, Poland*

Proceedings of the 2016 Federated Conference on Computer Science and Information Systems (FedCSIS)

September 11–14, 2016. Gdańsk, Poland

TABLE OF CONTENTS

Plenary Panel on the Legacy of Professor Zdzisław Pawlak

Data Mining Challenge: Predicting Dangerous Seismic Events in Active Coal Mines

2ND International Workshop on Artificial Intelligence in Machine Vision and Graphics

6TH International Workshop on Artificial Intelligence in Medical Applications

1ST International Workshop on AI aspects of Reasoning, Information, and Memory

6TH International Workshop on Advances in Semantic Information Retrieval

6TH International Workshop on Dealing with Spatial and Temporal Uncertainty and Imprecision

1ST International Workshop on Language Technologies and Applications

9TH International Workshop on Computational Optimization

8TH Workshop on Scalable Computing

Education, Curricula & Research Methods

1ST International E-education Symposium–Education of the Future

3RD Doctoral Symposium on Recent Advances in Information Technology

4TH International Conference on Innovative Network Systems and Applications

3RD Workshop on Emerging Aspects in Information Security

5TH International Symposium on Frontiers in Network Applications, Network Systems and Web Services

5TH International Conference on Wireless Sensor Networks

Information Technology for Management, Business & Society

7TH International Workshop on Advances in Business ICT

14TH Conference on Advanced Information Technologies for Management

11TH Conference on Information Systems Management

22ND Conference on Knowledge Acquisition and Management

4TH Conference on Multimedia, Interaction, Design and Innovation

The 36TH IEEE Software Engineering Workshop

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1
DOI: 10.15439/2016F598
ACSIS, Vol. 8. ISSN 2300-5963

Big Data Meets Big Water: Analytics of the AIS Ship Tracking Data

Stan Matwin
Institute for Big Data Analytics
Dalhousie University
Halifax, NS, Canada B3H 1W5 Canada
Email: stan@cs.dal.ca

IN THIS presentation we will argue that Big Data technologies can contribute in an important way to an unprecedented breakthrough in the understanding of oceans as a factor in climate change, in transportation, and in supplying humanity with its important food component.

Oceans cover almost 70% of the surface of the earth, and supply at least 15% of animal protein intake for 4.5 billion people. At the same time, ocean are an area of human interest that undergoes currently a massive infusion of information technology. As a result, ocean data and its challenges will become a fertile ground for data science.

After briefly introducing Big Data, we will argue that many of the emerging data sources focused on oceans are Big Data. We will use the global Automatic Identification System as an example. We will introduce the AIS system and characterize it quantitatively. We will then illustrate some of the Big Data projects under way in the Institute for Big Data Analytics, Dalhousie University. In particular, we will focus on the analysis of fishing ship trajectories available through AIS data, and will show how this analysis can lead in the future to unprecedented quality of estimates of the fish intake by global fisheries.

We will discuss AIS data management, data preprocessing techniques, data segmentation, data representation, and data modeling (point-wise and geometrically). We will demo a specific implementation of our data management solution. We will present our early experiences with some of the basic classification tasks (ship kind classification, fishing gear classification, fishing-non fishing classification) using Markovian approaches, standard data exploration approaches, and classifier induction approaches. We will also show how alternative methods from Natural Language Processing can assist in the same task. We will also discuss early results and challenges with the use of Deep Learning methods (e.g. Long Short Term Memory) on the AIS data.

Finally we will discuss the ongoing efforts in data integration, particularly in standardization of ocean data metadata under way as an IODE and Ocean Data Integration Project. International Oceanographic Data and Information Exchange.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 3–4
DOI: 10.15439/2016F597
ACSIS, Vol. 8. ISSN 2300-5963

How Digital Transformation Shapes Corporate IT: Ten Theses about the IT Organization of the Future

Frederik Ahlemann
University of Duisburg-Essen Essen, Germany
Email: frederik.ahlemann@uni-due.de

DIGITAL transformation is a major challenge for many organizations. IT managers in particular not only wonder what the next digital trends in their industry will be, they also need to understand how today's IT organizations will change in light of digital transformation. I will first discuss some foundations of digital transformation and will then present 10 theses on how digital transformation will influence corporate IT.

Thesis 1: *There will be no business without IT: IT is the indispensable driver and enabler of value creation*

IT is already the backbone of many enterprises and a key resource. At the same time, many executives still do not view it as a crucial competitive factor. Digital transformation will change this. IT will not only be used to automate internal and external business processes. It will also be used to realize new digital products, services, and business models. Beyond this, IT will fundamentally change the ways enterprises will be organized and managed. Even many demanding management tasks can be performed by AI systems fueled by machine learning. This will revolutionize how enterprises operate—in terms of speed, reliability, efficiency, and quality. At the same time, business will increase its dependency on IT. System crashes that cannot be fixed immediately will lead to insolvency faster than ever before owing to interrupted business operations.

Thesis 2: *Development and operations will lose weight: Tomorrow's IT functions will follow the paradigm of* Innovate-design-transform

Innovate-design-transform Classic corporate IT follows the *Plan-build-run* paradigm, which serves as a blueprint for the structure of an IT organization and places significant emphasis on long-term planning and subsequent implementation in a more or less stable environment. The focus is on efficiency. However, digital transformation requires enterprises to become more flexible so as to realize innovative business models. Thus, future IT functions will follow the *Innovate-design-transform* paradigm that stress the importance of innovation, subsequent design of IS with high acceptance and adoption rates, and a transformation of the organization to serve the new business model.

Thesis 3: Shadow *IT becomes normal: IT innovations are developed through joint interdisciplinary teams in the business departments*

Today, in most enterprises, IT projects are initiated by the business and then realized in the IT organization. However, this process neglects some key characteristics of innovation in the digital age: (1) innovations must be developed quickly, (2) innovations are the result of a (very) close collaboration of business and IT (and external parties), (3) requirements change rapidly, (4) communication is intense and frequent. Thus, in the future, IT professionals will be part of the business department to continuously work on digital innovations, whether these be process, product, or business model innovations. At the same time, the business can decide, with some constraints, on the IT/IS they need.

Thesis 4: *Innovations through networks: Strategic vendors become innovation partners Most of today's IT organizations struggle to implement*

Most of today's IT organizations struggle to implement disruptive IT innovations because they lack the required capabilities. For instance, many businesses have ideas regarding the use of big data and machine learning. However, they lack the data scientists necessary to implement these concepts. Thus, digital partnerships and digital innovation networks will become more important. These partnerships and networks will often be very different from classic vendor relationships. They are long-term, eye-level, strategic, and will also often involve benefits-sharing.

Thesis 5: *From applications to user: Development processes are agile, user-centric, and closely linked to IT operations*

Even today, software development is often based on the waterfall model. The timespan between an initial idea and handover to IT operations is fairly long. In times of digital innovation, this can significantly hinder the gaining of market shares quickly and the building of positive brand images. Future IT will be lightweight, characterized by agile development processes, and free from too many architectural and organizational constraints. It will allow developers to focus on user needs and user feedback. High acceptance rates and intense usage are primary goals of development. Developers will iteratively improve applications to satisfy users, and mostly in very short cycles. A close integration of development and IT operations is key for this approach (DevOps).

Thesis 6: *Infrastructure as a commodity: IT infrastructure services will be traded on free markets*

Despite the trend towards outsourcing, many companies still operate their own IT infrastructure in data centers. The decision to have an own IT infrastructure is based on a few assumptions. For instance, organizations believe that an own IT infrastructure allows for better controlling, higher security levels, better compliance, and a better cost structure. These assumptions will soon be invalid, if they are not already invalid. In the future, data centers will no longer be necessary, and corporate IT will completely be based on public cloud offerings, with very few exceptions. These public cloud offerings will largely be standardized, allowing for trading at new types of exchanges for IT services.

Thesis 7: *Digital transformation as a major risk: Security and business continuity will be primary cross- departmental functions The growing use of IT—even as parts of products and services*

The growing use of IT—even as parts of products and services—will increase dependency and vulnerability. Attacks against central IT will directly jeopardize a company's continued existence. Thus, security and business continuity management (SBCM) will gain importance. It will pervade all areas of an enterprise and will no longer be a field of only IT experts. Businesses will realize that SBCM is an indispensable tool for long-term business success, because attacks and threats from the outside as well as from the inside will become 'normal'.

Thesis 8: *Transformable IT landscapes: IT architectures will be standardized, modular, flexible, ubiquitous, elastic, cost-efficient, and secure*

Today's complex IT landscapes will undergo tremendous transformation. Via practices such as enterprise architecture management (EAM), increased standardization both at the industry and company levels, technological advances and the trends towards cloud computing, the IT architectures of the future will leave many current challenges behind. We expect them to be more standardized (through cloud computing and industry starndardization), more modular (through technological and architectural advances), more flexible (through technological advances), ubiquitous (through mobile computing, new device categories, more flexible architectures), elastic (through cloud computing), more cost- efficient (through cloud computing), and more secure (through technological advances and cloud providers' expertise). This will allow for more dynamic, fast, and easy implementation of new products, services, and business models.

Thesis 9: *The end of the IT department: IT experts will be part of business departments*

Given that the aforementioned trends come on-stream, the largest share of IT specialists will likely work on the specification, development, configuration/customization, and maintenance of applications. However, because close collaboration with business is key for the success of these activities, we expect them to become part of the business departments, where they can sit next to users and business managers and can develop new products, services, and processes necessary to make IT innovations happen. The IT department will shrink significantly, because fewer infrastructure experts are required (owing to the use of cloud services) and the move of IS specialists to the business. The remaining unit will focus on more strategic tasks such as EAM, SBCM, procurement, partner management, innovation management, and portfolio management. Since these functions are crucial for a corporate success, we expect them to be located alongside the board.

Thesis 10: *Demographics, digital natives, and individual entrepreneurship: Employees become a strategic competitive factor*

A key factor for the success of digital initiatives today and in the future will be access to skilled human resources. Digital transformation requires specific qualifications and skills that are currently fairly rare. Even in the future, with new study programs at universities (e.g. for data science), it is likely that the professionals required will be scarce. The demographic changes in many western societies, the changes in young professionals' value systems, and the growing desire for individualism and self-determination will make a dedicated and innovative IT human resource management necessary.

DOI: 10.15439/2016F596
ACSIS, Vol. 8. ISSN 2300-5963

Location always matters: how to improve performance of dynamic networks?

Michael Segal
Department of Communication Systems Engineering
Ben-Gurion University of the Negev, Israel
Email:segal@bgu.ac.il

IN OUR talk we will focus on networks with no predefined infrastructure (ad-hoc networks, sensor networks, vehicular networks). There are many optimization problems derived from the context of such networks including power assignment mechanisms, scheduling, data gathering, etc. We will discuss various techniques tacking these problems emphasizing the importance of mobile nodes locations and its influence on the tightness of the solutions.

In particular, we consider the following scenarios.

- Wireless Sensor Network. A wireless sensor network (WSN) consists of n wireless sensor nodes, $S = \{s_1, \ldots, s_n\}$, distributed in some area A. These nodes perform monitoring tasks and periodically report to a base station r which is located somewhere within the area A (we consider different locations throughout the paper). During the report phase, the sensor nodes propagate a message to the base station through a *data collection tree*, $T_S = (S \cup \{r\}, E_S)$, rooted at r. We consider *data collection with aggregation*, where every node $s \in S$ forwards a single unit size *report message* to its parent. The message holds an accumulated information collected from a subtree of T_S rooted at s. An example of this scenario can be found in temperature monitoring systems for fire prevention, intrusion detection, seismic readings, etc. **Minimizing the energy requirement** is one of the primary optimization objectives when deploying a WSN due to the very low battery reserves at the sensor nodes and the high costs that are associated with replacing these batteries (if at all possible). The second measure that we are interested in is **transport capacity**, $D(T_S)$, of the data collection tree T_S. Another critical aspect in the design of a WSN is the **hop-diameter** of T_S. We consider different approaches of T_S construction including: short-cutting Minimum Spanning Tree (MST) [1], identification of balance nodes [1], centroid-based constructions [2], (r, d)-index constructions [2].
- Wireless Ad-hoc Network. A wireless ad-hoc network consists of transceivers (nodes) that are located in the plane and communicate by radio. In contrast to wired networks, wireless ad-hoc networks have no fixed communication backbone. The temporary physical topology of the network is determined by the relative disposition of the wireless nodes and the transmission range assignment of each of the nodes. The combination of these two factors produces a directed communication graph where the nodes correspond to the transceivers and the edges correspond to the communication links. The topology of the induced communication graph has a strong effect on the routing algorithms' efficiency. In this talk we will discuss one of the key properties of the induced communication graph – energy stretch factor [3], [4]. Let $\gamma_{u,v}$ be the minimum energy required to send a message from u to v (using other nodes if necessary). The energy spanner is aimed at minimizing the energy stretch factor t_E of the induced communication graph, that is, for any u, v, the energy required to propagate a message from u to v is at most $t_E \cdot \gamma_{u,v}$.
- Vehicular Ad-Hoc Network. Vehicular ad-hoc network (VANET) is a promising branch of traditional MANET. VANET is designed to provide wireless communication between vehicles and between vehicles and nearby road-side equipment. This communication intends to improve both safety and comfort on the road. VANET has a number of difficulties regarding the traditional MANET. Due to the dynamic nature of VANET environments, configuration is always changing, where links may appear and disappear very quickly and vehicle density is constantly changing. In this talk, we also will discuss self-organizing hierarchical topology to serve as the infrastructure for beacon dissemination process in VANET by carefully partitioning the network into geographically optimized clusters with chosen clusterheads [5].

REFERENCES

[1] J. Crowcroft, M. Segal, and L. Levin, "Improved structures for data collection in static and mobile wireless sensor networks," *Journal of Heuristics*, vol. 21, no. 2, pp. 233–256, 2015.
[2] V. Milyeykovski, M. Segal, and V. Katz, "Using central nodes for efficient data collection in wireless sensor networks," *Computer Networks*, vol. 91, pp. 425–437, 2015.
[3] H. Shpungin and M. Segal, "Near-optimal multicriteria spanner constructions in wireless ad hoc networks," *IEEE/ACM Transansactions on Networking*, vol. 18, no. 6, pp. 1963–1976, 2010.
[4] ——, "Improved multicriteria spanners for ad-hoc networks under energy and distance metrics," *ACM Transactions on Sensor Networks*, vol. 9, no. 4, p. 37, 2013.
[5] Y. Allouche and M. Segal, "A cluster-based beaconing approach in vanets: Near optimal topology via proximity information," *ACM Mobile Networks and Applications*, vol. 18, no. 6, pp. 766–787, 2013.

11th International Symposium Advances in Artificial Intelligence and Applications

THE AAIA'16 will bring scientists, developers, practitioners, and users to present their latest research, results, and ideas in all areas of Artificial Intelligence. We hope that theory and successful applications presented at the AAIA'16 will be of interest to researchers who want to know about both theoretical advances and latest applied developments in AI.

TOPICS

Papers related to theories, methodologies, and applications in science and technology in this theme are especially solicited. Topics covering industrial issues/applications and academic research are included, but not limited to:

- Decision Support
- Machine Learning
- Fuzzy Sets and Soft Computing
- Rough Sets and Approximate Reasoning (see also FedCSIS'16 Plenary Panel commemorating Prof. Zdzisław Pawlak)
- Data Mining and Knowledge Discovery
- Data Modeling and Feature Engineering
- Data Integration and Information Fusion
- Hybrid and Hierarchical Intelligent Systems
- Neural Networks and Deep Learning
- Bayesian Networks and Bayesian Reasoning
- Case-based Reasoning and Similarity
- Web Mining and Social Networks
- AI in Business Intelligence and Online Analytics
- AI in Robotics and Cyber-Physical Systems
- AI-centered Systems and Large-Scale Applications

We also encourage researchers interested in the following topics to submit papers directly to the corresponding workshops, which are integral parts of AAIA'16:

- AI in Computational Optimization (see WCO'16 workshop)
- AI in Language Technologies (see LTA'16 workshop)
- AI in Machine Vision and Graphics (see AIMaViG'16 workshop)
- AI in Medical Applications (see AIMA'16 workshop)
- AI in Reasoning and Computational Foundations (see AIRIM'16 workshop)
- AI in Semantic Information Retrieval (see ASIR'16 workshop)
- AI in Spatial and Temporal Analytics (see DSTUI'16 workshop)

All papers accepted to the main track of AAIA'16 and to the above workshops will be treated equally in the conference programme and will be equally considered for the awards listed below.

PROFESSOR ZDZISLAW PAWLAK BEST PAPER AWARDS

We are proud to announce that we will continue the tradition started during the AAIA'06 Symposium and award two "Professor Zdzislaw Pawlak Best Paper Awards" for contributions which are outstanding in their scientific quality. The two award categories are:

- Best Student Paper—for graduate or PhD students. Papers qualifying for this award must be marked as "Student full paper" to be eligible for consideration.
- Best Paper Award for the authors of the best paper appearing at the Symposium.

In addition to a certificate, each award carries a prize of 300 EUR provided by the Mazowsze Chapter of the Polish Information Processing Society.

IFSA AWARD FOR YOUNG SCIENTIST

We are proud to announce that, as in the recent years, the International Fuzzy Systems Association (IFSA) Best Paper Award for Young Scientist, will be presented.

Candidates for all the above awards can come from the AAIA'16 and all workshops organized within its framework.

EVENT CHAIRS

- **Janusz, Andrzej,** University of Warsaw, Poland
- **Ślęzak, Dominik,** University of Warsaw & Infobright Inc., Poland

ADVISORY BOARD

- **Kacprzyk, Janusz,** Systems Research Institute, Warsaw, Poland
- **Kwaśnicka, Halina,** Wroclaw University of Technology, Poland
- **Markowska-Kaczmar, Urszula,** Wroclaw University of Technology, Poland
- **Skowron, Andrzej,** University of Warsaw, Poland

PROGRAM COMMITTEE

- **Artiemjew, Piotr,** University of Warmia and Mazury, Poland
- **Bartkowiak, Anna,** Wrocław University, Poland
- **Bazan, Jan,** University of Rzeszów, Poland
- **Bordogna, Gloria,** CNR IREA, Italy
- **Borkowski, Janusz,** Polish-Japanese Academy of Information Technology & Infobright Inc.
- **Błaszczyński, Jerzy,** Poznań University of Technology, Poland

- **Cetnarowicz, Krzysztof,** AGH University of Science and Technology, Poland
- **Chakraverty, Shampa,** Netaji Subhas Institute of Technology, India
- **Chen, Phoebe,** La Trobe University, Australia
- **Cheung, William,** Hong Kong Baptist University, Hong Kong S.A.R., China
- **Cyganek, Bogusław,** AGH University of Science and Technology, Poland
- **Czarnowski, Ireneusz,** Gdynia Maritime University, Poland
- **Czerniak, Jacek M.,** Casimir the Great University in Bydgoszcz, Poland
- **Czyżewski, Andrzej,** Gdańsk University of Technology, Poland
- **Dardzińska, Agnieszka,** Białystok University of Technology, Poland
- **Dey, Lipika,** Tata Consulting Services, India
- **do Carmo Nicoletti, Maria,** UFSCar & FACCAMP, Brazil
- **Duentsch, Ivo,** Brock University, Canada
- **Eklund, Patrik,** Umeå University, Sweden
- **Froelich, Wojciech,** University of Silesia, Poland
- **Holzinger, Andreas,** Graz University of Technology, Austria
- **Jatowt, Adam,** Kyoto University, Japan
- **Jin, Xiaolong,** Institute of Computing Technology of Chinese Academy of Sciences, China
- **Kayakutlu, Gulgun,** Istanbul Technical University, Turkey
- **Korbicz, Józef,** University of Zielona Góra, Poland
- **Kostek, Bożena,** Gdańsk University of Technology, Poland
- **Krasuski, Adam,** Main School of Fire Service (SGSP), Poland
- **Kryszkiewicz, Marzena,** Warsaw University of Technology, Poland
- **Lopes, Lucelene,** PUCRS, Brazil
- **Madalińska-Bugaj, Ewa,** University of Warsaw
- **Marek, Victor,** University of Kentucky, United States
- **Matson, Eric T.,** Purdue University, United States
- **Menasalvas, Ernestina,** Universidad Politécnica de Madrid, Spain
- **Mercier-Laurent, Eunika,** University Jean Moulin Lyon3, France
- **Mirkin, Boris,** Birkbeck University of London & NRU Higher School of Economics in Moscow, Russia
- **Miyamoto, Sadaaki,** University of Tsukuba, Japan
- **Moshkov, Mikhail,** King Abdullah University of Science and Technology, Saudi Arabia
- **Musiał-Gabryś, Katarzyna,** Bournemouth University, United Kingdom
- **Myszkowski, Paweł,** Wrocław University of Technology, Poland
- **Nguyen, Hung Son,** University of Warsaw, Poland
- **Nourani, Cyrus F.,** Akdmkrd-DAI TU Berlin & Munich Transmedia & SFU Burnaby, Germany
- **Nowostawski, Mariusz,** Norwegian University of Technology and Science (NTNU), Norway
- **Ohsawa, Yukio,** University of Tokyo, Japan
- **Peters, Georg,** Munich University of Applied Sciences, Germany
- **Po, Laura,** Università di Modena e Reggio Emilia
- **Porta, Marco,** University of Pavia, Italy
- **Proficz, Jerzy,** Academic Computer Center, Gdansk University of Technology, Poland
- **Przybyła-Kasperek, Małgorzata,** University of Silesia, Poland
- **Raghavan, Vijay,** University of Louisiana at Lafayette, United States
- **Rao, Raghavendra,** University of Hyderabad, India
- **Rauch, Jan,** University of Economics, Prague, Czech Republic
- **Reformat, Marek,** University of Alberta, Canada
- **Ruta, Dymitr,** Khalifa University, United Arab Emirates
- **Ryżko, Dominik,** Warsaw University of Technology, Poland
- **Santofimia, Maria Jose,** Universidad de Castilla-La Mancha, Spain
- **Schaefer, Gerald,** Loughborough University, United Kingdom
- **Sikora, Marek,** Silesian University of Technology, Poland
- **Su, Chang,** Chongqing University of Posts and Telecommuncations
- **Sydow, Marcin,** Polish Academy of Sciences & Polish-Japanese Academy of Information Technology, Poland
- **Szczęch, Izabela,** Poznań University of Technology, Poland
- **Szczuka, Marcin,** University of Warsaw, Poland
- **Szpakowicz, Stan,** University of Ottawa, Canada
- **Szwed, Piotr,** AGH University of Science and Technology, Poland
- **Tsay, Li-Shiang,** North Carolina A&T State University, United States
- **Unland, Rainer,** Universität Duisburg-Essen, Germany
- **Unold, Olgierd,** Wrocław University of Technology, Poland
- **Wang, Xin,** University of Calgary, Canada
- **Weber, Richard,** Universidad de Chile, Chile
- **Wieczorkowska, Alicja,** Polish-Japanese Academy of Information Technology, Poland
- **Woźniak, Michał,** Wrocław University of Technology, Poland
- **Wróblewski, Jakub,** Infobright Inc.
- **Zadrożny, Sławomir,** Systems Research Institute of Polish Academy of Sciences, Poland
- **Zakrzewska, Danuta,** Łódź University of Technology, Poland
- **Zielosko, Beata,** University of Silesia, Poland
- **Ziółko, Bartosz,** AGH University of Science and Technology, Poland

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 9–15
DOI: 10.15439/2016F543
ACSIS, Vol. 8. ISSN 2300-5963

Quality of Histograms As Indicator Of Approximate Query Quality

Agnieszka Chądzyńska-Krasowska
Polish-Japanese Academy of Information Technology
Koszykowa 86, 02-008 Warsaw, Poland
Email: honzik@pjwstk.edu.pl
Infobright Inc.
ul. Krzywickiego 34/219, 02-078 Warsaw, Poland
Email: agnieszka.chadzynska-krasowska@infobright.com

Marcin Kowalski
Infobright Inc.
ul. Krzywickiego 34/219, 02-078 Warsaw, Poland,
Email: marcin.kowalski@infobright.com

***Abstract*—We consider concept of *approximate query* in RDBMS i.e. query that returns results which may differ from common (exact) query results in a way but its evaluation requires less resources. In the work we focus mostly on time and storage space aspects. We follow one of the state-of-the-art trends using synopses of data as the input of approximate query evaluation. We propose some measures of approximate query results quality. Basing on them we present steps of adaptive elaboration of synopses quality measure that should be mutually corresponding.**

***Index Terms*—Approximate Query, Quality Measures, Histograms.**

I. Introduction

APPROXIMATE query concept emerged as a tool of coping with continuous growth of data volume gathered in databases. Disposing limited budget of resources like time, storage space, computing power etc. database user wanted to gain information from the data quickly but accepting fact that achieved results may not be crisp. Two main classes of solutions are present in the literature: data sampling techniques [1], [2] and data synopses calculation [3], [4]. The former one utilizes statistical apparatus for choosing representative sample of the data, considerably smaller than whole data set, and using it to estimate results of query for whole data. The latter is based on concepts of data synopses which are data descriptions built once during load, stored and exclusively used during query evaluation. Some compact comparison between both approaches is contained in e.g. [5].

In the world of RDBMSs synopses are considered as the descriptions of e.g. columns value data sets. We may consider both one or multicolumns descriptions. The most common and well-examined types of data synopses described in the literature are histograms. In standard approaches, histograms are built per whole column (or more columns) and inference based on histograms is done for whole relation.

Our approach differs from this while we utilized some elements of *granular computing*[1] in the query evaluation process. Despite our approach may be treated as an example of the synopses calculus trend, we create synopses and inference on their basis not on whole relation level but on their parts. This may require some additional operations during query evaluation like compound partial results of evaluated query from each data packs. On the other hand it enables reflecting potential changes or differences in columns value sets in time and avoiding complex operations on synopses when e.g. consecutive loads into relations are considered.

Other consequence of this fact is that the size (in bytes) of generated synopses must be of orders of magnitude smaller than compressed data themselves and synopses build in standard approaches for whole column. For effectiveness sake and in order to easily control storage we assume also existing size budget for single synopsis. That means in particular that for most cases we cannot afford storing e.g. exact histogram of value set from part of the table (i.e. histogram in which all intervals are single-valued). The idea of utilizing granularity concepts may go further in building hierarchical structures of synopses e.g. for subsets of row/data packs.

The presented analysis concerns measures of histogram quality. We will say that the histogram is of good quality when applying used methodology led to achieve good approximation of query results. That raises immediately question of quality of the latter one. In the article we introduce measures of quantifying approximate query results and their selected aspects. The main purpose of the work remains though developing and analyzing measures of histogram quality that will directly translate to proposed quality measures of the query results.

Section II describes basics of the Infobright RDBMS engine that we had chosen for experiments, explains terminology we use in the work and presents methodology used in experiments framework. We discuss problem of measuring approximate query results in Section III and propose some formulas for it. In Section IV we describe standard methods of 1dim histogram generation and classic quality measures for them. As the experiments result presented in latter section run for standard approaches were not satisfactory, we propose in Section V some modifications of them.

[1]https://en.wikipedia.org/wiki/Granular_computing

II. Basic Definitions and Methodology

A. Definitions

We consider *1–dimensional histograms* as a collection of *bars* (*buckets*), each of which then consists of: interval (determined by 2 numbers - interval's begin and end) and *frequency* which in our work is cardinal (but finite) number. If single histogram bars' intervals are disjoint, frequencies should be interpreted as number of elements which occur in the interval. Exemplary 1dim histograms are in Figures 3 – 5 . We may easily extend this definition to *multidimensional histograms* changing intervals to *cubes* however we do not consider multidimensional case in the article so we omit it here. Histograms might be generated using many algorithms. Some of them are presented in Section IV. In the work we may refer to information contained in synopses as to *rough data* or *granuled data*.

B. Infobright Basics

In the considered RDBMS engine [6], [7], during the data load process, the sets of relation tuples are split into chunks of equal and fixed cardinality (apart from the last chunk perhaps) which form *row packs*. Sequence of values from row pack for single column forms *data pack* for this column. As *data pack range* we mark interval from minimal to maximal values of data pack. In actual implementation of the database engine chunk size is equal to 2^{16}. Data packs are then compressed separately and stored. For each of formed data pack there is additionally built and stored compact information (data synopsis) of values from data pack. Such synopsis may be considered as the information granule of the data pack. Prepared data synopses are loaded into memory during query evaluation and used in several ways e.g. to filter out irrelevant data (e.g. to avoid costly I/O operations for corresponding data packs).

As pointed out, in the described engine, both compressed data packs and synopses are stored. In some of our previous works, we investigated how to further decrease the amounts of accessed data packs and focus more on synopses-based computations in order to accelerate computations, on the cost of possible inexactness of query results [8], [9]. However, in the case of the new Infobright's engine dedicated to approximate queries only synopses are stored, while exact data are accessible only during their load (when synopses are being generated) and then forgotten.[2] This is also the assumption that we follow in this paper. Another aspect that distinguishes our approach from most of standard methods is the fact that our data synopses are built for each data pack – not for each column only.

C. Experiments Framework

As there is no room for describing methods of using synopses in internal operations of the query engine and, on the other hand, this is not the main purpose of this article, we decide to emulate behaviour of the engine designed for approximate queries. Generally speaking, histogram reflects probability distribution and therefore might be the input of random value generator. To make histogram the distribution we assume that (1) inside each interval all values are uniformly distributed and (2) frequencies of each value in domain should be normalized (their sum should be equal to 1).

During query evaluation histograms were read from storage and data pack is constructed of randomly chosen 2^{16} column values according to distribution described by histogram.

For the purpose of this work we assume, that each approximate query evaluation consists of 2 stages: (1) generating data according to calculated 1dim histograms and (2) evaluate the query using such prepared (approximate) table with exact engine. We created set of benchmark queries which were evaluated on data sets generated according to specific histograms. The overall rating of specific histogram depended on query results quality gained from experiments on these data sets (we will call it *approximate data*).

D. Storage Budget Discussion

As we had chosen histograms as types of data synopsis used in our framework, we should indicate how one can control existing storage budget in histogram construction. Remind that we consider budget per 1 data pack. Few question may be raised here. The most natural way to control budget in the case of histograms is to adjust number of stored buckets. This parameter affects not only storage factor but has also influence on calculation complexity (number of performed operations). The other factor may be also method of storing histogram. In general, as presented in definition, each bucket of histogram is described by left and right boundary. However with additional assumption (made also in this work) that set of histogram intervals covers whole data pack domain, apart from storing data pack minimum and maximum, to identify the bar in histogram suffice to store only its e.g. right boundary (left can be induced from prior interval). So each bucket can be stored as 2 numbers - its maximum and its frequency. Deeper optimizations of storage may depend on type of synopsis or on implementation method. In the work we focus on number of buckets as the main control parameter.

III. How to Compare Query Results?

In order to compare query results evaluated on exact (real) data and on rough data one needs to elaborate quality measures of query results. Below we describe measures used in the article. One can easily notice that they are to some extent inspired by the notions of fuzzy similarity and multi-label classification, now adopted for new purposes.

The result of every SELECT has a tabular form. We will denote it as $R = (U, C)$, where U and C are sets of its tuples and columns. U can refer to original rows or groups, possibly with limit.

In C, we distinguish columns A that are results of aggregate functions and columns G used in group by clause. Alternatively, G can gather columns that are primary key and A - all

[2]https://infobright.com/introducing-iaq/

other columns. Some columns in A can be used in ORDER BY clause. In such case the new rank function is added to C.

Let us consider two results of the same query, real result $R_r = (U_r, C)$ and an approximate result $R_a = (U_a, C)$. The idea is to check to what extent tuples in U_r and U_a match with each other with regard to columns in G and, for those tuples which can be matched, how similar are their values over A. Similarity of R_r and R_a should be within [0,1], and equal to 1 only if R_r and R_a are identical.

Consider a pair of tuples $t_r \in U_r$ and $t_a \in U_a$, such that there is $g(t_r) = g(t_a)$ for every g in G.

A score of similarity of t_r and t_a is as follows:

$$s(t_r, t_a) = \prod_{c \in A} s_c(c(t_r), c(t_a))$$

where $\prod$ denotes multiplication and:

$$s_c(c(t_r), c(t_a)) = 1 - \frac{|c(t_r) - c(t_a)|}{|c(t_r)| + |c(t_a)| + 1}$$

If rank is added, it can take a form of:

$$r(t_r, t_a) = 1 - \frac{|rank(t_r) - rank(t_a)|}{|rank(t_r)| + |rank(t_a)|}$$

While rows (and groups) from results achieved from exact query and approximate query may differ we distinguish two types of mismatched rows (groups):

- False Positives (FP) - rows (groups) which shouldn't occur but did

$$FP = card(U_a \backslash U_r) = card(U_a) - card(U_r \cap U_a)$$

- True Negatives (TN) - rows (groups) which should occur but didn't

$$TN = card(U_r \backslash U_a) = card(U_r) - card(U_r \cap U_a)$$

We define ***Aggregation Similarity*** as follows:

$$AggSim(R_r, R_a) = \frac{\sum_{t_r \in U_r, t_a \in U_a : G(t_r) = G(t_a)} s(t_r, t_a)}{card(U_r)}$$

We define ***Ranking Similarity*** as follows:

$$RankSim(R_r, R_a) = \frac{\sum_{t_r \in U_r, t_a \in U_a : G(t_r) = G(t_a)} r(t_r, t_a)}{card(U_r)}$$

And finally, we define ***Total Similarity*** as follows:

$$TotSim(R_r, R_a) = \\ = \frac{\sum_{t_r \in U_r, t_a \in U_a : G(t_r) = G(t_a)} s(t_r, t_a) \cdot r(t_r, t_a)}{card(U_r) + card(U_a \backslash U_r)}$$

In Fig. 1 there is a simple example of calculation of presented measures.

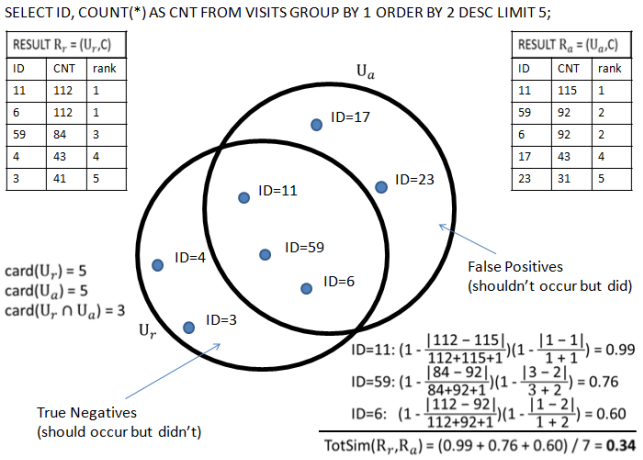


Fig. 1. Example of comparison between results of exact and approximate queries

IV. Standard Approach

A. Generation of 1-Dimensional Histogram

We had started experiments with three standard approaches to histogram generation.

As mentioned earlier we built separate histogram for each data pack (value set of 1 column from 2^{16} consecutive rows) i.e. exact data were the input to create each 1-dim histogram. After being build each histogram was stored.

Looking for the best correspondence between histogram quality measures and approximate query results quality we tested many different methods of histogram generation. Here we present results of experiments on 3 most commonly used in databases types of histograms: EquiWidth histogram (classic), EquiDepth histogram (quant) and MaxDiff histogram (diff).

Each type of histogram splits data pack range into k buckets.

- EquiDepth histogram divides the set of values into k ranges such that each range has the same number of values [10].
- EquiWidth histogram divides the set of values into k buckets of equal width [11].
- In MaxDiff histogram boundaries of intervals are chosen after analyzing differences between frequencies of adjacent bars from exact histogram (adjacency is induced by natural order of values). $k - 1$ largest differences determine split points of histograms intervals [12].

As inside each histogram's interval we assume uniform distribution for contained values, and intervals from each considered 1-dim histograms cover whole range of data pack, we may achieve in natural way from random generation values which did not occur in original data pack (False Positives). As turned out they constituted the biggest challenge.

Each type of histogram splits data pack range into k buckets, where k is established parameter corresponding to storing budget. In our experiments we took $k = 64$.

All experiments were run on the table containing 33 columns and $10 \cdot 2^{16}$ rows.

B. Standard Quality Measures for Histogram

In order to use approximate query in most efficient way not only results comparison aspect should be examined but also methods of prediction to what extent type or scale of used synopsis could affect query results quality. Such knowledge would let one to choose optimal type/size of histogram and also to control by the user threshold between storage footprint of created synopses (or on the other hand: volume of data processed during query evaluation) and query results quality. We assume that the database user is aware of the existence of such threshold.

That brings us to subproblem of defining measure of histogram quality which is correspondent to approximate query results.

We acknowledged as fundamental here the ability of generated histogram to reflect original distribution of the data pack's values. As mentioned earlier due to existing budget, in most cases, we are not able to store exact histogram of value set (such case we recognize as optimal), so such possibility is crucial to make more sophisticated inference of approximations successfully.

At the first stage we applied standard measures of similarity between distributions based on deviations from mean, which work fine for the cases without generated false positives [13], [14].

Here is the first examined measure:

$$Q1(c,p) = \sum_{v \in Dom(c)} min\{\frac{freq(p_v)}{ALL_{p_v}}, freq(v)\},$$

where c is a column, p - data pack range split (set of intervals), p_v - interval (from split) containing v, $freq(p_v)$ - frequency of p_v, $freq(v)$ - number of occurences of v and $ALL_{p_v} = end(p_v) - start(p_v) + 1$. Intuitions of measure $Q1$ are presented at Fig. 2

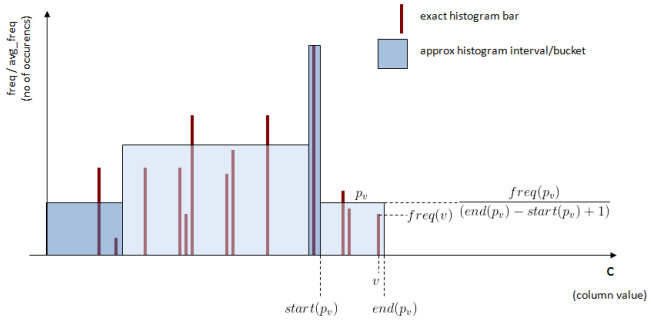


Fig. 2. Measure $Q1$ components.

The second analyzed measure was the sum of squares of deviations of exact value frequencies from corresponding bar frequency, with additional assumption that all values from intervals range is included in the sum (those which did not exists in original data have frequency equal to 0)

$$Q2(c,p) = \sum_{v \in ALL_{Dom(c)}} (freq(v) - \frac{freq(p_v)}{ALL_{p_v}})^2,$$

where c is a column, p - data pack range split (set of intervals), p_v - split interval for v, $freq(p_v)$ - frequency of interval containing v, $freq(v)$ - number of occurrences of v in data pack (non existing values have freq = 0), $ALL_{p_v} = end(p_v) - start(p_v) + 1$ and $ALL_{Dom(c)} = end(Dom(c)) - start(Dom(c)) + 1$.

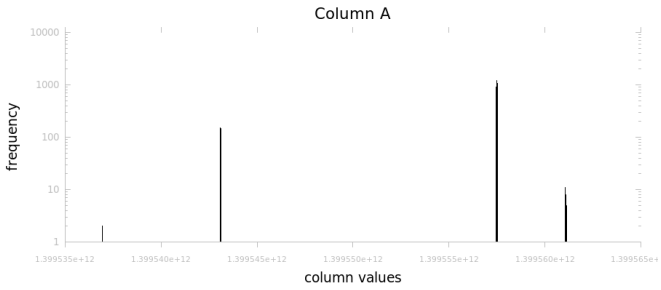


Fig. 3. Exact histogram (distribution) on 1 data pack of real-life column (Column A; logarithmic scale of frequencies)

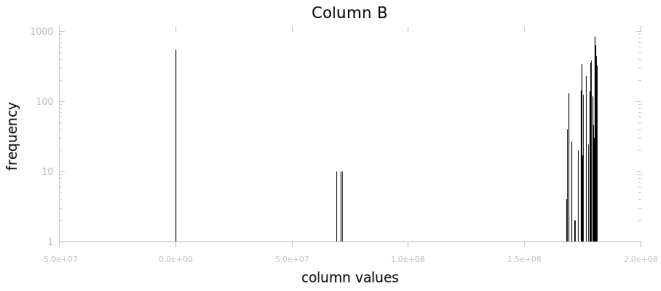


Fig. 4. Exact histogram (distribution) on 1 data pack of real-life column (Column B; logarithmic scale of frequencies)

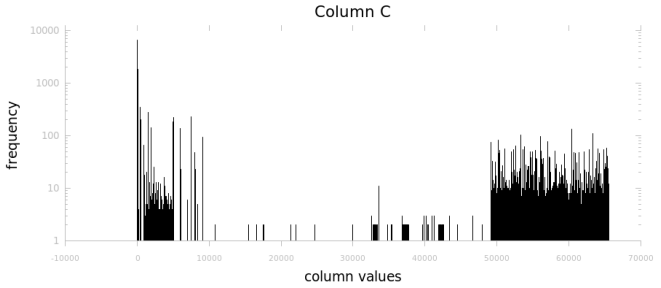


Fig. 5. Exact histogram (distribution) on 1 data pack of real-life column (Column C; logarithmic scale of frequencies)

At the first stage of experiments we calculated $Q1$ for every considered type of histograms. Next we evaluated set of prepared queries on exact data and on data generated from every histogram. We found achieved results encouraging, however we identified many hard cases for presented approach.

Lack of satisfactory correspondence between $Q1$ and query results quality for each histogram was visible both on simplest single-column queries and more complex queries. Fig. 6–9 illustrate results of testing query that simulate calculation of exact distribution of column values:

```
SELECT col, count(*) FROM t GROUP BY col
```

We chose exemplary columns from real data set and mark them as A, B, C. Their real distributions are presented on Fig. 3–5.

Calculated value of $Q1$ was additionally divided by number of rows in table t. The higher the value of $Q1$ the better quality of the histogram.

Column	Metric	Diff	Classic	Quant
A	Q1 Measure	0.8620	0.0601	0.8460
	TotSim	0.0281	0.0014	0.0351
B	Q1 Measure	0.1086	0.0004	0.0925
	TotSim	0.0036	0.0002	0.0075
C	Q1 Measure	0.4812	0.3612	0.3674
	TotSim	0.4562	0.5192	0.4520

Fig. 6. Illustration of lack of satisfactory correspondence between $Q1$ and quality of approximate queries results. Diff, Classic and Quant stand for MaxDiff, EquiWidth and EquiDepth histograms respectively. Values of $Q1$ are averages from 10 data packs.

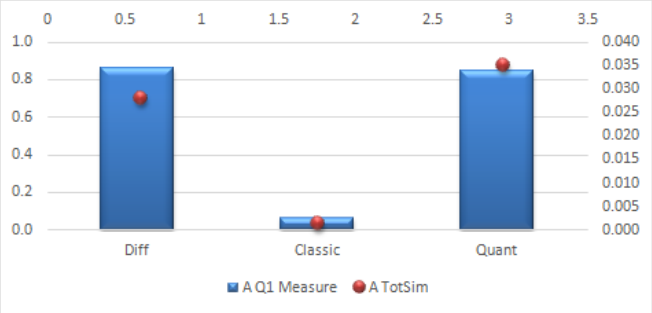


Fig. 7. Lack of satisfactory correspondence between $Q1$ and quality of approximate query on column A

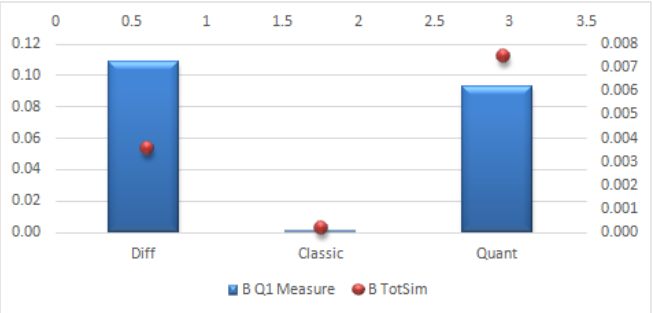


Fig. 8. Lack of satisfactory correspondence between $Q1$ and quality of approximate query on column B

Similar tests were run for $Q2$. Here also observed correspondence between histogram quality and quality of approximate query results was disappointing. Results are presented

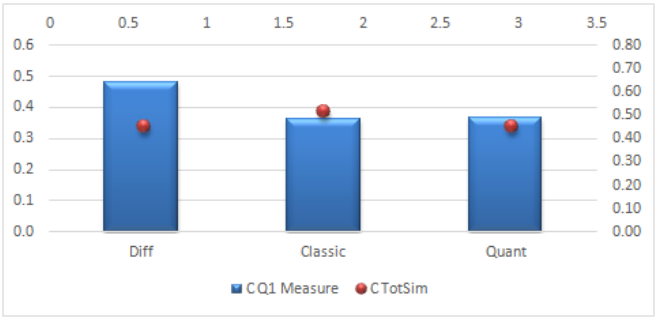


Fig. 9. Lack of satisfactory correspondence between $Q1$ and quality of approximate query on column C

on figures 10–13. Analogously value of $Q2$ was divided by number of rows in the table. In contrast to the measure $Q1$, the higher the value of $Q2$, the worse the quality of the histogram.

Column	Metric	Diff	Classic	Quant
A	Q2 Measure	330.669	981.971	351.434
	TotSim	0.028	0.001	0.035
B	Q2 Measure	2 785.017	2 837.871	2 821.522
	TotSim	0.004	0.0002	0.007
C	Q2 Measure	4 446.830	5 694.313	5 587.474
	TotSim	0.456	0.519	0.452

Fig. 10. Table illustrating lack of satisfactory correspondence between $Q2$ and quality of approximate queries (on data generated according to specified histogram)

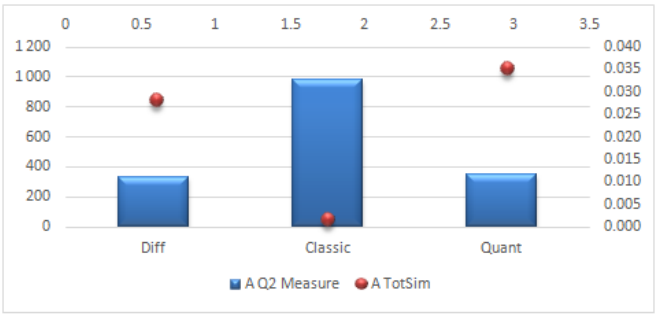


Fig. 11. Lack of satisfactory correspondence between $Q2$ and quality of approximate query on column A

V. Specific Histogram Quality Measures

Because of discrepancies between both $Q1$, $Q2$ and approximate query results quality, some new histogram quality measures had to be developed. Discrepancies analysis revealed that the main reason of inadequacy of applied measures was assumption of uniform distribution of values frequencies inside each interval. In particular no information of gaps in data packs domain was involved. By *gap* (if exists) we take interval between two consecutive values that exists in original data set. More formally:

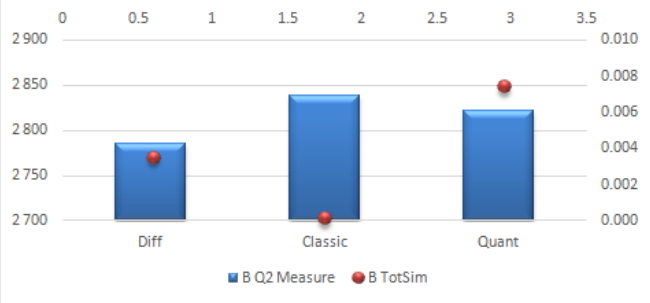


Fig. 12. Lack of satisfactory correspondence between $Q2$ and quality of approximate query on column B

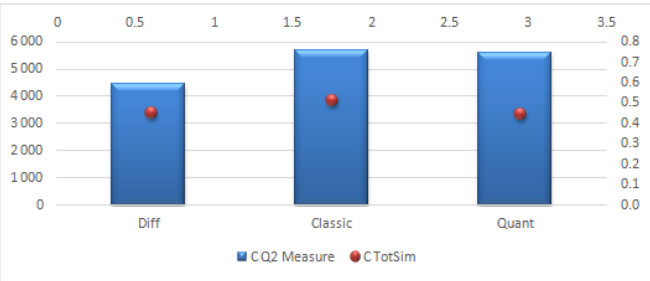


Fig. 13. Lack of satisfactory correspondence between $Q2$ and quality of approximate query on column C

Definition. Let $v, w \in Dom(c)$ be two consecutive (in the sense of linear order on values of column c) values that exist in considered value set. Then by *gap* we take set $\{x \in ALL_{Dom(c)} : x > v \wedge x < w\}$

As pointed out earlier because of method of data generation from considered types of histograms, one is exposed to generating false-positive cases. Gaps (especially wide ones), present in original data set would cause deterioration of quality of approximate query results in two ways. First – due to numerous false-positive cases and in the consequence also true-negative cases – it will reduce Total Similarity value. On the other hand too voluminous domain of interval makes average frequency lower and as a result may decrease Aggregation Similarity.

After testing some preliminar candidates, we introduced measure of the histogram which indicates difficulty of gaining proper correspondence between histogram quality and approximate query quality:

$$Q_{HG}(c,p) = \sum_{i=1}^{p_{cnt}} freq(p_i) \cdot FALSE(p_i)$$

, where c stands for column, p - split (set of intervals), p_{cnt} - number of intervals, $freq(p_i)$ - frequency of ith interval, and $FALSE(p_i)$ - number of false-positives generated in ith interval (which does not exist in original data set).

Q_{HG} was divided by number of rows in table t. The higher the value of Q_{HG}, the worse the quality of the histogram. Figures 14–16 show that Q_{HG} does not correspond to approximate query quality in direct way, however we may notice some regularities. First, calculated value of Q_{HG} explains poor quality of query results for columns A and B. For column C considered query returned better approximations and corresponding value of Q_{HG} was significantly lower than for A and B.

Column	Metric	Diff	Classic	Quant
A	Q_{HG} Measure	66 441 005.00	13 120 751.37	71 920 778.98
	TotSim	0.028	0.001	0.035
B	Q_{HG} Measure	71 474 878.37	36 296 002.79	60 186 447.39
	TotSim	0.004	0.0002	0.007
C	Q_{HG} Measure	1 087.03	588.48	859.77
	TotSim	0.456	0.519	0.452

Fig. 14. Illustration of the rule: high value of Q_{HG} corresponds to poor approximate query results quality

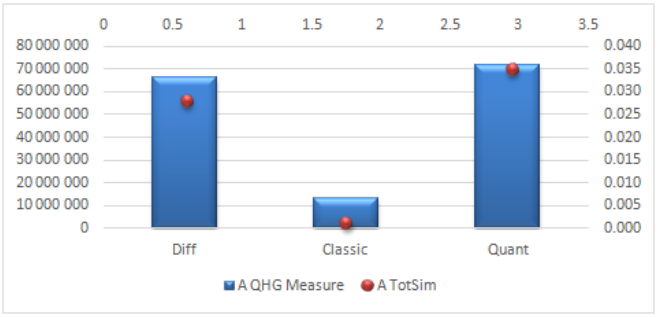


Fig. 15. Illustration of the rule: high value of Q_{HG} corresponds to poor approximate query results quality - for column A

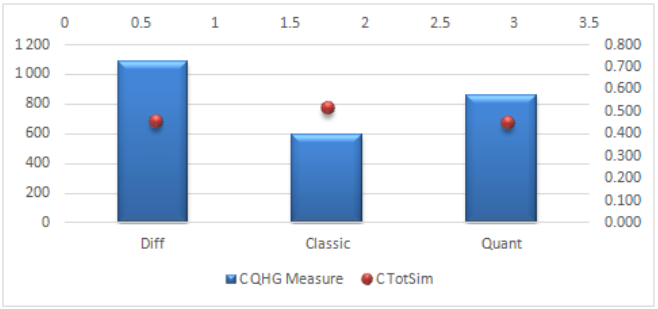


Fig. 16. Illustration of the rule: low value of Q_{HG} corresponds to well-approximated query results quality - for column C

To confirm hypothesis of correspondence between high value Q_{HG} and poor query results, we perform more experiments. We defined importance ranking formula for gaps and added to synopsis information of 100 most important gaps per data pack.

Definition. Gap importance ranking for each data pack (of column c) with given split p is defined as follow:

$$RANK\_GAP_{c,p}(gap) = freq(i(gap)) \cdot FALSE(gap),$$

where $freq$ stands for interval frequency, $i(gap)$ stands for index of interval containing gap , and $FALSE(gap)$ stands for number of false-positives covered by gap.

We consider also the case where data packs differ between themselves w.r.t. contained gaps. We observed such situation in real life data. In such case we may consider aggregated budget for gaps for several e.g. consecutive data packs (not per single data pack like in assumptions made at the beginning). Such modification enable to assign parts of the budget to data packs in more flexible way (bigger part to data packs with a lot of gaps, smaller - to the data packs with a few). In next series of experiments instead of 100 most important gaps per data pack we add information of 1000 most important gaps but for 10 data packs (whole column value set). Importance we calculate using formula:

Definition. Holistic gap importance ranking for column c and given split p is defined as follow:

$$BIGRANK\_GAP_{c,p}(gap) = \\ = \sum_{j=1}^{10} freq_j(i(gap)_j) \cdot FALSE_j(gap),$$

where $freq_j$ stands for frequency of interval in jth data pack, $i(gap)_j$ stands for index of interval in jth data pack containing gap, and $FALSE_j(gap)$ stands for number of false-positives covered by gap in jth data pack.

At Fig. 17 we presented dependency between order of magnitude of Q_{HG} and approximate query results quality. For clarity sake dependency was shown only for MaxDiff histogram (for other types dependency is the same). We can observe that if value of Q_{HG} is not low there is no chance to achieve good results of approximate query. However relatively small values of Q_{HG} would not guarantee good quality of approximate query results, and strength of correspondence is depended on chosen column. Therefore, we conclude there are some other factors that affect approximate query results quality.

Column	Metric	Diff	RANK Diff	BIGRANK Diff
A	Q_{HG} Measure	66 441 005.00	14.02	19.95
	TotSim	0.028	0.489	0.713
B	Q_{HG} Measure	71 474 878.37	206 719.29	844.31
	TotSim	0.004	0.011	0.171
C	Q_{HG} Measure	1 087.03	553.17	561.79
	TotSim	0.49	0.59	0.69

Fig. 17. Correspondence between order of magnitude of value Q_{HG} and approximate query results quality.

VI. Conclusions and Future Work

In the article we present preliminary method for calculation of quality of histogram representing original column values in data set. Such measure should in the assumption correspond to quality of approximate query run over data generated from the histogram. Development of such measure would allow at the loading stage to construct histogram reflecting the input data in the best way with limited storage budget maintained.

Performed experiments confirm that many features of input histograms may have influence on quality of approximate query results. Some of them are not identified yet. We suspect such characteristics might be related to intervals width or - like in case of gaps - to distributions of exact values frequencies inside each interval. Probably most of these factors may be expressible in terms of statistical measures of dispersion, symmetry or skewness. We can adapt them to histogram quality formulas however proper choice and the way of applying chosen measure requires much more experimental work.

References

[1] V. Ganti, M.-L. Lee, and R. Ramakrishnan, "Icicles: Self-tuning samples for approximate query answering." in *VLDB*, vol. 176. Citeseer, 2000, p. 187.

[2] S. Chaudhuri, G. Das, and V. Narasayya, "Optimized Stratified Sampling for Approximate Query Processing," *ACM Trans. Database Syst.*, vol. 32, no. 2, p. 9, 2007.

[3] K. Chakrabarti, M. N. Garofalakis, R. Rastogi, and K. Shim, "Approximate Query Processing Using Wavelets," *VLDB J.*, vol. 10, no. 2-3, pp. 199–223, 2001.

[4] G. Cormode, M. N. Garofalakis, P. J. Haas, and C. Jermaine, "Synopses for massive data: Samples, histograms, wavelets, sketches," *Foundations and Trends in Databases*, vol. 4, no. 1-3, pp. 1–294, 2012. doi: 10.1561/1900000004. [Online]. Available: http://dx.doi.org/10.1561/1900000004

[5] B. Mozafari and N. Niu, "A handbook for building an approximate query engine," *IEEE Data Eng. Bull.*, vol. 38, no. 3, pp. 3–29, 2015. [Online]. Available: http://sites.computer.org/debull/A15sept/p3.pdf

[6] D. Ślęzak and V. Eastwood, "Data warehouse technology by infobright," in *Proceedings of the 2009 ACM SIGMOD International Conference on Management of data*. ACM, 2009, pp. 841–846.

[7] D. Ślęzak, P. Synak, A. Wojna, and J. Wróblewski, "Two database related interpretations of rough approximations: Data organization and query execution," *Fundam. Inform.*, vol. 127, no. 1-4, pp. 445–459, 2013. doi: 10.3233/FI-2013-920. [Online]. Available: http://dx.doi.org/10.3233/FI-2013-920

[8] D. Ślęzak and M. Kowalski, "Towards approximate SQL–Infobright's approach," in *Rough Sets and Current Trends in Computing*. Springer, 2010, pp. 630–639.

[9] M. Kowalski, D. Ślęzak, and P. Synak, "Approximate assistance for correlated subqueries," in *Proceedings of the 2013 Federated Conference on Computer Science and Information Systems, Kraków, Poland, September 8-11, 2013.*, M. Ganzha, L. A. Maciaszek, and M. Paprzycki, Eds., 2013, pp. 1455–1462. [Online]. Available: http://fedcsis.org/2013/

[10] S. Chaudhuri, "An overview of query optimization in relational systems," in *Proceedings of the seventeenth ACM SIGACT-SIGMOD-SIGART symposium on Principles of database systems*. ACM, 1998, pp. 34–43.

[11] R. P. Kooi, "The optimization of queries in relational databases," 1980.

[12] V. Poosala and Y. E. Ioannidis, "Selectivity estimation without the attribute value independence assumption," in *VLDB*, vol. 97, 1997, pp. 486–495.

[13] Y. E. Ioannidis and V. Poosala, "Balancing histogram optimality and practicality for query result size estimation," in *ACM SIGMOD Record*, vol. 24, no. 2. ACM, 1995, pp. 233–244.

[14] H. V. Jagadish, N. Koudas, S. Muthukrishnan, V. Poosala, K. C. Sevcik, and T. Suel, "Optimal histograms with quality guarantees," in *VLDB*, vol. 98, 1998, pp. 24–27.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 17–20
DOI: 10.15439/2016F266
ACSIS, Vol. 8. ISSN 2300-5963

Verifying cuts as a tool for improving a classifier based on a decision tree

Łukasz Dydo, Jan G. Bazan, Sylwia Buregwa-Czuma, Wojciech Rząsa
Interdisciplinary Centre for Computational Modelling, University of Rzeszow, Pigonia 1, 35-310 Rzeszow, Poland
Email: {ldydo, bazan, sczuma, wrzasa}@ur.edu.pl

Andrzej Skowron
Institute of Mathematics, The University of Warsaw
Banacha 2, 02-097 Warsaw, Poland and
Systems Research Institute Polish Academy of Sciences
Newelska 6, 01-447 Warsaw, Poland
Email: a.skowron@mimuw.edu.pl

***Abstract*—This article is a continuation of previous work, in which a new method of decision tree construction was presented. That method is based on the use of so-called verifying cuts, which can provide knowledge obtained from the attributes frequently eliminated when greedy methods of the choice of singleton best cuts are applied. Till now only one strategy of choosing verifying cuts was examined. It exploits a measure based on a number of pairs of objects discerned by a chosen cut. In this paper, we examine two additional measures used for determining the best verifying cuts. They are based on Gini's Index and Entropy. The paper includes the results of experiments that have been performed on data obtained from biomedical database and machine learning repositories.**

I. Introduction

DECISION tree with verifying cuts [1] (denoted by v-tree) is a method of decision tree construction formed in response to the problem of classification data with a large number of attributes. Such data can contain a lot of attributes that bear similarity with respect to the quality of potential cuts but significantly different with respect to domain knowledge represented. In contrast to the method of classifier construction based on the decision tree with local discretization techniques known from literature (see, e.g., [2], [4], [8]), for which always singleton best cuts are used and therefore there are serious doubts as to the validity of such approach, v-tree uses the so-called verifying cuts. They are additional cuts, which enable to evaluate the quality of cuts in tree nodes during classification of objects. Our experiments conducted on real data and described in [1] have shown that data with numerous set of attributes constitutes a class of data where v-tree outperforms the conventional approach. However, only one technique of choosing verifying cuts was tested, namely the one based on the maximization of the number of discerned object pairs with different decision class membership. Accordingly, the following hypothesis arose. Perhaps, using another measures of determining the quality of verifying cuts v-tree can enhance its effectiveness. Another question is whether new techniques of choosing verifying cuts may be helpful in the case of input data with non-numerous set of attributes. In this paper two another techniques of verifying cuts construction based on the Entropy and the Gini's Index are tested. Furthermore, we conducted comparative experiments using these two approaches and the one described in [1].

II. Classical discretization tree

Decision tree of the local discretization [2] is a technique of a binary tree construction based on supervised discretization which introduces iterative binary partitioning of data set into groups with respect to the value of certain attribute. This algorithm is well known from literature (see, e.g., [4], [8]), therefore we will refer to it as the *classical method*.

The greedy method of choosing a pair - an attribute and its value (for numeric attributes often called the cut), which are used in the process of data partitioning - is a key element of the discussed local discretization tree construction method and is taking into consideration decision attribute values of training objects. In construction of local discretization tree we decided to use two various measures of best cut, i.e., Information Gain and Gini's Index.

1) Information Gain measure: First method for calculating quality of cuts that was chosen for our research is Information Gain - approach used in C4.5 algorithm [9]. The method uses concept of Entropy which was described by Claude Shannon in his work on information theory [10]. In relation to construction of decision trees of the local discretization, this measure represents diversity of objects set that corresponds to particular node in tree. Thus, let X be the set of objects which comprises of two decision classes - C_0 and C_1. Furthermore, $p_0 = \frac{|C_0|}{|X|}$ and $p_1 = \frac{|C_1|}{|X|}$ are the distribution of C_0 and C_1 in the set X. Therefore the entropy is calculated by the following expression: $Entropy(X) = -\sum_{i=0}^{1} p_i * \log_2 p_i$
The quality of a binary partition, which is defined according to the cut value c in the set X of objects, is computed by Information Gain measure as follows.

$$Gain(c, X) = Entropy(X) - \sum_{i=0}^{1} \frac{|X_i|}{|X|} * Entropy(X_i) \quad (1)$$

where X_i for $i = 0, 1$ are subsets of X, that corresponds to split which is defined by cut value c.
The value of information gain is determined for all possible cuts and next a cut is greedily chosen which maximizes that measure. Surely, this method can be generalized to greater number of decision classes than 2.

978-83-60810-90-3/$25.00

2) Gini's Index measure: An alternative example of measuring the quality of cuts, that was used in our work, is the method used in CART algorithm [5] - Gini's Index. For indications such as in the previous paragraph, if X contains examples from classes C_0 and C_1, the measure of diversity of X set is defined as $Gini(X) = 1 - \sum_{i=0}^{1} p_i^2$
where p_i is the class distribution in X. Moreover, the quality of cut can be calculated as fallows.

$$G(c, X) = Gini(X) - \sum_{i=0}^{1} \frac{|X_i|}{|X|} * Gini(X_i) \qquad (2)$$

As previously, the best cut is chosen greedily from all possible cuts. Furthermore, this approach can also be generalized to more than two decision classes.

Binary tree classifiers for which Information Gain or Gini's Index were used in the procedure of best cut finding we call in this paper the Entropy-C classifier and the Gini-C classifier, respectively.

III. Decision tree with verifying cuts

As in the previous article [1], the motivation for our work concerns the validity of classical approach used to data sets with large number of attributes. We recall that the method chooses only one split (for a single attribute) with the best quality based on the selected measure, at the given step of searching for optimal binary partitions. In such case, the method would greedily eliminate the information contained in attributes, which are similar in terms of quality of potential cuts, but are different with respect to domain knowledge, which they represent. The main idea presented in [1] is based on the fact that at a given stage of searching for partitions of a set of attributes, the family of k-binary verifying partitions is determined after construction of the optimal binary partition of a set of objects. Obviously, it refers to family of partitions which are similar to the optimal partition and concerns other attributes than the attributes used in the optimal partition. Moreover, the concept of similarity depends on measure which is used to determine the best split. Thus, in the case called *MaxDiscPair*, the similarity means to distinguish between set of pairs of objects of different decision classes as similar as possible to the optimal partition. Whereas in the case of measures based on Gini's Index and Information Gain, verifying partitions should separate objects in possible the same manner as the main split. The differences in selecting verifying cuts between all three measures are such that in case of discernibility-based measure we determine objects that are separated by both cuts simultaneously, while in case of Gini's Index and Information Gain based measures, the candidate for additional cut divides the set of object independently of the main cut.

The algorithm for construction of a decision tree with verifying cuts [1], has been enhanced to use all three measures. Assume that a decision table $\mathbf{A} = (U, A, d)$, a parameter k (in our experiments the value of k was empirically chosen) belonging to natural numbers, template T_p defined by the optimal cut and template T_{p_i} (for $i = 1, \ldots, k$) defined by the verifying cuts are given. Depending on a chosen measure the following criteria are optimized during the v-tree construction:

MaxDiscPair (see [1]) – criterion is maximized.
Entropy based measure – criterion is minimized:

$$EM(p_i) = \begin{cases} 0 & \text{for } \frac{|W|}{|\mathbf{A}|} \geq t_w \\ \big|ES\big(\mathbf{A}(T_p), \mathbf{A}(\neg T_p)\big) - \\ \quad ES\big(\mathbf{A}(T_{p_i}), \mathbf{A}(\neg T_{p_i})\big)\big| & \text{otherwise,} \end{cases}$$

where:

- W is a set of objects that at the same time are not matching patterns T_p and T_{p_i} as also patterns $\neg T_p$ and $\neg T_{p_i}$ (for $i = 1, \ldots, k$),
- t_w is a fixed threshold (t_w was equal 0.1 and 0.05 in our experiments for "microarray" and "normal" data, respectively),
- $ES\big(\mathbf{A}(T_q), \mathbf{A}(\neg T_q)\big) = \frac{\big|\mathbf{A}(T_q)\big|}{\big|\mathbf{A}\big|} * Entropy\big(\mathbf{A}(T_q)\big) + \frac{\big|\mathbf{A}(\neg T_q)\big|}{\big|\mathbf{A}\big|} * Entropy\big(\mathbf{A}(\neg T_q)\big)$ ($q \in \{p, p_1, \ldots, p_k\}$) is the weighted sum of entropies of partitions p and p_i, respectively ($q \in \{p, p_1, \ldots, p_k\}$).

Gini's Index based measure – criterion is minimized:

$$GM(p_i) = \begin{cases} 0 & \text{for } \frac{|W|}{|\mathbf{A}|} \geq t_w \\ \big|GS\big(\mathbf{A}(T_p), \mathbf{A}(\neg T_p)\big) - \\ \quad GS\big(\mathbf{A}(T_{p_i}), \mathbf{A}(\neg T_{p_i})\big)\big| & \text{otherwise} \end{cases},$$

where:

- W is a set of objects that at the same time are not matching patterns T_p and T_{p_i} as also patterns $\neg T_p$ and $\neg T_{p_i}$ (for $i = 1, \ldots, k$),
- t_w is a fixed threshold (t_w was equal 0.1 and 0.05 in our experiments for "microarray" and "normal" data respectively),
- $GS\big(\mathbf{A}(T_q), \mathbf{A}(\neg T_q)\big) = \frac{\big|\mathbf{A}(T_q)\big|}{\big|\mathbf{A}\big|} * Gini\big(\mathbf{A}(T_q)\big) + \frac{\big|\mathbf{A}(\neg T_q)\big|}{\big|\mathbf{A}\big|} * Gini\big(\mathbf{A}(\neg T_q)\big)$
is the weighted sum of ginis of partitions p and p_i, respectively ($q \in \{p, p_1, \ldots, p_k\}$).

The stop condition mentioned in algorithm of v-tree construction is separation of all possible pairs of objects from different decision classes. It is worth pointing out, that the only part of the above algorithm, which would increase the time complexity compared to the classical algorithm from Section II is step 3. This step can be performed in time $O(n \cdot log\ n \cdot m)$, where n is the number of objects and m is the number of attributes.

The determination of the best verification split for the symbolic attributes can be done in time $O(n \cdot l)$, where l is the number of values of symbolic attribute.

Below we present the algorithm for selection of verifying partition for the constructed earlier binary partition p. We assume that the verifying split is determined by a numerical attribute. For ease of discussion, we consider a situation that

there are only two decision classes C_0 and C_1 in the data. This method can be easily generalized to the case of more than two decision classes. The output of this algorithm is the computed collection of cuts that verify partition p.

Algorithm *Selection of verifying cut*
Step 1 Sort the values of the numerical attribute a.
Step 2 Browsing the a attribute values from the smallest to the largest, determine for each appearing cut c the following numbers and store them into a memory (about cuts) M:
$V_L(a, c, C_0), V_L(a, c, C_1)$ - number of objects from decision class C_0 or C_1 with values of attribute a smaller then c,
$L(a, c, C_0, T_p), L(a, c, C_1, T_p)$ - number of objects from decision class C_0 or C_1 with values of attribute a smaller than c and at the same time matching the pattern T_p.
Step 3 Browsing the a attribute values from the highest to the lowest, determine for each appearing cut c the following numbers and place them in a memory (about cuts) M:
$V_H(a, c, C_0), V_H(a, c, C_1)$ - number of objects from decision class C_0 or C_1 with values of a greater then or equal c,
$H(a, c, C_0, \neg T_p), H(a, c, C_1, \neg T_p)$ - number of objects from decision class C_0 or C_1 with values of attribute a greater than or equal to c and at the same time matching the pattern $\neg T_p$.
Step 4 Using information from the memory M, determine the quality of cuts on a in the manner that depends on measure selected for determining the quality of cuts:
MaxDiscPair (see [1])
Entropy:
1. determine the size of set W:
$|W| = |\mathbf{A}| - \big(L(a, c, C_0, T_p) + L(a, c, C_1, T_p) + H(a, c, C_0, \neg T_p) + H(a, c, C_1, \neg T_p)\big)$,
2. discard cuts for which $\frac{|W|}{|\mathbf{A}|} > t_w$,
3. determine the sizes of tables designated by cut c:
$|\mathbf{A}(T_c)| = V_L(a, c, C_0) + V_L(a, c, C_1)$,
$|\mathbf{A}(\neg T_c)| = V_H(a, c, C_0) + V_H(a, c, C_1)$,
4. compute the weighted sum of entropies for partition designated by cut c from $ES\big(\mathbf{A}(T_c), \mathbf{A}(\neg T_c)\big)$,
5. determine the optimum cutting such that the value of $\big|ES\big(\mathbf{A}(T_p), \mathbf{A}(\neg T_p)\big) - ES\big(\mathbf{A}(T_c), \mathbf{A}(\neg T_c)\big)\big|$ is the smallest.
Gini:
1. determine the size of set W:
$|W| = |\mathbf{A}| - \big(L(a, c, C_0, T_p) + L(a, c, C_1, T_p) + H(a, c, C_0, \neg T_p) + H(a, c, C_1, \neg T_p)\big)$,
2. discard cuts for which $\frac{|W|}{|\mathbf{A}|} > t_w$,
3. determine the sizes of tables designated by cut c:
$|\mathbf{A}(T_c)| = V_L(a, c, C_0) + V_L(a, c, C_1)$,
$|\mathbf{A}(\neg T_c)| = V_H(a, c, C_0) + V_H(a, c, C_1)$,
4. compute the weighted sum of Ginis for partition designated by cut c: $GS\big(\mathbf{A}(T_c), \mathbf{A}(\neg T_c)\big)$,
5. determine the optimum cutting such that the value of $\big|GS\big(\mathbf{A}(T_p), \mathbf{A}(\neg T_p)\big) - GS\big(\mathbf{A}(T_c), \mathbf{A}(\neg T_c)\big)\big|$ is smallest.

Assuming that the memory about cuts M is accessible in constant time, the above algorithms runs in time $O(n \cdot log\ n)$, where n is the number of objects (due to the sorting of objects on the basis of the a attribute).

The algorithm for an object classification, using a v-tree with verifying partitions was introduced in [1].

The classifiers constructed with the use of v-decision tree will be called here the *MaxDiscPair-V* classifier, *Entropy-V* classifier or *Gini-V* classifier – depending on used measure during construction, respectively. Note that the algorithm [1] to classify the object in the node utilizes a single tree only when all verifying cuts classify the object just as the main partition p. In other cases, the classification is done by both subtrees. Then the following two cases are considered. The first case refers to the situation when the two subtrees returned the same decision value. Then the value of the node is returned as the decision. The second case refers to a situation where one of the subtrees returned one decision value, and the second subtree the other one. Then that node returns a decision coming from the subtree, which is associated with a greater number of such verifying patterns that classify a test object for this tree. If the numbers of verifying cuts are equal, then the decision comes from subtree selected nondeterministically.

IV. Experiments and results

To verify the effectiveness of classifiers based on our approach, we have implemented classifiers based on the verifying cuts in the programming library CommoDM (Common Data Minnig), which is a continuation of the RSES-lib library (forming the kernel of the RSES system [3]). The experiments have been performed on the data sets obtained from Kent Ridge Biomedical Dataset [7], UCI ML repository (see [11]) and website of The Elements of Statistical Learning book (Statweb)(see [6]). 6 data collections from the first source relates to microarray experiments and they are characterized by a large number of attributes. Our experiments were conducted on the merged original training and testing data sets. The objective of conducted experiments was to test the quality of the classification algorithms discussed in this paper. Table I presents the experimental results received for given data sets and two discretization methods (Entropy and Gini Index based ones) applied to classical tree and v-tree. The counterparts received for discretization method based on maximum number of discernible pairs is presented in [1].

For determining quality of classifiers we applied 10 fold cross-validation technique, which was repeated 10 times for every data set (i.e., 100 cycles of a train-and-test scheme was conducted). The final result of the algorithm is the average of 100 cycles. Popular parameters accuracy (ACC) and coverage (COV) were used to measure the classification success. It is easy to observe that in most cases better results were obtained when the v-tree classifier was applied, both for entropy and Gini's Index based discretization method. That observation is confirmed by the Wilcoxon mached pairs test with 0,05 level of significance in the following cases: (1) [ACC, Entropy-V classifier, num] > [ACC, Entropy-C classifier, num], i.e., the classification quality expressed by ACC coefficient and entropy based discretization method for v-tree is better than for c-tree when applied for data with numerous sets of

TABLE I
THE AVERAGE ACC AND COV WITH STD. DEV. OF EXPERIMENTS FOR C-TREE AND V-TREE AND 2 DISCRETIZATION METHODS

	Entropy-C classifier				*Entropy-V* classifier				*Gini-C* classifier				*Gini-V* classifier			
Method	Acc	Std dev	Cov	Std dev	Acc	Std dev	Cov	Std dev	Acc	Std dev	Cov	Std dev	Acc	Std dev	Cov	Std dev
lymphoma	0.788	0.041	0.945	0.022	**0.836**	0.043	1.0	0.0	0.795	0.042	0.943	0.021	**0.845**	0.047	1.0	0.0
leukemia	0.803	0.037	1.0	0.0	**0.91**	0.023	1.0	0.0	0.819	0.035	1.0	0.0	**0.9**	0.04	1.0	0.0
colon	0.75	0.045	1.0	0.0	**0.756**	0.033	1.0	0.0	**0.766**	0.03	1.0	0.0	0.765	0.045	1.0	0.0
lung	0.925	0.014	1.0	0.0	**0.957**	0.013	1.0	0.0	0.925	0.014	1.0	0.0	**0.956**	0.02	1.0	0.0
prostate	0.837	0.026	1.0	0.0	**0.876**	0.024	0.999	0.002	0.84	0.033	1.0	0.0	**0.847**	0.014	1.0	0.0
ovarian	0.976	0.004	1.0	0.0	**0.981**	0.004	0.999	0.002	0.976	0.004	1.0	0.0	**0.98**	0.006	1.0	0.0
audiology	**0.625**	0.025	0.74	0.017	0.538	0.039	0.996	0.004	**0.66**	0.02	0.827	0.026	0.618	0.021	1.0	0.0
biodeg	0.817	0.009	1.0	0.0	**0.818**	0.009	1.0	0.0	0.809	0.008	1.0	0.0	**0.813**	0.011	1.0	0.0
conn.bench	0.74	0.03	1.0	0.0	**0.752**	0.024	1.0	0.0	0.695	0.02	1.0	0.0	**0.722**	0.024	1.0	0.0
cylinder	0.708	0.015	0.813	0.011	**0.73**	0.013	1.0	0.0	0.703	0.014	0.811	0.014	**0.74**	0.015	1.0	0.0
dermatol.	0.945	0.007	1.0	0.001	**0.954**	0.008	1.0	0.0	0.939	0.005	0.998	0.001	**0.952**	0.006	1.0	0.0
mushroom	**1.0**	0.0	1.0	0.0	0.985	0.0	1.0	0.0	**1.0**	0.0	0.787	0.0	**1.0**	0.0	1.0	0.0
flags	0.629	0.023	1.0	0.0	**0.632**	0.019	0.999	0.002	0.605	0.017	1.0	0.0	**0.609**	0.019	1.0	0.0
ozone	0.953	0.003	0.843	0.004	**0.96**	0.002	1.0	0.0	0.947	0.004	0.822	0.002	**0.96**	0.003	1.0	0.0
parkinsons	0.865	0.016	1.0	0.0	**0.873**	0.029	1.0	0.0	0.86	0.025	1.0	0.0	**0.882**	0.025	1.0	0.0
SAheart	0.626	0.013	1.0	0.0	**0.647**	0.013	1.0	0.001	0.613	0.009	1.0	0.0	**0.652**	0.015	1.0	0.001
segmentat.	**0.953**	0.002	1.0	0.0	0.945	0.002	1.0	0.0	**0.955**	0.003	1.0	0.0	0.942	0.003	1.0	0.0
spam	**0.921**	0.002	1.0	0.0	0.915	0.003	1.0	0.0	**0.913**	0.002	1.0	0.0	0.893	0.003	1.0	0.0

attributes; (2) [ACC, Gini-V classifier, num] > [ACC, Gini-C classifier, num]; (3) [ACC*COV, Entropy-V classifier, num] > [ACC*COV, Entropy-C classifier, num]; (4) [ACC*COV, Gini-V classifier, num] > [ACC*COV, Gini-C classifier, num]; (5) [ACC*COV, Gini-V classifier, non-num] > [ACC*COV, Gini-C classifier, non-num];

We have also checked, separately for c-tree and v-tree classifiers, whether one of the three tested discretization methods leads to better classification quality. We used the Friedman test. It showed that none of the three methods has such property. Both Wilcoxon matched pairs test and Friedman test were used in the form implemented in Statistica program ver. 10.

V. CONCLUSION

In the paper, we presented Entropy based measure and Gini's Index based one applied to determining decision tree with verifying cuts classifier. We checked usefulness of those algorithms on - 18 input data sets. Experiments have confirmed (with statistical significance) that v-tree is relevant classifier for data with a large number of attributes. Used 12 input data with non-numerous set of attributes was too little family of data to express analogous observation when input data do not have really many attributes. Moreover, none of three methods of local discretization proved to be better than remaining ones. The novelty of the paper is important because the experimental results showed that the employment of the knowledge contained in the redundant attributes increases the quality of the classifiers not only for the previously used measure. The conducted experiments have proved the correctness of our assumptions that our method will be also effective for the use of new measures. We expect that the methods may be used in various fields.

ACKNOWLEDGEMENT

This work was partially supported by two following grants of the Polish National Science Centre: DEC-2013/09/B/ST6/01568, DEC-2013/09/B/NZ5/00758, and also by the Centre for Innovation and Transfer of Natural Sciences and Engineering Knowledge of University of Rzeszów, Poland. Andrzej Skowron was also partially supported by the Polish National Science Centre (NCN) grants DEC-2011/01/D/ST6/06981, as well as by the Polish National Centre for Research and Development (NCBiR).

REFERENCES

[1] Bazan, J., G., Bazan-Socha, S., Buregwa-Czuma, Dydo, L., Rzasa, W., Skowron, A.: A classifier based on a decision tree with verifying cuts. Fundamenta Informaticae, vol. 143, no. 1-2, pp. 1-18, 2016

[2] Bazan, J.G., Bazan-Socha, S., Buregwa-Czuma, S., Pardel, P.W., Sokolowska, B.: Predicting the presence of serious coronary artery disease based on 24 hour Holter ECG monitoring. In: M. Ganzha, L. Maciaszek, M. Paprzycki (eds.), *Proceedings of the Federated Conference on Computer Science and Information Systems*, 2012, pp. 279-286, IEEE Xplore - digital library.

[3] Bazan, J. G., Szczuka, M.: The Rough Set Exploration System. Transactions on Rough Sets, III, LNCS 3400, 2005, pp. 37âĂŞ56.

[4] Bazan, J. G., Nguyen, H. S., Nguyen, S. H., Synak, P., Wróblewski, J.: Rough set algorithms in classification problems. In: L. Polkowski, T. Y. Lin, S. Tsumoto (eds.), "Rough Set Methods and Applications: New Developments in Knowledge Discovery in Information Systems," *Studies in Fuzziness and Soft Computing*, Springer-Verlag/Physica-Verlag, vol. 56, 2000, pp. 49–88.

[5] Breiman, L. et. al., Classification and Regression Trees. Wadsworth, Belmont, 1984.

[6] The Elements of Statistical Learning repository, `http://statweb.stanford.edu/ tibs/ElemStatLearn/ datasets/`

[7] Kent Ridge Biomedical Dataset repository, `http://datam.i2r.a-star.edu.sg/datasets/krbd/`

[8] Nguyen, H. S.: Approximate Boolean Reasoning: Foundations and Applications in Data Mining, Transactions on Rough Sets, **V**, LNCS 4100, 2006, pp. 334–506.

[9] Quinlan, J. R.: C4.5: Programs for machine learning, Morgan Kaufmann, San Mateo, California (1993)

[10] Shannon, C.E.: A mathematical theory of communication, Bell System Technical Journal, 27 (1948), pp. 379-423.

[11] UC Irvine Machine Learning Repository, `http://archive.ics.uci.edu/ml/`

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 21–30
DOI: 10.15439/2016F104
ACSIS, Vol. 8. ISSN 2300-5963

A* Heuristic Based on a Hierarchical Space Model Extracted from Game Replays

Bartłomiej Józef Dzieńkowski
Faculty of Computer Science and Management
Wroclaw University of Science and Technology
Wyb. Wyspianskiego 27, 50-370 Wroclaw, Poland
Email: bartlomiej.dzienkowski@pwr.edu.pl

Urszula Markowska-Kaczmar
Faculty of Computer Science and Management
Wroclaw University of Science and Technology
Wyb. Wyspianskiego 27, 50-370 Wroclaw, Poland
Email: urszula.markowska-kaczmar@pwr.edu.pl

***Abstract*—The paper presents a new method of building a hierarchical model of the state space. The model is extracted fully automatically from game replays that store executed plan traces. It is used by a novel approach for estimating the distance between states in a state-space graph. The estimate is applied in the A* algorithm as a heuristic function to reduce the search space. The method was validated using the game *Smart Blocks*. It is a testbed environment for studying methods that benefit from game replay analysis. The proposed heuristic is dedicated to difficult classical planning problems, for which problem-specific or automated heuristics are difficult to obtain.**

I. Introduction

FOR A long time, AI experts have been developing new methods of imitating intelligent behaviors that can be observed in games. Their goal is to give the player the impression that a computer-controlled unit is an intelligent being. Planning plays an important role in such a task because it enables us to solve complex problems automatically. In classical planning, state search methods are applied to find a sequence of actions between an initial and a goal state. Depending on a particular application, the solution should be optimal. However, it must be the best one that can be provided fitting in a limited computation time, because planning problems in games are solved during play. In this work, we propose a new and promising approach for solving difficult problems in the field of classical planning, which is aimed at application in games. The method introduces a novel technique of extracting and storing information from game replays to increase the performance of planning by providing a new heuristic of estimating the distance to the goal.

Supporting the planning process by information extracted from game replays is a promising direction. This is because, for the majority of games, accumulating the recordings is relatively easy. The data is often used for collecting statistical information that models a player. Among many other benefits, this enables us to analyse players' behaviors and predict their actions.

This work focuses on a particular aspect of the game replay analysis that is a reduction of the search space of a state-space search algorithm. It is assumed that the input data contains traces of plans executed by players. The proposed method processes observed plans to build a general model of the state space. Then, the model is employed in the heuristic of the A* algorithm [1]. It is used for estimating the distance to the goal in a state-space graph. The phase of replay processing is separate. It can be done earlier, so the planning process is not slowed down.

In contrast to the popular planners, our method does not require a STRIPS-like representation of a game state [2]. Instead, it operates on an abstract state-space graph, which is representation-independent. This simplification is significant because providing a symbolic model is time-consuming and difficult in many cases. In comparison with other methods that use plan traces, our approach does not require manual annotations [3]. The proposed method is characterized by a high level of automation. It uses a minimum amount of knowledge about a game and its rules.

The approach is original, and its evaluation requires an adequate testbed environment. The environment should have a nontrivial planning problem. There are relatively few research environments accumulating game replays. Usually, stored replays enable us to reproduce a match visually. However, they do not allow us to access full information about the game state. Adding proper replay storing mechanisms to a complex game is a rather large undertaking. Many of the obstacles can be avoided by developing a new game environment that focuses on the research aspects. Thus, we introduce the *Smart Blocks* game [4]. This light-weight environment enables us to store replays in a simple format, investigate a game state easily, and conduct experiments quickly. It was designed for a range of studies related to planning.

To summarize, the goal of the research was to build a method of accelerating the planning process using information retrieved from the plan traces provided by human players. The original contribution of this paper is:

- the novel method of building a hierarchical model of the state space from game replays,
- the heuristic estimate that relies on the hierarchical space model,
- the new testbed environment designed for analysing actions of players solving difficult planning tasks.

The document is divided into eight sections. At the beginning, references to the related works are provided. Next, the *Smart Blocks* environment is characterized. Subsequently, statistics of the collected replays are presented. In the follow-

978-83-60810-90-3/$25.00

ing section, the algorithm of building a hierarchical model of the state space is introduced. Then, application of the model as a heuristic in A* is thoroughly discussed. In the experimental study, the proposed method and a traditional approach are compared. Finally, features of the approach are summarized, and future development directions are indicated.

II. Related Work

This section sets the proposed approach in the planning context. The term of planning is used differently depending on the application domain [5]. In control theory and especially robotics, planning methods are usually applied for motion and trajectory planning [6]. In video games, the topic often refers to pathfinding algorithms [7]. In this paper, we refer to planning as problem solving. It is a process of choosing and organizing actions by anticipating their outcomes. The process relies on basic concepts like states and actions. Plans come from a decision maker, and they are executed by agents.

It is assumed that actions have deterministic effects, and states are fully known. Therefore, we use the taxonomy of classical planning [8]. Non-classical planning refers to partially observable or stochastic environments.

Planning is applied in many games to solve complex tasks. A survey of the approaches currently used in games can be found in [9]. According to the author, STRIPS, Hierarchical Task Networks (HTN), utility systems, and behavior trees are applied in most cases. However, apart from STRIPS-like planners, the discussed methods are mainly used for modelling the reactive behavior of AI-controlled players. In these cases, the provided model is a plan, and its execution is defined by designers. Referring to them as planning methods, which search for a path to a defined goal state, is misleading. Goal-Oriented Action Planning is closer to the discussed understanding of planning [10]. In general, planning gives better perception of AI players' intelligence, but it is more complex to apply in practice. Therefore, our research is dedicated to increasing the applicability of planning in games.

A popular hierarchical approach is Hierarchical Pathfinding A* (HPA*) [11]. HPA* searches for a path on terrain. A state-space graph is represented by a mesh of nodes on the terrain surface. The improved method HPA* clusters places that lay in the geographical neighborhood (inside rectangles). The method also considers graphs with different levels of abstractions as we do in our tree. However, the cost between groups of states cannot be easily determined if the Euclidean distance between state variables does not reflect the transition cost, which is the case addressed in this work. HPA* solves problems in the geographical space while our method is applicable for any abstract state space.

Analysis of executed plan traces is a subject that is discussed in many fields, e.g., robotics [12], business [13], and games [14], [15]. With a few exceptions, there is not much attention given for supporting state search in classical planning by observed plans [16], [17]. More often the observations are applied to plan recognition [18], [19] or player action prediction [20], [21]. In the work of Wang [3], whose method is old but close to our idea of solving a planning problem, plans are learned from observation. However, the described method is inefficient because plan traces must be examined and annotated by experts manually. In addition, it suffers from STRIPS negative preconditions that are generated without limitations. Our method avoids these problems.

Another approach that is somehow related to our problem is presented in the work of Hogg [17]. The method learns hierarchical planning knowledge to solve tasks in HTN. It takes as input a set of planning states and a set of semantically-annotated tasks. Our approach is different because the knowledge model does not require information about tasks and goals.

III. Smart Blocks

Smart Blocks is a testbed environment. It was designed to study planning methods that can learn from player actions. The project includes two subsystems. The first one is a video game in which a player solves planning problems. The game sends observed solutions to the server, where they are stored. The second one is a simulation. It enables us to reconstruct saved plans and test planning algorithms.

The game can be classified as a single-player logic game. Its mechanics was inspired by *Sokoban* [22]. It was implemented with *Unity3D* [23]. One of the priorities was to make the game attractive to the players because the more they play, the more data is collected. The game offers a visual interface, simple gameplay, and easy online access[1] [4].

A. Game Rules

The gameplay relies on simple box-pushing mechanics. In the game, a player controls a team of agents, each of which is represented by a block. The blocks are characterized by different sizes and shapes. The main goal of the team is to reach a golden artifact by any of the blocks. The task is complicated by the maze of triggers and gates that block the path. The triggers usually require different blocks working together to open a gate. The paint of a block is also important when matching a trigger pattern. The pattern comprises a shape together with a colour that must be satisfied by the blocks standing on the trigger. The paint is obtained from a colour portal, and it mixes with the current colour of a block.

A player has to take into account an energy reserve, which is limited and consumed in each action step. This constraint prevents infinite game duration. Unlike the time constraint, it does not enforce a player to act hastily. Energy consumption depends on the size of a block, and it can be increased by entering a ground obstacle. The energy level can be refilled by an energy cell – it disappears once it is used. A player's score depends on the energy reserve at the moment of level completion.

An example of a planning problem that involves agent cooperation is illustrated in detail by Fig. 1. The example shows a part of a stage that contains three user-controlled agents: a ring, a box, and a small cylinder. Their objective

[1] *Smart Blocks* game is available at http://unity3ddev.net/smartblocks/. The source code can be obtained by contacting the authors.

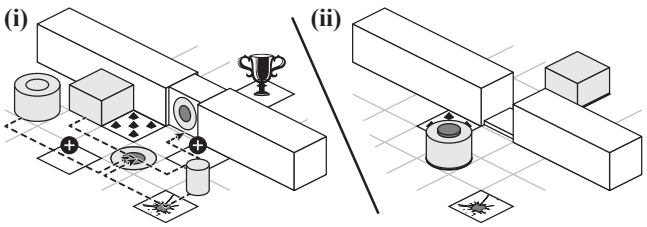


Fig. 1. An example of a planning problem solved in *Smart Blocks*.

is to reach the artifact that is placed behind the wall (the goal is marked with a cup). Initially, the path is blocked by the closed gate **(i)**. It can be opened by using the trigger. First, the box agent approaches the gate avoiding the ground obstacle and collecting one of the energy cells. Next, the small cylinder goes to the trigger through the field where it changes a paint colour to the one that is accepted by the trigger. Then, the big ring collects the last energy cell, and it ends its move in the same place. As soon as the key of the trigger is satisfied, the gate is opened. Now, the box agent is free to reach the goal **(ii)**. If any of the two agents moves from the trigger, the gate will close (unless the third agent is blocking the gate by standing on it).

The presented scheme shows only one of many possible tasks that can be present on a stage. The game levels were designed to be challenging by joining chains of tasks that must be completed before the goal can be reached.

The first version of the game includes ten levels of different difficulty. Subsequent levels can be accessed without solving the previous ones. However, following the order helps to familiarize with the game rules. In the first level, a player learns how to use a simple cooperation to open a gate and reach the artifact. The next level gives an example of multiple blocks interacting with a single trigger. Later, a player is introduced to paint colours and how they mix. In level 4, a player has to use energy cells for the first time. It is a simpler level, but potentially more challenging for a state search algorithm. Level 5 has an alternate route, which is shorter but more challenging to guess. The subsequent level contains many possible routes, and it is difficult to estimate which one will consume the least energy. Level 7 has a long chain of tasks that have to be executed in a certain order. In the next level, misleading trails are present. The last two levels are characterized by a high complexity and a very large state space.

B. Planning Problem

There are several arguments justifying why the planning problem in *Smart Blocks* is nontrivial. The first one is the difficulty of measuring the distance to the goal. For instance, in *Sokoban*, it is possible to count the number of crates on the spots as smaller tasks [22]. It is a good heuristic for estimating the progress of the goal accomplishment. For our problem, the goal progress cannot be measured easily. One of the blocks must reach the location of the artifact. It is unknown which block, what combination of gates and trigger should be used, or how much energy and how many actions it will take. Sometimes it is not necessary to visit locked rooms in order to reach the artifact. However, it may be required to save some energy. The order of blocks in small corridors is also important because the agents can collide with each other. The world can be modified by agents: temporarily by opening a gate or permanently by collecting an energy cell. Each modification applied to the environment builds a subtree in a state search tree (multiplies the search space).

A significant complication is the presence of energy cells. They introduce edges with negative cost to the state-space graph. This type of a state space requires an algorithm that traverses every edge in the graph to ensure optimality (e.g., Bellman-Ford method) [24]. However, the number of states is usually too large to perform a brute-force search. The first three game levels have neither energy cells nor negative cycles in their state spaces.

The problem stated in *Smart Blocks* is the centralized planning of cooperation of agents with limited resources in a mutable environment [25]. It is located in a group of planning problems in which a heuristic estimate is difficult to provide. An abstract version of this problem can be found in a number of real games. Many analogies are present. However, the specificity of the problem in practical application can differ. In some cases, the problem can be solved offline in the phase of game design. Designers can embed a solution schema of a planning problem in a game rigidly. In this research, we are aiming at cases in which the problems can appear dynamically – for instance, problems invented by players during an online game. Therefore, we minimize the amount of predefined knowledge and try to improve planning performance by relying on the observations.

C. Testbed Environment

The simulation is a C# console program. Its function is to parse the recorded plans, execute planning algorithms, and yield the statistics. It contains the model of game rules and a light-weight representation of a game state. Therefore, it enables us to conduct experiments efficiently.

The fundamental motivation for building the environment was a very small number of planning benchmarks that work with observed plan traces. It is a common practice that game replays store only visual effects of player actions. Therefore, they require a lot of reverse engineering to extract actual game states. Our game stores full-information game states in an easy readable format.

Another argument is complexity. In *Smart Blocks*, planning problems are small and flexible. It means that they are easy to understand by players and simple to scale by the level designer. It is easy to follow the execution of a tested method. In environments like *WarCraft (Wargus)* or *StarCraft (BWAPI)* game rules are complicated, and the game state is large [26], [27]. Their state spaces are vast. For most of the tasks, information about optimal plans is unavailable (or practically incomputable). It is difficult to analyse recorded

TABLE I
REPLAY DATABASE STATISTICS

Level No.		*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
Replays		230	169	71	56	57	49	30	13	7	7
Steps	min.	23	18	57	24	40	32	69	60	246	429
	avg.	39.8	34.7	64	27.3	70	58.8	82.7	75	304.6	441.6
	max.	99	88	89	44	108	101	95	125	391	457
Energy	min.	3	3	2	0	3	9	19	30	152	19
	avg.	180.7	121.7	49.8	9.6	70.1	130.3	133.1	124.7	343.7	101.4
	max.	231	161	66	11	138	205	187	182	503	141

plans and evaluate planning algorithms. In our environment, plans can always be compared with optimal solutions or even evaluated visually. It should be noted that *Wargus* and *BWAPI* lack simulation modes (iterative and game-time-independent execution). It is crucial for testing algorithms that traverse a state-space graph.

Finally, our project allows researchers to focus purely on the planning procedure. In another popular environment, *RoboCup*, the mechanics rely on continuous state and physics [28]. It is a good benchmark for robotic applications because it reflects the real world. However, planners usually work with a simplified world model. Providing the model is a challenging task. Our environment abstracts from uncertainty and state discretization issues. These are important problems, but they are located in the conceptual phase of preparing an environment for planning. Our study focuses on traversing the state space, which is discrete by definition.

D. Data

A solution of a stage provided by a player is recorded in a replay file, and then it is submitted to the server. Only complete replays are stored – unfinished or partial solution sequences are discarded. Each recorded and stored solution sequence allows game states to be reproduced fully with no uncertainty. A replay consists of a sequence of steps. Each step is an atomic action committed by an agent. Steps are deterministically ordered – simultaneous actions are not allowed. There are no additional complications behind the described process of collecting data. In *Smart Blocks*, recording the game state is almost as simple as in chess. Actions are animated for the purpose of the presentation, but in fact, they are discrete.

Simple statistics of replays stored in the database are given in Table I. They include for each game level:

- the total number of collected replays,
- a number of action steps to solve a stage,
- an amount of energy at the moment of reaching the final goal.

The statistics provide a good reference for the evaluation of planning algorithms.

It can be seen that the largest number of replays has been collected for the initial levels. It is because the difficulty grows rapidly, and many players resign. The data also gives clues how the levels differ from each other and what is the spread of possible outcomes.

IV. HIERARCHICAL SPACE

To have a better understating of what a hierarchical division of the state space is, we can imagine a city, its district, a residential block placed in there, an apartment in the building, and its room. Searching for a specified room is much easier if we know the name and part of the city, the building address, and the apartment number. The model of our hierarchical space can be perceived as a map that is discovered according to visited places. However, it is not necessary to visit every place to outline a region on the map. The same rule applies to game states and state subspaces.

In formal terms, a hierarchical space is a tree structure. It divides the state space into subspaces that contain groups of states. Subspaces on a higher level are nesting the ones on a lower level. Organizing states as graph nodes inside this kind of structure enables us to traverse and search in the graph more efficiently. For instance, we can reduce the search space by simply skipping whole groups of states that are not leading us to the solution. This procedure can be done on different levels of detail – starting from the most general to the most detailed.

Our approach for building the model relies on *state descriptors*, which is a term introduced in this paper. A state descriptor refers to a selected part or some feature of a state. It can be said that a descriptor partially *describes* a game state. Formally, it is a predicate, and it holds a rule or expression that returns a boolean value depending on whether it is satisfied or not. Descriptors are usually related to intermediate objectives and goals in a game. A descriptor enables us to group a set of states based on a defined criterion. In practice, state descriptors can be added to the implementation easily as boolean functions that accept a state as input. They do not impose formal requirements on problem representation.

In *Smart Blocks*, state descriptors are defined by the designer. They are closely related to the game rules and player's objectives. Therefore, they have a simple and intuitive form. A set of descriptors is generated by descriptor classes – their complete list is presented in Table II. They group states taking into account, for example, the colour of an agent, its position in a room, a gate state, or the goal accomplishment. Each game state can be partially depicted by a subset of descriptors that are satisfied at a given moment. For instance, a state can satisfy a group of three descriptors: {⟨`agent 1 is in room 1`⟩, ⟨`agent 1 is on trigger 3`⟩, ⟨`gate 2 is open`⟩}.

In general, the idea that stands behind state descriptors is to provide a degree of flexibility to the approach. A set of descriptors can be optimized for a problem by machine learning methods. It is the aim of the future study. However, here we use rigidly defined descriptors to focus on the heuristic and provide a proof-of-concept.

Each group of descriptors defines a subspace that covers a part of the state space. The more descriptors work together, the smaller part of the space they cover. Less detailed subspaces nest inside more detailed subspaces. The more general a group of descriptors is, the more children it has. However, it is assumed that state subspaces, covered by descriptors, are not

TABLE II
A LIST OF DESCRIPTOR CLASSES IN SMART BLOCKS

Descriptor Class	Description
agent {id} in room {nr}	informs whether a specified agent is placed in a defined area
agent {id} on colour portal {nr}	true if a specified agent stands on a defined colour portal
agent {id} has {R|G|B} component	checks whether a colour of a specified agent contains one of the base colours
agent {id} on trigger {nr}	true if a specified agent stands on a defined trigger
agents {id}, {id} ... {id} stand together	valid while a specified list of agents stays on the same field
trigger {nr} is valid	informs whether a pattern of a specified trigger is satisfied
gate {nr} is open	true if a specified gate is open
gate {nr} is held	checks whether one or more agents is standing on a specified gate
goal ok	true if one of the agents reached the golden artefact; groups the goal states

required to nest each other fully, but they are allowed to intersect. The method of handling the intersections is provided later in the document.

Theoretical foundations and the algorithm of building a hierarchical model of the state space using state descriptors are presented in the following subsections.

A. Formalization

Below are the theoretical assumptions stated. They are required to discuss optimality of the introduced heuristic.

1) State Space: A classical planning problem can be formulated as finding a path between two arbitrarily specified states in an abstract state space. Let us define a state space as a directed graph $G = \langle S, E \rangle$, where S is a set of nodes and E is a set of edges. Each node $s \in S$ is a system state represented by a tuple of state variables v, Eq. 1:

$$s = \langle v_1, v_2, \ldots \rangle. \tag{1}$$

Each edge $e \in E$ has a transition cost $c \in \mathbb{R}_{>0}$ associated with it. It is assumed that the state space is vast and the cost (or distance) estimate function $\delta : S \times S \to \mathbb{R}_{\geq 0}$ between any two noncontiguous states is unknown or complex. Consequently, state search in the defined space is a nontrivial problem.

2) Plan Observations: A set of executed plans is provided by the system users. An observed plan is a sequence $x_i \in X$ of state transitions, Eq. 2:

$$x_i = \langle s_1, s_2, \ldots, s_n \rangle, \tag{2}$$

where $s_1 \in S$ is an initial state, and $s_n \in S$ is some goal state ($s_1 \neq s_n$). Subsequent state nodes in the sequence are connected by edges.

It is assumed that the observed plans are valid but not cost-optimal. A set X of observations does not cover the entire state space, but it allows us to collect information about its structure.

3) State Descriptor: Let $S_i \subseteq S$ denote a subset of the space states, and $d_i : S \to \{0, 1\}$ is a function that determines membership of a state in this subset, Eq. 3:

$$S_i = \{s \in S : d_i(s)\}. \tag{3}$$

Then, the function $d_i \in D$ is called a state descriptor that classifies and groups states by their selected features. Thus, each state $s_j \in S$ is assigned to a set $D_j \subseteq D$ of descriptors that are valid for s_j, Eq. 4:

$$s_j \Rightarrow D_j = \{d \in D : d(s_j)\}. \tag{4}$$

Based on a set $S_X \subset S$ of states appearing in a set X of plan observations, a set D_X of descriptor sets is extracted, Eq. 5:

$$D_X = \{D_j \subseteq D : s_j \in S_X\}. \tag{5}$$

Descriptor sets enable us to discover a hierarchical structure of the state space, and they play an analogous role as transactions in data-mining.

4) Subspace Tree: State groups separated by descriptor sets are used to build a tree of state subspaces. A state subspace $h_k \in H$ is a pair $h_k = \langle D_k, p_k \rangle$ comprising a set $D_k \subseteq D$ of state descriptors together with an index p_k of a parent subspace. The set D_k determines a set S_k of states that belong to the subspace h_k. The set S_k is the intersection of states grouped by each $d_j \in D_k$, Eq. 6:

$$h_k = \langle D_k, p_k \rangle \Rightarrow S_k = \{s \in S : \left(\forall_{d_j \in D_k} d_j(s)\right) \vee D_k = \emptyset\}. \tag{6}$$

If a subspace h_i is a parent of a subspace h_k, then the set of states covered by the child is a subset of the parent's set, Eq. 7:

$$(p_k = i) \Rightarrow S_k \subset S_i, \tag{7}$$

which is equivalent to Eq. 8:

$$(p_k = i) \Rightarrow D_k \supset D_i. \tag{8}$$

Considering the parent-child relation, it is worth noticing that the implication operator does not have to be applicable in the opposite direction. In other words, a set of states representing a part of the space can be associated with more than one subspace, and thus it has more than one possible parent. For instance, the intersection $S_k = S_i \cap S_j$ of descriptor sets D_i and D_j can be nested by h_i as well as h_j. The selection of a parent subspace is disambiguated algorithmically.

B. Algorithm

The algorithm starts from collecting a set D_X of all possible descriptor groups, which can be found in a set S_X of observed game states. Additionally, the common parts of the groups that intersect are added to D_X. The nesting relations are then stored. Next, a tree structure is assembled by disambiguating the parent-child relations. The result is a tree expressed by a set $H = \{h_1, h_2, \dots\}$ of subspaces, which models a hierarchical structure of the state space.

The basic steps of the algorithm are expressed in pseudocode in Alg. 1. The algorithm accepts a set of game states observed in a replay database as input. A state contains complete information about a temporary situation in a game. Global variables are declared in the context of all methods. In the beginning, unique groups of descriptors are collected (line 2). In the same process, relations between the groups are determined. Next, based on these groups, a hierarchical model of the state space is built (line 8). The sorting in line 7 will be explained later.

Alg. 1: BuildHierarchy ($states$)

```
Data: set of observed states
Result: hierarchical model of state space
1  global descriptorSets ← ∅
2  CollectDescriptorSets ( states )
3  var root ← ⟨ϵ, ϵ⟩
4  global subspaces ← {root}
5  global usedDescSets ← ∅
6  global usedStates ← ∅
7  var descriptorSetsSorted ← Sort ( descriptorSets )
8  foreach descSet ∈ descriptorSetsSorted do
9      BuildSubspaces ( descSet, root )
10 return root
```

At the beginning of the routine, unique groups of descriptors are collected by iterating over every input state (Alg. 2). A descriptor group consists of all the descriptors that are satisfied by a state (line 2). Observed descriptor groups usually overlap (as well as state subspaces). Thus, additional groups are created by intersecting new groups with the already observed ones (line 7), and then stored (line 9). An intersection of two descriptor groups separates a subspace that can be nested by the subspaces of both operands.

Alg. 2: CollectDescriptorSets ($states$)

```
Data: set of observed states
Result: list of descriptor sets
1  foreach s ∈ states do
2      var newDescSet ← DescriptorSetFromState ( s )
3      if InsertDescriptorSet ( newDescSet ) then
4          var intersections ← ∅
5          foreach descSet ∈ descriptorSets do
6              if descSet ≠ newDescSet then
7                  intersections ←
                       intersections ∪ {descSet ∩ newDescSet}
8          foreach descSet ∈ intersections do
9              InsertDescriptorSet ( descSet )
```

It can be concluded that different game states represented by exactly the same groups of descriptors are redundant, and they do not contribute to the model. Therefore, only one state is sufficient for discovering a subspace. The more unique descriptor groups are observed, the richer the structure of the model is.

While the new descriptor groups are added to the list (Alg. 3, line 3), parent-child relations are assigned (line 4). A group of descriptors is a parent of another one if the first one is a subset of the second one. In other words, all descriptors from a parent group can be found in its child. In this relation, a parent will always cover an equal or bigger number of states than its child. The information about subspace nesting will be used in the next step.

Alg. 3: InsertDescriptorSet ($newDescSet$)

```
Data: new descriptor set
Result: stores new descriptor set and assigns parent-child links
1  if newDescSet ∈ descriptorSets then
2      return 0
3  descriptorSets ← descriptorSets ∪ {newDescSet}
4  foreach descSet ∈ descriptorSets do
5      if descSet ⊂ newDescSet then
6          descSet.children ← descSet.children ∪ {newDescSet}
7      else if newDescSet ⊂ descSet then
8          newDescSet.children ←
               newDescSet.children ∪ {descSet}
9  return 1
```

Having prepared such a set of descriptor groups, we can proceed with assembling a data structure that expresses a hierarchy of state subspaces – Alg. 4. Each state subspace is created based on a corresponding descriptor group. State subspaces maintain the same parent-child relation as descriptor groups. At this point, the relation links are copied and disambiguated. A descriptor group may have many possible parents if the group is a product of the intersection operation. However, a subspace in a tree can have only one parent, and it can appear in the structure only once.

In the process of disambiguation, each subspace receives one parent. The process must be performed in a certain order. For instance, if a small subspace is attached to a big one too early, we may lose an opportunity to add an intermediate layer. Therefore, the procedure begins from bigger subspaces that nest the largest number of smaller subspaces. To do so, each time a list of descriptors is sorted descending by the number of unassigned descendants (children, grandchildren, etc.) – line 7 in Alg. 1, and line 6 in Alg. 4.

The recursive procedure in Alg. 4 begins from the root, and it expands the children down to the tree leafs. A subspace can be added to the structure only if it nests (directly or through a descendant) at least one observed state that was not used previously (line 15). Otherwise, the subspace is discarded (line 32). Near the end of the procedure, if there is a subspace that holds a child subspace, and it contains any states at the same time, an additional subspace is created inside the scope of the current subspace to take over these states (line 21).

Alg. 4: BuildSubspaces ($descSet$, $parent$)

```
Data: descriptor set and its parent subspace
Result: subspace tree
1  if descSet ∈ usedDescSets then
2      return 0
3  usedDescSets ← usedDescSets ∪ {descSet}
4  var subspace ← ⟨descSet, parent⟩
5  var anyStateInChild ← 0
6  var descriptorSetsSorted ← Sort ( descSet.children )
7  foreach childDescSet ∈ descriptorSetsSorted do
8      if BuildSubspaces ( childDescSet, subspace ) then
9          anyStateInChild ← 1
10 var anyStateHere ← 0
11 foreach s ∈ descSet.states do
12     if s ∉ usedStates then
13         anyStateHere ← 1
14         break
15 if anyStateInChild ∨ anyStateHere then
16     subspaces ← subspaces ∪ {subspace}
17     parent.children ← parent.children ∪ {subspace}
18     if anyStateHere then
19         var leaf ← subspace
20         if anyStateInChild then
21             leaf ← ⟨descSet, subspace⟩
22             subspaces ← subspaces ∪ {leaf}
23             subspace.children ← subspace.children ∪ {leaf}
24         foreach s ∈ descSet.states do
25             if s ∉ usedStates then
26                 usedStates ← usedStates ∪ {s}
27                 leaf.states ← leaf.states ∪ {s}
28                 s.subspace ← leaf
29                 if IsGoal ( s ) then
30                     leaf.hasGoal ← 1
31     return 1
32 else
33     delete subspace
34     return 0
```

This action ensures that the states are placed only in leaf subspaces. Observed states are assigned to leaf subspaces in line 27. Finally, subspaces that contain goal states are marked (line 29).

The next section explains how the resulting structure can be utilized in planning.

V. Hierarchical Space Based Estimate

In a planning task, the least expensive (shortest) path between an initial state and any state that satisfies the goal is being searched. In our case, the result is a sequence of actions that solves a stage consuming the least amount of energy.

In the considered problem, the best performance can be achieved by applying the A* algorithm [29]. However, it requires a heuristic estimate function to approximate the distance to the goal. Providing such an estimate is not an easy task. In *Smart Blocks*, it is difficult because the progress of solving a stage is hard to measure. The problem structure makes relaxation heuristics futile, which was explained in Section III.B. The goal progress could be calculated using the designer's knowledge about all possible solutions of a level. In this benchmark environment, a solution pattern can be generated, but in real problems, it is an unrealistic assumption.

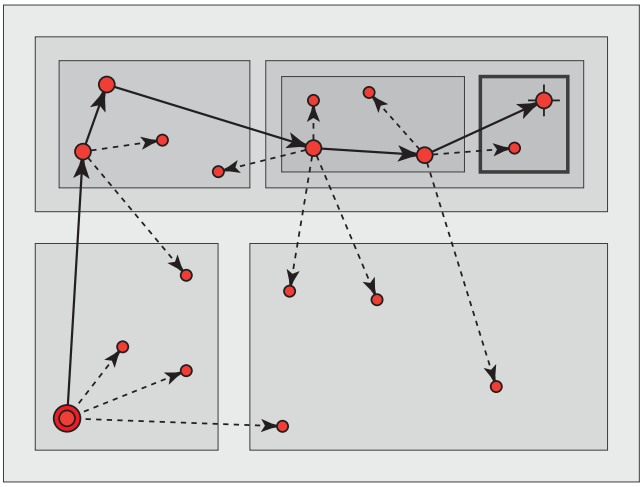


Fig. 2. A visualization of state search supported by the hierarchical model of the state space.

Our idea is to use a hierarchical state-space model to roughly estimate how close a state is to the goal. The distance is calculated based on the number of parent subspaces that a state shares with the nearest goal subspace. The more mutual nodes in a tree they have, the closer to the goal the state is. It is an abstract measure. It supports state search by leading it to more promising regions of the state space.

Fig. 2 provides a visual example of state search that uses information stored in the hierarchical model of the state space. In the picture, the state search begins in the bottom left corner of the figure, and it ends in the top right one. The initial state shares only one parent (the root) with the goal subspace (marked as the bold rectangle). At this point, there are four possible state transitions. One of the subsequent states shares two parents with the goal subspace. This one has a higher priority, and it should be expanded next. The procedure is repeated until the goal subspace is reached.

The formula for calculating the proposed *Hierarchical Space Based Estimate* (HSBE) is defined in Eq. 9:

$$\Delta(s) = 1 - \frac{max_i\Big(\big|parents(s) \cap parents(g_i)\big|\Big)}{d}, \qquad (9)$$

where:

- s is a state,
- g_i is a goal subspace (one of many),
- d is a depth of a hierarchical space tree,
- $parents(\cdot)$ is a function that returns a set of parent subspaces up to the root,
- $max_i(\cdot)$ iterates over all goal subspaces and returns the biggest value.

First, the method finds a parent subspace for a state. It is a leaf subspace with the biggest number of descriptors that match the state. Then, the method collects a set of parent

subspaces of the state (up to the root). The set is intersected with a chain of subspaces from the root to a goal subspace (inclusively). The two sets are intersected to count the number of shared subspaces. The number is divided by the tree depth to normalize the outcome. If there is more than one goal subspace, the most promising one is chosen. The range of $\Delta(\cdot)$ is in $[0, 1)$. The normalized value can be scaled by a minimum transition cost in a state-space graph to ensure that the heuristic never overestimates the cost. Therefore, it is admissible. It should be emphasized that the discussed admissibility is the property of a heuristic that ensures optimality in a state space that does not contain negative cycles, as it was assumed in the formal description.

The main operation of the estimate is tree search, and its computational complexity is logarithmic. A tree structure enables us to quickly collect a chain of nodes between a leaf and the root. In addition, parents of goal subspaces can be stored before the planning phase. Finding a parent subspace for a newly expanded state is linear in the number of leaf subspaces. A significant impact on the computational overhead has the operation of set intersection, which is frequently used. Complexity of this operation depends on the implementation [30]. With some effort, it can be done efficiently.

VI. Experiments

The goal of this study was aimed at checking whether this early concept of a planning method is worth further development. The proposed method is intended for problems in which a heuristic is unavailable, ineffective, or characterized by a high computational complexity [31]. Therefore, Dijkstra's algorithm was chosen as a reference [24]. The algorithm is often used as a benchmark, because it is optimal and represents the worst-case scenario in domain-independent planners [32].

The performance of the methods was measured by:

- the number of iterations in the main loop,
- the number of visited states,
- the maximum size of the state queue (the open set),
- the mean wall-clock execution time.

The results are presented in Table III.

The experiments were conducted on a typical hardware setup: Win 7 x64, Intel i7 @ 3.4 GHz, 8 GB RAM. The wall-clock times were measured for the state search procedure in the main loop. For Dijkstra's algorithm, the measurements were simply averaged from 10 runs. However, HSBE relies on a subspace tree. It can have different sizes, because it can be built from different numbers of plan traces. Thus, the smallest subset of the collected plan traces, which was necessary to obtain full efficiency, was used. Usually, 10% of the set was enough. The subset was chosen randomly, and the procedure was repeated 10 times.

The tests were conducted for the first six game levels. The remaining levels are more complex, their state spaces are larger, and the number of states grows very rapidly. The explosion of states is caused by a bigger number of dynamic objects on the stage. The experiment could not be finished in an acceptable time on the present hardware. However, these levels are somewhat analogous to the preceding ones, and their contribution should not conflict with the initial results.

As expected, the compared methods are equal in the quality of returned solutions. For most of the levels, they provide cost-optimal plans. The exception is level 5 and 6 (compare with the replay statistics in Table I). A slight deviation from the optimal paths is observed. These two levels contain energy cells, which introduce negative-cost transitions to the state-space graph. In this case, solution optimality is not ensured by the algorithms, and the returned solutions may vary depending on the order of states in the queue.

The discussion regarding the performance should begin from comparing the execution times and the number of visited states. In most cases, A*+HSBE is slower than Dijkstra's algorithm. This is caused by the fact that the researched heuristic uses many operations involving complex data structures, while Dijkstra's algorithm is all about adding and removing an item from a queue. On the other hand, the proposed method is characterized by a smaller number of visited states. The execution times include the time of visiting states and the overhead of the method. The more expensive visiting a state is, the lesser part the overhead in the execution time has.

For instance, let us consider level 5. Dijkstra's algorithm is approximately two times faster than A*+HSBE, but it also visits twice as many states. If the cost of visiting a state was tripled, then both algorithms would have almost the same execution time. A*+HSBE is faster if the number of states exceeds this threshold.

The reduction of visited states increases for the larger game levels. The density of the state space division is constant, and it might be too sparse in the simpler ones. If the partition of the state space is low then the heuristic is less informative. On the other hand, if it is too high, then the overhead is considerable.

In the final part of the study, the relation between the number of plan traces, used for building the hierarchical model, and the performance of the proposed method was examined. The experiment was conducted for level 5, and the results are shown in Fig. 3. The subsets of replays in the database were chosen randomly, the process was repeated 10 times, and the measurements were averaged. The quality of the returned solutions remains constant. The reduction of visited states increases as the number of plan traces grows. It appears that a small number of observations is sufficient to discover a large part of the subspace tree. Similar results were observed for the remaining game levels.

Although, the results do not show incontestable superiority of the proposed heuristic over the algorithm used as a benchmark, it should be noted that applying the method in practice may be justified by the reduction of the search space. The profit of employing the introduced approach depends on the computational cost of visiting a state in a particular system. For simple planning problems, the execution time of a complex method can take longer than using a trivial state-search algorithm. *Smart Blocks* as a testbed environment is characterized by a medium-size state space and a very light state so that the experiments could be conducted swiftly.

TABLE III
COMPARISON OF DIJKSTRA AND A*+HSBE (THE LAST TWO COLUMNS CORRESPOND TO BOTH METHODS)

Level No.	Algorithm iterations		Visited states		Max. queue size		Avg. execution time [ms]		Solution length	Solution energy
	Dijkstra	*A**	*Dijkstra*	*A**	*Dijkstra*	*A**	*Dijkstra*	*A**		
1	164	153 **(-6.71%)**	180	163 **(-9.44%)**	13	13 (+0.00%)	0.6	1.3 (+116.67%)	23	231
2	5 271	4 538 **(-13.91%)**	5 800	5 062 **(-12.72%)**	593	593 (+0.00%)	71.3	253.1 (+254.98%)	18	161
3	825 193	747 690 **(-9.39%)**	885 508	806 562 **(-8.92%)**	58 866	58 874 (+0.01%)	12 616.5	13 134.5 (+4.10%)	57	66
4	1 098	440 **(-59.93%)**	1 608	930 **(-42.16%)**	493	493 (+0.00%)	9.3	6.9 **(-25.81%)**	27	11
5	32 609	15 341 **(-52.95%)**	38 025	17 865 **(-53.02%)**	35 45	2 571 **(-27.48%)**	491.7	976.6 (+98.62%)	42	94
6	19 973	12 174 **(-39.05%)**	29 549	19 539 **(-33.88%)**	7 423	7 367 **(-0.75%)**	366.2	993.1 (+171.19%)	53	190

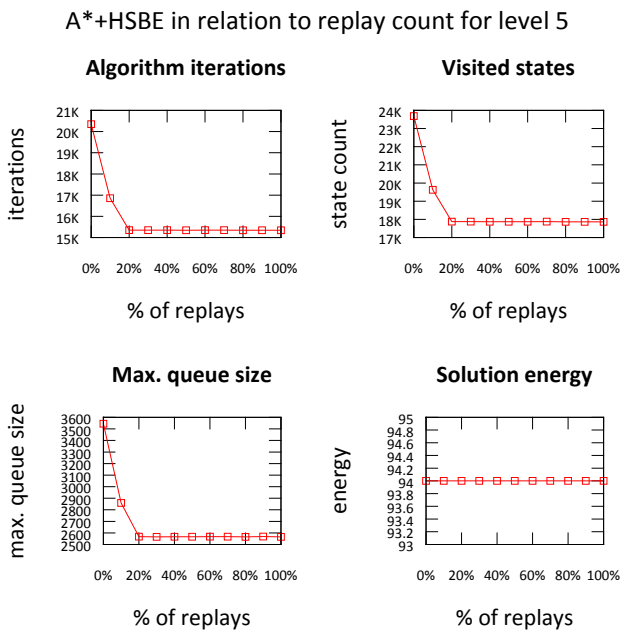


Fig. 3. The set of charts shows the statistics of A*+HSBE for level 5 depending on different numbers of replays used for building a subspace tree.

However, for real planning problems, if the state space is vast and visiting a state is expensive then the method is most certainly worth applying.

VII. CONCLUSIONS

The original contribution of this paper is a new method of building the hierarchical model of the state space from game replays. Information stored in the model represents general knowledge about the structure of the state space. The method accepts executed plan traces as input. The observations are not required to be annotated by experts. The model abstracts from state representation methods. It relies on state descriptors that refer to selected features of a state, and their form is unrestricted. It does not oblige system designers to rewrite a planning problem in PDDL. Instead, they can use the original implementation of a game. It is more convenient and efficient.

The method is used for grouping states hierarchically, and it relies on hyperspace nesting. The proposed heuristic estimates the distance between any two states in the state space to effectively guide the state search towards the goal and reduce the search space. It is intended for difficult cases in which a heuristic estimate of distance in a state-space graph cannot be easily provided and domain-independent heuristics are inefficient.

The proof-of-concept was validated using *Smart Blocks*. It is a testbed environment designed for conducting experiments that involve game replay analysis. The game is modeled as a variant of a multi-agent system. It enables a researcher to focus on classical planning problems. Unlike commercial games, it comes with tools that simplify method testing.

It should be emphasized that the approach is general. Formal assumptions of the method, which include the planning problem definition and the model of the hierarchical state space, are abstract. The same rule applies to the introduced algorithms, because they do not assume any specific properties of the system in which a planning problem is solved. Although the method relies on state descriptors that were implemented according to domain-specific knowledge for this particular study, their formal form is abstract, and it is possible to extract them automatically (see the Future Work section). To summarize, the approach can be applied to any classical planning problem for which input observations are available.

VIII. FUTURE WORK

The presented results mark an important milestone in the development of the approach. There are still many opportunities for improvement, but their examination is a rather large undertaking. This section outlines possible development directions.

In the course of ongoing research, it has been discovered that a concept (Galois) lattice is a more suitable model for expressing relations between state subspaces [33]. In a hierarchical structure described by the lattice, the intersection of state subspaces has many parents, rather than a single one like in the tree structure. Formal foundations of the Formal Concept Analysis (FCA) are consistent with the descriptors introduced in the discussed model of the state space. Thus, a lattice can be built by an algorithm commonly used in FCA [34]. Preliminary study results show a good performance of a heuristic supported by this model. Therefore, a tree will be replaced by a lattice in the subsequent study.

At the current stage of development, issues related to processing a very large replay database have not been taken into account. It is a secondary problem because the process is separated from planning, and its execution time is acceptable. Nonetheless, pruning methods that protect against the overgrowth of the structure should be researched.

Reduction of the search space depends on the quality of the state space division. The algorithm for building a hierarchical space model ensures that the formal assumptions are satisfied. However, the division quality is affected by state grouping, which is handled by state descriptors. Adapting state descriptors automatically can lead to better results. The research is aimed at inventing an adaptive descriptor model that can be tuned by machine learning methods.

REFERENCES

[1] R. Dechter and J. Pearl, "Generalized best-first search strategies and the optimality of A*," *J. ACM*, vol. 32, no. 3, pp. 505–536, 1985. doi: 10.1145/3828.3830

[2] R. E. Fikes and N. J. Nilsson, "STRIPS: A new approach to the application of theorem proving to problem solving," in *Proceedings of the 2Nd International Joint Conference on Artificial Intelligence*. Morgan Kaufmann Publishers Inc., 1971, pp. 608–620.

[3] X. Wang, "Learning by observation and practice: An incremental approach for planning operator acquisition," in *Proceedings of the 12th International Conference on Machine Learning*. Morgan Kaufmann, 1995. doi: 10.1.1.36.7719 pp. 549–557.

[4] "Smart Blocks," http://unity3ddev.net/smartblocks, accessed: 2015-11-11.

[5] S. M. LaValle, *Planning Algorithms*. Cambridge University Press, 2006. ISBN 0521862051

[6] D. Nau, M. Ghallab, and P. Traverso, *Automated Planning: Theory & Practice*. Morgan Kaufmann Publishers Inc., 2004. ISBN 1558608567

[7] D. M. Bourg and G. Seemann, *AI for game developers*. O'Reilly & Associates Inc., 2004. ISBN 0-596-00555-5

[8] D. Bryce and S. Kambhampati, "A tutorial on planning graph based reachability heuristics," *AI Magazine*, vol. 28, no. 1, pp. 47–83, 2007.

[9] A. J. Champandard, "Planning in games: An overview and lessons learned," http://aigamedev.com/open/review/planning-in-games/, 2013, accessed: 2015-11-11.

[10] J. Orkin, "Applying goal oriented action planning in games," in *AI Game Programming Wisdom 2*. Charles River Media, 2002, pp. 217–229. [Online]. Available: http://web.media.mit.edu/~jorkin/GOAP_draft_AIWisdom2_2003.pdf

[11] A. Botea, M. Muller, and J. Schaeffer, "Near optimal hierarchical path-finding," *Journal of Game Development*, vol. 1, pp. 7–28, 2004. doi: 10.1.1.112.314

[12] L. Mosenlechner, N. Demmel, and M. Beetz, "Becoming action-aware through reasoning about logged plan execution traces," in *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*, 2010. doi: 10.1109/IROS.2010.5650989. ISSN 2153-0858 pp. 2231–2236.

[13] I. Hulpuş, M. Fradinho, and C. Hayes, "On-the-Fly Adaptive Planning for Game-Based Learning," in *Case-Based Reasoning. Research and Development*, ser. LNCS, I. Bichindaritz and S. Montani, Eds. Springer, 2010, vol. 6176, pp. 375–389.

[14] S. Sahasrabudhe and H. Munoz-Avila, "Mining cause-effect sequential patterns from action traces," FLAIRS Conference, 2013. [Online]. Available: http://www.cse.lehigh.edu/~munoz/Publications/flairs04.pdf

[15] S. Breining, H. Kriegel, M. Schubert, and A. Züfle, "Action sequence mining," Workshop on Machine Learning and Data Mining in Games, 2013. [Online]. Available: http://www-kd.iai.uni-bonn.de/dmlg11/program.html

[16] N. Nejati, P. Langley, and T. Konik, "Learning hierarchical task networks by observation," in *Proceedings of the 23rd International Conference on Machine Learning*. ACM, 2006. doi: 10.1145/1143844.1143928. ISBN 1-59593-383-2 pp. 665–672.

[17] C. Hogg, H. Munoz-Avila, and U. Kuter, "Learning hierarchical task models from input traces," *Computational Intelligence*, vol. 32, no. 1, pp. 3–48, 2016. doi: 10.1111/coin.12044

[18] H. Xu, B. T. R. Savarimuthu, A. Ghose, E. D. Morrison, Q. Cao, and Y. Shi, "Automatic BDI plan recognition from process execution logs and effect logs." in *Engineering Multi-Agent Systems*, ser. LNCS, M. Cossentino, A. E. Fallah-Seghrouchni, and M. Winikoff, Eds., vol. 8245. Springer, 2013. doi: 10.1007/978-3-642-45343-4_15. ISBN 978-3-642-45342-7 pp. 274–291.

[19] G. Sukthankar and K. Sycara, "Robust and efficient plan recognition for dynamic multi-agent teams," in *Proceedings of the 7th International Joint Conference on Autonomous Agents and Multiagent Systems*, ser. AAMAS, vol. 3, 2008. doi: 10.1.1.149.7677. ISBN 978-0-9817381-2-3 pp. 1383–1388.

[20] J. Hsieh, C. Sun, C. Wang, and C. Cheng, "Mining replays of real-time strategy games to learn player strategies," Canadian Artificial Intelligence Conference, 2008. [Online]. Available: http://gamelearninglab.nctu.edu.tw/ctsun/RTS%20player%20modeling.pdf

[21] B. Weber and M. Mateas, "A data mining approach to strategy prediction," IEEE Symposium on Computational Intelligence and Games, 2009. [Online]. Available: http://alumni.soe.ucsc.edu/~bweber/pubs/cig_2009.pdf

[22] A. Botea, M. Muller, and J. Schaeffer, "Using abstraction for planning in Sokoban," in *Proceedings of the 3rd International Conference on Computers and Games*. Springer, 2003. doi: 10.1.1.70.8121 pp. 360–375.

[23] "Unity3D," http://unity3d.com, accessed: 2015-11-11.

[24] T. H. Cormen, C. Stein, R. L. Rivest, and C. E. Leiserson, *Introduction to Algorithms*, 2nd ed. MIT Press and McGraw-Hill, 2001. ISBN 0070131511

[25] M. Woolridge and M. J. Wooldridge, *Introduction to Multiagent Systems*. John Wiley & Sons, Inc., 2001. ISBN 978-0470519462

[26] "Wargus," http://wargus.sourceforge.net, accessed: 2015-11-11.

[27] "BWAPI," http://bwapi.github.io, accessed: 2015-11-11.

[28] "RoboCup," http://www.robocup.org, accessed: 2015-11-11.

[29] S. J. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, 2nd ed. Pearson Education, 2003. ISBN 0137903952

[30] R. Baeza-Yates, "A fast set intersection algorithm for sorted sequences," in *Combinatorial Pattern Matching*, ser. LNCS, S. Sahinalp, S. Muthukrishnan, and U. Dogrusoz, Eds. Springer, 2004, vol. 3109, pp. 400–408. ISBN 978-3-540-22341-2

[31] J. Seipp, F. Pommerening, G. Roger, and M. Helmert, "Correlation complexity of classical planning domains," in *Proceedings of the 8th Workshop on Heuristics and Search for Domain-independent Planning (HSDIP)*, 2016, pp. 12–20. [Online]. Available: http://icaps16.icaps-conference.org/proceedings/hsdip16.pdf

[32] R. Liang, "A survey of heuristics for domain-independent planning," *Journal of Software*, vol. 7, pp. 2099–2106, 2012. doi: 10.4304/jsw.7.9.2099-2106

[33] R. Wille, *Formal Concept Analysis: Foundations and Applications*. Springer, 2005, ch. Formal Concept Analysis as Mathematical Theory of Concepts and Concept Hierarchies, pp. 1–33. ISBN 978-3-540-31881-1

[34] S. O. Kuznetsov and S. A. Obiedkov, *Algorithms for the Construction of Concept Lattices and Their Diagram Graphs*. Springer, 2001, pp. 289–300. ISBN 978-3-540-44794-8

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 31–40
DOI: 10.15439/2016F53
ACSIS, Vol. 8. ISSN 2300-5963

Employing Game Theory and Computational Intelligence to Find the Optimal Strategy of an Autonomous Underwater Vehicle against a Submarine

Bartłomiej Józef Dzieńkowski
Faculty of Computer Science and Management,
Wroclaw University of Science and Technology
Wyb. Wyspianskiego 27,
50-370 Wroclaw, Poland
Email: bartlomiej.dzienkowski@pwr.edu.pl

Christopher Strode
Centre for Maritime Research and Experimentation (CMRE)
La Spezia, Italy
Email: strode@cmre.nato.int

Urszula Markowska-Kaczmar
Faculty of Computer Science and Management,
Wroclaw University of Science and Technology
Wyb. Wyspianskiego 27,
50-370 Wroclaw, Poland
Email: urszula.markowska-kaczmar@pwr.edu.pl

***Abstract*—Game theory is a tool that may be used to model a player as an intelligent being – one who seeks to optimize his own performance while taking into account the performance of his opponent. However, it is often challenging to apply the theory in practice. In the naval environment, this approach may be used, for instance, to find the best strategy for an Autonomous Underwater Vehicle (AUV) while considering the intelligence of the submarine opponent. Classic approaches based on Minimax suffer from an explosion of states, and they are difficult to use in real-time. The paper introduces an approach that improves the Minimax algorithm in a complex naval environment. It assumes limited and scalable computational resources. The approach takes advantage of a flexible utility function based on a neural network with parameters tuned by a genetic algorithm.**

I. Introduction

AN AUTONOMOUS Underwater Vehicle (AUV) is a robot that travels underwater (Fig. 1)[1]. Compared to other Unmanned Underwater Vehicles (UUVs) such as Remote Operating Vehicles (ROVs), it is not guided by an operator. AUVs are mostly employed in the field of oceanography and are beginning to be considered for military use [2]. For instance, they can be used for patrolling and monitoring the vicinity of a naval port, or for supporting a surface platform in its search for a submarine. An AUV has an advantage over an ROV because it does not reveal its location by continuous communication with an operator.

A submarine is a very difficult opponent to detect because of its ability to exploit the complicated nature of underwater sound propagation. As a result, it can stay hidden for a long time owing to its superior endurance and speed. By comparison, an AUV has limited mobility, sonar capabilities, computational power, and battery life. More importantly, a submarine is commanded by qualified personnel making it a deliberate and intelligent competitor. Therefore, the robot controller must meet high requirements. AUV's software should provide as much intelligent behavior as possible to mitigate its limited resources.

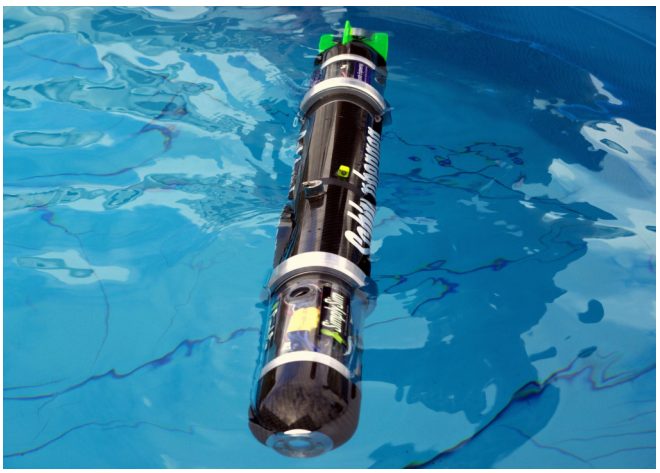


Fig. 1. The picture shows an example of AUV used in military [3]. The Blackghost AUV can attack with no outside control.

In order to maximize the ability of the AUV to detect the submarine, we consider here a bistatic sonar employment. This means that the sonar source and receiver are separated – rather than collocated on a single platform in the more traditional monostatic case. In this study, the sonar source

The project was commissioned and financed by the Visiting Researcher Programme organized by the Centre for Maritime Research and Experimentation (CMRE) located in La Spezia (Italy) – the Centre is an executive body of NATO's Science and Technology Organization (STO) along with the NATO Collaboration Support Office. The research was partially supported by the statutory funds of the Department of Computational Intelligence, Faculty of Computer Science and Management, Wroclaw University of Science and Technology.

is in a fixed location while the receiver is assumed to be a linear array of hydrophones towed by the AUV. Importantly this means that the AUV can receive bistatic sonar contacts – with range and bearing information – without needing to transmit. Sonar transmissions are easily counter detected by a submarine, thereby revealing the location of the asset and allowing the submarine the opportunity to evade.

To achieve an efficient strategy model for an AUV, it must be validated against a challenging foe. Thus, to find the best strategies for both naval units we use game theory [4]. The problem is described as a naval version of a pursuit-evasion game. The players hold different properties; therefore, a skirmish is asymmetrical. It is a multistage non-zero-sum game placed in a complex environment with uncertainty and incomplete information about the game state. The **non**-zero-sum assumption is introduced because the players may have specific and unique objectives. In addition, having a limited access to information about an opponent's state, they might unconsciously cooperate in some cases.

Finding the game equilibrium, which is the problem solution, is a difficult task. In theory, the game should be defined as an extensive-form game. However, adding equivalent state nodes that denote the space of hidden possibilities would greatly increase the problem complexity by extending the game tree, which is already very large. In other words, it is nearly impossible to traverse the entire game tree (extensive or not) in practice. The approach considered here is to employ a pure strategy form of a sequential game with discrete time.

Another complication is that each player receives input that is not readily convertible to a utility value, which expresses how desirable a given state is for a player. There is a significant distance in the state space between actions in the past and their real effect in the future.

In our work, we solve the problem by using a flexible and trained utility function model that is optimized according to specified criteria. A utility function then converts player's input into a utility value of the game state. Both players use the Minimax decision rule to choose the best action. They have their own utility function model that is tuned to gain the best outcome assuming that an opponent is doing the same. In the presented approach, a neural network was chosen as the utility modeling function [5]. Its weights are trained by a genetic algorithm to maximize each player's fitness [6]. The fitness is calculated taking into account the players' objectives.

In the following section, a general overview of the method framework is introduced. Next, a short survey of the existing works related to the stated problem is provided. Subsequently, important properties of the naval environment are described. The following part of the document provides formal foundations and describes the problem as a pursuit-evasion game. Next, naval players are characterized, and their cost functions are defined. A utility function model and its training method are presented in the next part of the document. Finally, results of the experimental study are collated and described.

II. Method Overview

This section provides a general overview on the proposed approach. It briefly summarizes the method framework and its components (Fig. 2).

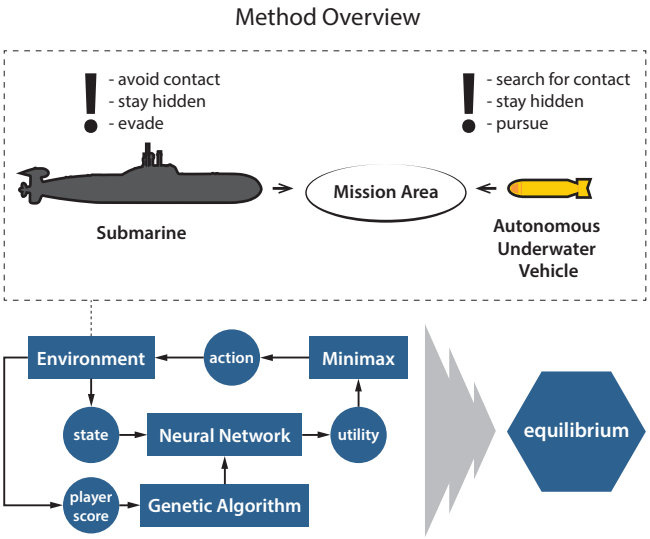


Fig. 2. The picture visualizes the game and the method of solving it.

The proposed method searches for the optimal strategies of players operating in a simulated environment. The problem description uses game-theoretic formalism to model the naval environment as a game. The solution is a saddle-point equilibrium. It is a state in which no player can gain by changing his strategy. In accordance with theoretical foundations, it is assumed that players are rational. The quality of a strategy is represented by its cost value. It is an accumulative cost calculated for a sequence of states from an initial state to a final one.

For the adopted game model, an optimal strategy is expressed by the Minimax decision rule. Application of the algorithm is not straightforward, because a utility value for a given game state is unknown. The utility is a system-wide feedback to the Minimax algorithm. In this approach, a utility value is returned by the output of a multilayer neural network, which is employed as a utility function. Its input is fed with a game state perceived by a player. Each player has a separate neural network, because the players have different capabilities, and they do not share the same view of the environment and its state.

Because the desired (optimal) output of a neural network is unknown, weights inside the network are tuned using an evolutionary approach, rather than the backpropagation learning algorithm. In this method, a genetic algorithm optimizes the cost of the Minimax strategy, which is guided by the output of the neural network. The evaluation procedure requires the game to be simulated over a number of steps. During the optimization process both players improve their strategies until

they reach the equilibrium. It is a theoretical state in which the players have found their optimal strategies.

III. State of the Art

The problem stated in this work is fresh and specific. In the literature, there is a range of works directly addressed to this field but not this particular problem. Most of the papers discuss fundamental issues related to the naval environment [7]. One of the most elementary is modeling of underwater signal propagation and communication [8], [9]. Another is a classic flaming datum problem, in which a fleeing submarine is relocated after momentarily revealing its position [10].

Modern navies are beginning to procure and test multistatic sonar systems at sea. Locations of the sensors are optimized based on a game-theoretic approach, and their efficiency is evaluated in the previous works [11], [12]. Recent technological progress enables us to employ an AUV as a mobile signal receiver instead of using a fixed-position sensor [13]. Despite the fact that an AUV can be considered as a more effective tool against a submarine, the topic is rarely undertaken so far. Related works are focused on the problem of building a probability density map of possible locations of an opponent [14]. Many of the approaches employ simplified behavioral models [15], rather than considering the best strategies for both players and optimizing towards the equilibrium.

Solutions designed for similar problems are often bounded to their domains and cannot be directly applied to this specific case [16]. They are intended for less demanding environments. In the air, aerial vehicles can easily communicate with the command and exchange information. They rely on radar and visual information, which can be processed more efficiently than sonar input. Underwater communication is slow, has a very limited bandwidth, and above all, using it immediately exposes the location of a vehicle to an opponent. Underwater vehicles exploit properties of underwater signal propagation to remain stealthy. They must plan their route very carefully while drones can maneuver more freely, because their environment is mostly uniform. A very limited computational power makes it impossible to run an expensive optimization process on-board while the requirements regarding the method are still high.

The discussed problem is focused on tracking and spying an intelligent enemy unit, rather than patrolling or engaging in combat according to a clearly defined protocol. The task is specific, and it consists of a range of problems that are not addressed by the works related to autonomous vehicles.

The proposed method is based on a known idea of employing an evolutionary algorithm for finding an optimal strategy in game theory [17], [18]. In these works, a population of agents is optimized to achieve the best performance against an opponent who is following a defined strategy. However, that approach cannot be employed in this case, because the behavior of an opponent cannot be easily described by a closed set of rules. Here, both competing sides are optimized simultaneously. To avoid executing an expensive optimization process on-board, the optimization is aimed at tuning a neural network that is used for evaluating the game state and computing utility (reward). The neural network plays the role of a utility function in an on-board decision process. In the light of the above facts, the problem setup and solution are considered as original.

IV. Naval Environment

The underwater environment in which our scenario takes place is both harsh, from a technological standpoint, and complex in terms of the physics that govern the propagation of sound used to detect a submarine [8]. The performance of a sonar system is a complex function of transmission loss, reverberation levels and noise levels, all affected by various oceanographic, surface, and bottom parameters. Not least of which is the sound speed profile within the water column which causes a bending of the sound propagation paths and can result in large regions of water from which sound may be diverted. These shadow zones may then be exploited by an intelligent submarine greatly decreasing its chance of being detected.

The eventual signal-to-noise ratio (SNR), the primary metric used to assess the probability of detecting a submarine by sonar, depends on water temperature and salinity, surface roughness, depth, and shape of the sea bottom. This study considers the use of multistatic sonar whereby performance is a function of the specific geometry between the separated sonar source and receiver (in this case the AUV) together with the submarine location. As a result, detailed (and time consuming) acoustic propagation models are required to accurately predict multistatic sonar performance within range-dependent environments.

Irrespective of the complexities of actually detecting a submarine, the simple operation of an AUV in the ocean also has its challenges. In addition to basic underwater physics, the AUV has limitations imposed on its motion trajectory due to the stability of its receiving array of hydrophones [19]. The complicated propagation effects already discussed also serve to limit the ability of the AUV to send and receive messages through the water. This is in fact a driving force behind the need for better autonomous decision making – since the vehicle cannot rely on regular intervention by operators.

The goal of this study is to provide a proof-of-concept and consequently, some simplifications are introduced to avoid vast computations, which significantly improves the computation time of the experiments. The system is modeled in such a manner that a general characteristic of the underwater environment is captured. Earlier experiments have shown that the acoustic model is a bottleneck for the calculations. It was replaced by a set of rules that cover only the general features.

The operational area is rectangular, and it contains the players' start positions, signal source position, and mission area, which is circular (Fig. 3). It is assumed that the mission area holds an objective that is important to a submarine. The objective is accomplished if the unit closes to within a defined radius of the goal. In the meantime, the AUV's objective is to keep as close as possible to the submarine. While the AUV is

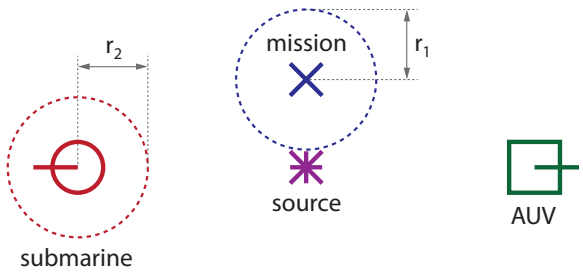


Fig. 3. The image presents an example of game arrangement. The objective of a submarine is to travel to the mission area. The AUV's objective is to be in close surroundings of the opponent. The nearer a unit approaches the signal source, the easier it can be detected.

not intended to prosecute the submarine, thereby preventing it reaching its mission goal, it should nevertheless detect its presence and alert the defenders as soon as possible. Both players benefit from staying undetected. However, the players have different priorities, so the zero-sum property is not valid in this case.

V. Game Theory

The pursuit-evasion game (PEG) is a well-known problem in the class of differential games in game theory [20]. It is often referred to as a simple lion-zebra or cop-robber case.

Let us define a naval version of PEG as a state-feedback discrete-time dynamic game with two non-cooperating players, where P_1 is a pursuer (AUV) and P_2 is an evader (a submarine) executing actions sequentially in each game stage k, Eq. 1:

$$a(k) = \begin{cases} a_k^1 & \text{if } k \text{ is odd} \\ a_k^2 & \text{if } k \text{ is even} \end{cases}, \tag{1}$$

where $a(k)$ is an action at stage k, a_k^1 is P_1's action, and a_k^2 is P_2's action. The game corresponds to dynamics of the form (Eq. 2):

$$s_{k+1} = \Delta_k\Big(s_k, a(k)\Big), \ \forall k \in \{1, 2, \ldots, K\}, \tag{2}$$

where:

- s_k is an entry state node at stage k,
- Δ_k is a transition function modeling dynamics at stage k,
- K is a finite time horizon.

A finite horizon stage additive cost is (Eq. 3):

$$\sum_{k=1}^{K} c_k\Big(s_k, a(k)\Big), \tag{3}$$

that P_1 wants it to minimize, and P_2 wants it to maximize. A state-feedback information structure corresponds to policies of the form (Eq. 4):

$$a_k^1 = \gamma_k(s_k), \quad a_k^2 = \sigma_k(s_k), \tag{4}$$

where γ and σ are state-feedback policies for P_1 and P_2, respectively. The corresponding value of the cost in Eq. 3 for the policies is denoted by $C(\gamma, \sigma)$. A saddle-point pair of equilibrium policies (γ^*, σ^*) satisfies (Eq. 5):

$$C(\gamma^*, \sigma) \leq C(\gamma^*, \sigma^*) \leq C(\gamma, \sigma^*), \ \forall \sigma, \gamma. \tag{5}$$

For games with a finite state space, an optimal policy cost can be found using Minimax Theorem. It is solved algorithmically. However, the players perceive the state differently, and they do not have strictly opposite objectives. Therefore, the game is not considered as zero-sum. For this reason, Alpha-Beta pruning, which would reduce the complexity, cannot be applied [21].

VI. Naval Players

Each player has a set of discrete move actions to choose in his turn. For the sake of simplicity, a move action can be executed with a finite number of speeds and headings (Eq. 6):

$$a(k) \in \{\vec{m}_0, \vec{m}_1, \ldots \vec{m}_n\}, \ \vec{m} = s * \vec{h}, \tag{6}$$

where $\vec{m}$ is a move vector as a product of scalar speed s and unit heading vector $\vec{h}$. In order to avoid inaccuracy and provide more flexibility, player coordinates are expressed by floating point values rather than grid cells. Units cannot currently change depth.

To imitate sonar features, both detection and counter detection ranges were introduced. A naval player does not have access to information about his opponent as long as the distance between them is bigger than the opponent's detection range. The detection range is not constant, and it changes depending on the distance to the signal source. The closer to the source, the more acoustic energy is reflected by a naval unit and so it is easier to detect. The signal source can be received from a far distance and, therefore, its position is always known to all players.

At some point, the algorithm must estimate the opponent's actions to calculate utilities of the future states. Unfortunately, it cannot be easily done if a player does not know the exact position and heading of the opponent. To deal with the problem an approximate map of possible locations of the opposing unit is generated based on its operational range (Fig. 4). The operational range is a circular area placed in the center of the last revealed position with a radius equal to the maximum distance the unit could travel since the last contact time. The map of possible states of an opponent is limited by the complexity of the algorithm.

VII. Cost

A cost value describes how well a player performed during the whole game. AUV is referred to as a pursuer whose behavior is characterized by minimizing the distance to its opponent. The pursuer minimizes cost for staying within a specified range to an evader. The robot is penalized if

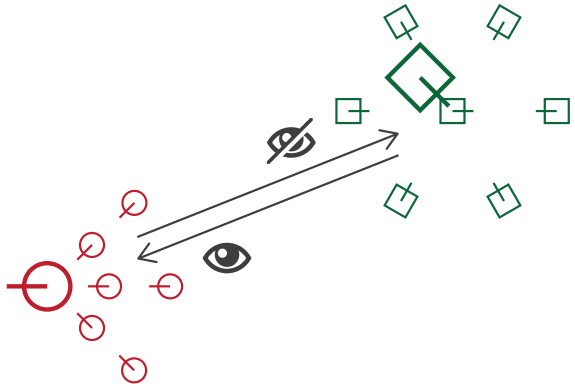


Fig. 4. The picture shows how the algorithm handles limited access to information about the current state of the opponent. The big circle on the left is the actual location of the submarine. The rectangle on the opposite side is the AUV. Small shapes represent their possible positions in the next game step. The state of the player on the left is exposed while the opposite player stays unrevealed, but his last position is known. A map of several possible positions inside the operational range of the hidden player is generated.

its position is revealed. Formally, the cost for the AUV is calculated by the following formula (Eq. 7):

$$C_1(k) = C_1(k-1) + d_o + \\ - \begin{cases} b_1 & \text{if } d_o \leq r_1 \\ 0 & \text{otherwise} \end{cases} + \begin{cases} p_1 & \text{if } v_1 = 1 \\ 0 & \text{otherwise} \end{cases}, \quad (7)$$

where:

- k is a game stage number,
- $C_1(\cdot)$ is the pursuer's (AUV) cost function,
- d_o is a distance to the opponent,
- b_1 is an award given to the pursuer if distance d_o is smaller than threshold r_1,
- p_1 is a penalty if the pursuer reveals his position to the opponent,
- v_1 is the pursuer's state of visibility to the opponent.

By analogy, the submarine is referred to as an evader who maximizes the distance to his opponent and moves towards the mission area. If the unit is sufficiently close to the mission, a positive cost is awarded. Whenever a player is exposed to his opponent, a penalty cost is applied. In order to clarify, the cost for the submarine is calculated by the following formula Eq. 8:

$$C_2(k) = C_2(k-1) + d_o - d_m + \\ + \begin{cases} b_2 & \text{if } d_m \leq r_2 \\ 0 & \text{otherwise} \end{cases} - \begin{cases} p_2 & \text{if } v_2 = 1 \\ 0 & \text{otherwise} \end{cases}, \quad (8)$$

where:

- $C_2(\cdot)$ is the evader's (submarine) cost function,
- d_m is a distance to the mission center,
- b_2 is an award given to the evader if distance d_m is smaller than threshold r_2,
- p_2 is a penalty if the evader reveals his position to the opponent,
- v_2 is the evader's state of visibility to the opponent.

Cost parameters b, r, and p affect the behavior of players by defining objectives and their relative importance. Distance effects included in the cost functions serve to steer the training towards the desired outcome. Ultimately, award and penalty values drive the tactics of a player, and they can be adjusted to observe different behaviors.

VIII. Utility

The utility value of the game state is calculated by a neural network based on information held by a player. Among many benefits of neural networks is an application for non-linearly separable problems and generalization of acquired information [22]. In this study, Multi-layer Perceptrons (MLP) were used as a popular model often applied in a broad class of problems [5]. The neural network has three layers. It gives one hidden layer with eight neurons. The activation function is the hyperbolic tangent.

The model is fed by eight inputs provided by a player:

- angle and distance to the mission area,
- angle and distance to the signal source,
- angle and distance to the opponent,
- heading relative to the opponent's heading,
- opponent contact – informs whether the opponent is revealed to a player or one of his possible states should be considered.

The input contains only relative values to ensure that as much general information as possible is acquired by the neural network. Thus, the training process is more efficient and less constrained to a particular case.

A. Minimax

Each player uses a classic Minimax algorithm with a limited depth of the state tree to choose the best action (Alg. VIII-A).

IX. Evolutionary Training

The utility function model (a neural network) is trained by a classic variant of a genetic algorithm [6]. It is a powerful tool suitable for complex optimization problems characterized by many local extrema. The algorithm optimizes a vector of neural network weights considered as a chromosome without further encoding. To evaluate an individual in a population, it must be decoded to his phenotype level, which means that a neural network is created based on his chromosome.

The optimization process is conducted according to the fitness function that is equal to the cost calculated for each player (Eq. 7 and Eq. 8). Because the players have different goals and fitness functions, the genetic algorithm (Alg. IX) holds two populations – one for each player type. An individual cannot be evaluated without an opponent. Therefore, these two populations periodically pass the copies of their best entities after a defined number of GA iterations. Each population adds a new opponent to a list, and the final fitness is the average cost achieved against a set of foes in the list. This is done to

Algorithm 1 Minimax (node, player, depth)

```
if depth ≤ 0 then
    ComputeUtilities ( node )
    return nil
else if IsRevealed ( player ) then
    actions ← GetActions ( player )
else
    actions ← GetPossibleStates ( player )
end if
nextPlayer ← (player + 1) % playerCount
bestChild ← nil
for child in Expand ( node, actions ) do
    Minimax ( child, nextPlayer, depth - 1 )
    if bestChild = nil ∨ GetUtility ( bestChild, player ) <
    GetUtility ( child, player ) then
        bestChild ← child
    end if
end for
if bestChild = nil then
    ComputeUtilities ( node )
else
    CopyUtilities ( node, bestChild )
end if
return bestChild
```

Algorithm 2 GeneticAlgorithm ( )

```
populations ← { evaders, pursuers }
for pop in populations do
    Evaluate ( pop )
end for
while g++ < generations do
    if g mod passBestPhase = 0 then
        PassBest ( populations )
    else
        for pop in populations do
            Select ( pop )
            Cross ( pop )
            Mutate ( pop )
        end for
    end if
    for pop in populations do
        Evaluate ( pop )
    end for
end while
```

optimize against a variety of enemy strategies rather than a single one.

X. Experiment

The experimental study includes a series of tests carried out to validate the approach. This section describes one of those experiments that was aimed to check the characteristics of a medium-long training process. Thus, it should be emphasized that in the best case scenario obtained results may only represent a near-optimal solution. This study has been preceded by a series of short experiments to select the training parameters. Because of the scale of the problem and the large amount of data, this section includes only selected results regarding the optimization process of the players' strategies.

Taking into account the nature of the research, realistic units of measure have not been preserved. They were adjusted to obtain easily observable behaviors and test the approach. In the experiment, a game arrangement is the same as presented in Fig. 3. However, the picture does not show an additional free space below and above the signal source placed in the center of the stage. A single game was simulated for a limited number of turns that was sufficient for both players to fulfill their objectives. A realistic degree of asymmetry between the players was set by giving to the submarine a better speed and the AUV greater stealth. Another difference is that the submarine is more interested in being undetected than the AUV. More detailed parameters of the experiment setup are provided in Tab. I.

Training results are shown in Fig. 5 and Fig. 6. Surprisingly, training of both populations begins from a relatively high level of fitness. Notwithstanding, an initial population is random, and there is always a chance that one of population entities will accidentally advance to the goal. It is, however, a chaotic behavior, and it changes drastically depending on the input. Thus, any opponent reaction can cause that change.

Further fitness decreases are caused by filling the list of opponents with better enemies. New opponents are added to the list every 20 generations, which is referred to as the pass best phase in the algorithm. Sometimes the period between the transfers of the best entities is too short to invent a better opponent. Nevertheless, it prevents from overtraining to a current set of opponents, which evolve quickly. Forcing an entity to achieve the best average cost against a number of opponents allows it to acquire a more general strategy but may also cause conservative and protective behaviors.

While observing behaviors of both players at the end of the training, it turned out that they acquired interesting strategies. The submarine does not immediately proceed to the mission. First, it glides far below the source, where it has a good cover. Next, it quickly travels toward the mission. It can be said that the submarine takes advantage of its speed and lures the robot. The AUV is not able to catch up with the submarine nor detect the opponent near the mission, because the submarine visits the mission only for a short moment at the end of the game. On the other hand, the AUV tries to follow the trail of the submarine. The robot minimizes its distance to the opponent but never reaches a sufficient distance to receive an award.

During the experimental study, a number of long training runs have been conducted. However, the simulation progress and outcome turned out to be very sensitive to the random nature of the genetic algorithm. Therefore, it was difficult to observe regularity, and averaging results did not lead to clear conclusions. The results from long training runs are placed in Fig. 7 and Fig. 8. For some of the experiments, fitness tend to stabilize in time (Fig. 7) while in other cases, the balance was lost and the fitness of both populations fluctuated (Fig. 8).

Minimax					
tree depth: 3	max. actions: 7	hidden opp. states: 7	game turns: 25		
Genetic Algorithm					
pop. size: 100	passBestPhase: 20	selection: tourn.	tourn. size: 10%		
elitist model	cross: uniform	cross prob.: 0.6	cross factor: 0.5		
mut. prob.: 0.1	mut. factor: 0.01	opp. list: 10			
Evader					
max. velocity: 2	max. turn angle: $\frac{\Pi}{4}$	detection range: 15	bistatic		
Pursuer					
max. velocity: 1	max. turn angle: $\frac{\Pi}{4}$	detection range: 5	bistatic		
Cost parameters					
$r_1 = 4$	$b_1 = 100$	$p_1 = 50$	$r_2 = 4$	$b_2 = 100$	$p_2 = 10$

TABLE I
THE TABLE SHOWS A DETAILED PARAMETER SETUP OF THE EXPERIMENT.

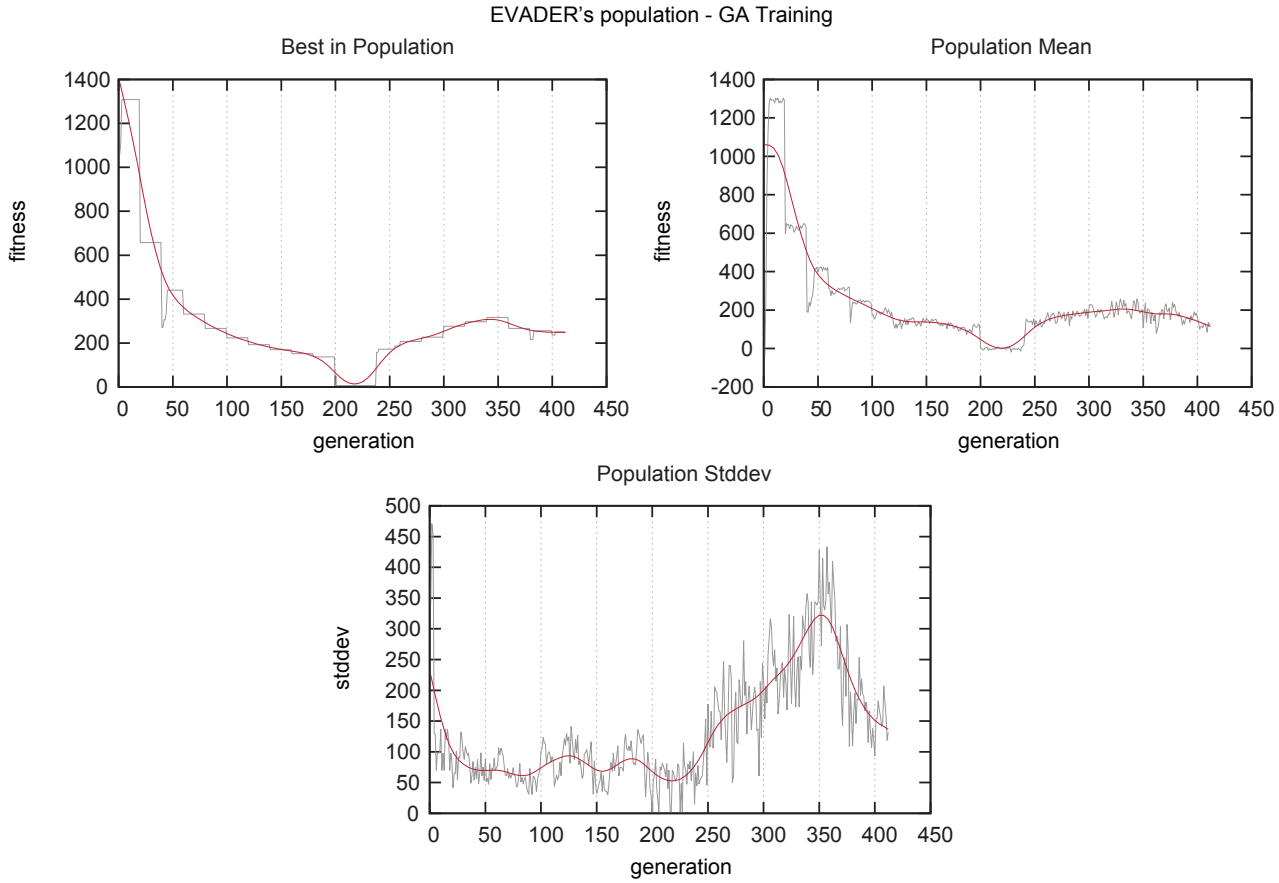


Fig. 5. The plots show the best fitness, average population fitness, and fitness standard deviation over generations of evader's population training.

The solution space is vast, and it has may local extrema. Therefore, it is difficult to clearly state if players could perform any better. Surely, training results may vary depending on the initial parameter setup. However, an important lesson is the training can be adjusted until a satisfactory outcome is achieved.

XI. SUMMARY

Through the training process, the players acquire knowledge that is encoded in the utility model. The knowledge can be used easily, requiring only reduced calculation. However, the information stored in a neural network cannot simply be exported to a human-readable form, unless it is a rule-based model. The study showed that the method can be successfully

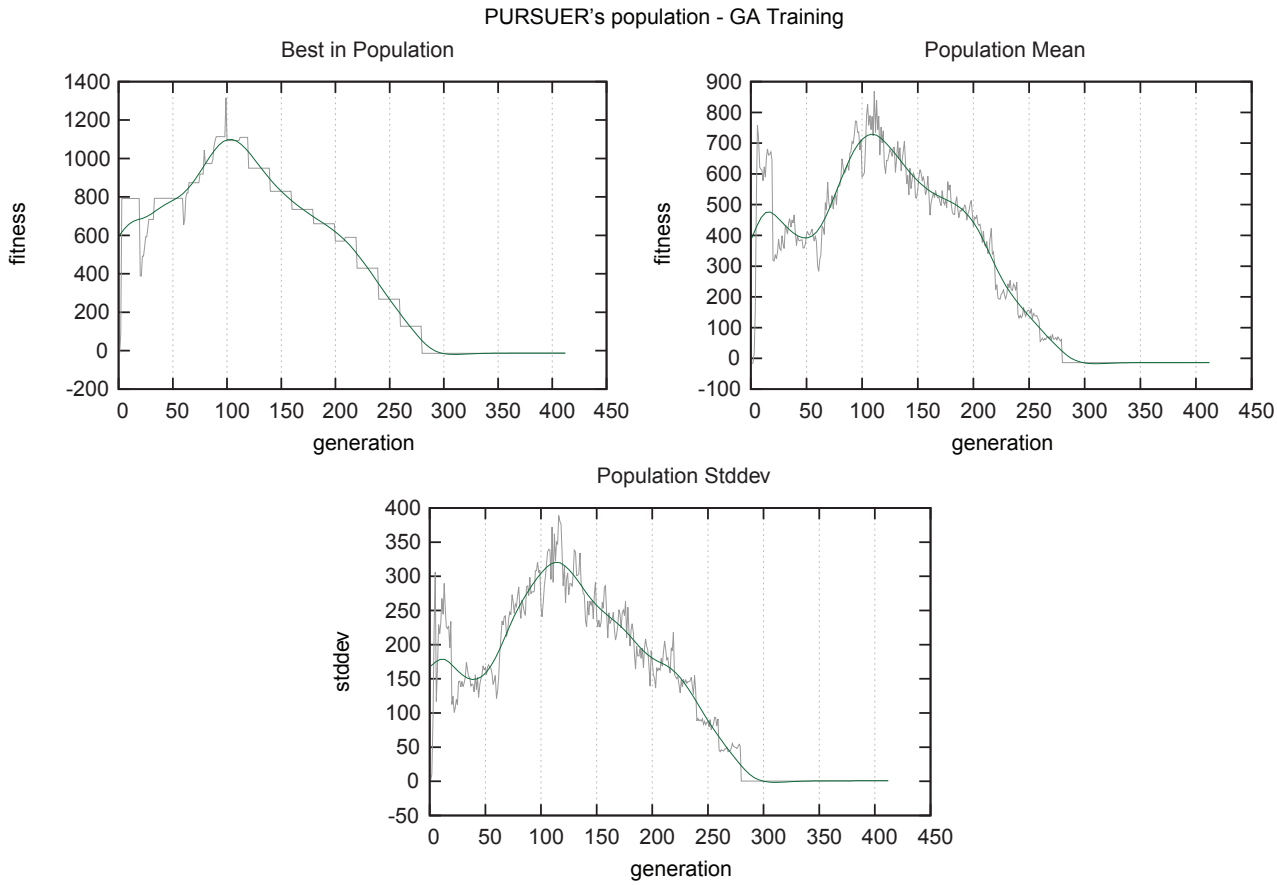


Fig. 6. The plots show the best fitness, average population fitness, and fitness standard deviation over generations of pursuer's population training.

used for observing how the system behaves depending on the game configuration.

Earlier studies have shown that the deeper Minimax algorithm penetrated a game tree, a more intelligent behavior of a player was observed, which is consistent with the theory. Also, it should be noted that the number of possible states of an unrevealed opponent taken into account by the algorithm has a congruent impact.

Another practical observation from the experiment is that the training phase is computationally expensive. Regardless of the fact that the implementation was using C++11 standard and calculations were efficiently distributed over several threads [23], the study was strongly impeded by time-consuming experiments. Accompanying tests proved that the process cannot be easily accelerated by GPU computing [24], [25]. The bottleneck is the Minimax algorithm that involves multiple calculations of the neural network's output. Despite the fact that computations in each network layer can be parallelized by GPU, the gain compared to CPU calculations was hard to observe.

An important achievement in the experimental study is that the players' strategies stabilize at some point. Strategies of competing players often oscillate when a pure-strategy equilibrium cannot be found. Although the stabilization is only one of the conditions that have to be satisfied to obtain a good solution, this should be considered as a significant success.

XII. Future Work

From a theoretical point of view, adopting a pure strategy may not be the most suitable model for this game. Nonetheless, it is one of the simplest approaches, and proved to be sufficient. One of the interesting directions is to check a mixed strategy and stochastic decision rules. Subsequently, player positions and environment parameters should be randomized. Undoubtedly, it will substantially increase the training process since additional evaluation repetitions are required to obtain an acceptable statistical significance level. However, the new model addresses environments with uncertain information, and it should give better results.

Because of the overall problem difficulty, the initial study has employed very basic tools that are commonly used in the field of computational intelligence. In the next step, it would be beneficial to use Minimax hybrids to increase the performance of the tree search [26]. Deep Learning methods could be applied to improve the training process and the generalization capabilities of a neural network [27].

At the current stage of development, the system cannot be used for building a universal model of AUV strategy that could

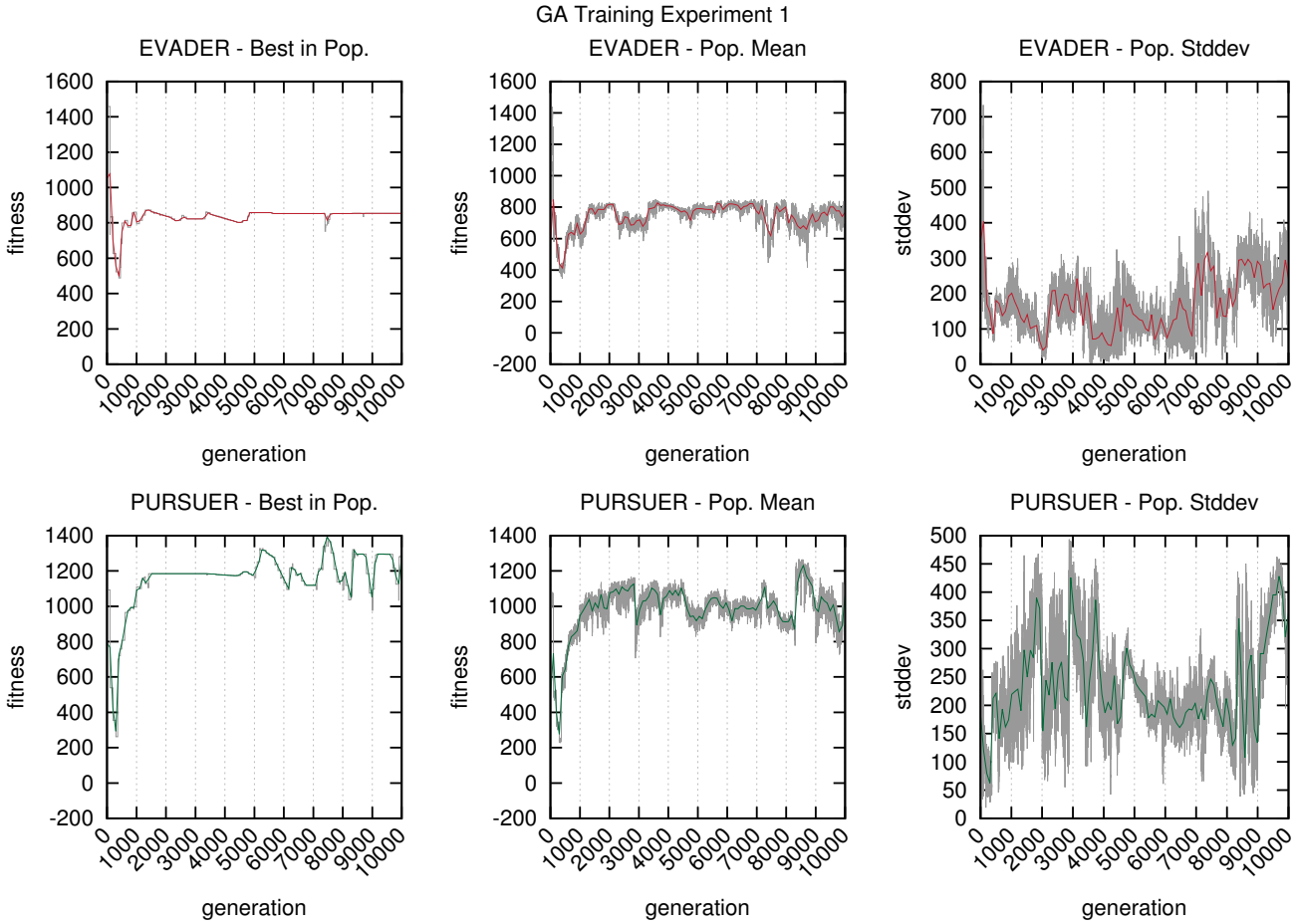


Fig. 7. A long training run 1.

be tested on actual hardware. Preparation of a reliable strategy requires a well-organized training process. The training should therefore introduce more realistic environmental parameters and cover various game scenarios including different player and mission arrangements. It should also simulate noisy and misleading information. Therefore, one of the primary goals is employing an accurate model of underwater signal propagation. It should be emphasized that the quality of the system must be evaluated according to the military knowledge and real-life scenarios.

REFERENCES

[1] P. Kachroo, *Autonomous Underwater Vehicles: Modeling, Control Design and Simulation*. CRC Press, 2010. ISBN 978-1-4398-1831-2

[2] G. Griffiths, Ed., *Technology and Applications of Autonomous Underwater Vehicles*. CRC Press, 2002. ISBN 978-0-415-30154-1

[3] Wikipedia. Blackghost photo. [Online]. Available: http://en.wikipedia.org/wiki/Autonomous_underwater_vehicle

[4] T. Basar and G. Olsder, *Dynamic Noncooperative Game Theory*, 2nd ed. Society for Industrial and Applied Mathematics, 1999. ISBN 978-0898714296

[5] S. Haykin, *Neural Networks: A Comprehensive Foundation*, 2nd ed. Prentice Hall PTR, 1998. ISBN 0-13-273350-1

[6] D. E. Goldberg, *The Design of Innovation: Lessons from and for Competent Genetic Algorithms*. Norwell, MA, USA: Kluwer Academic Publishers, 2002. ISBN 1402070985

[7] D. Wagner, W. Mylander, and T. Sanders, *Naval Operations Analysis*. Naval Institute Press, 1999. ISBN 1-557-50956-5

[8] C. Harrison, "Closed form bistatic reverberation and target echoes with variable bathymetry and sound speed," *Oceanic Engineering, IEEE Journal of*, vol. 30, no. 4, pp. 660–675, Oct 2005. doi: 10.1109/JOE.2005.862095

[9] A. Sehgal and D. Cernea, "A multi-auv missions simulation framework for the usarsim robotics simulator," in *Control Automation (MED), 2010 18th Mediterranean Conference on*, June 2010. doi: 10.1109/MED.2010.5547632 pp. 1188–1193.

[10] A. Washburn and R. Hohzaki, "The diesel submarine flaming datum problem," *Military Operations Research*, vol. 6, no. 4, pp. 19–30, September 2001. doi: 10.5711/morj.6.4.19

[11] C. Strode, "Optimising multistatic sensor locations using path planning and game theory," in *Computational Intelligence for Security and Defense Applications (CISDA), 2011 IEEE Symposium on*, April 2011. doi: 10.1109/CISDA.2011.5945938 pp. 9–16.

[12] C. Strode, B. Mourre, and M. Rixen, "Decision support using the Multistatic Tactical Planning Aid (MSTPA)," *Ocean Dynamics*, vol. 62, no. 1, pp. 161–175, 2012. doi: 10.1007/s10236-011-0483-7

[13] J. Borges de Sousa, K. H. Johansson, A. Speranzon, and J. Silva, "A control architecture for multiple submarines in coordinated search missions," in *Proceedings of the 16th IFAC world congress*. IFAC, 2005. doi: 10.1.1.65.3030

[14] A. Antoniades, H. Kim, and S. Sastry, "Pursuit-evasion strategies for

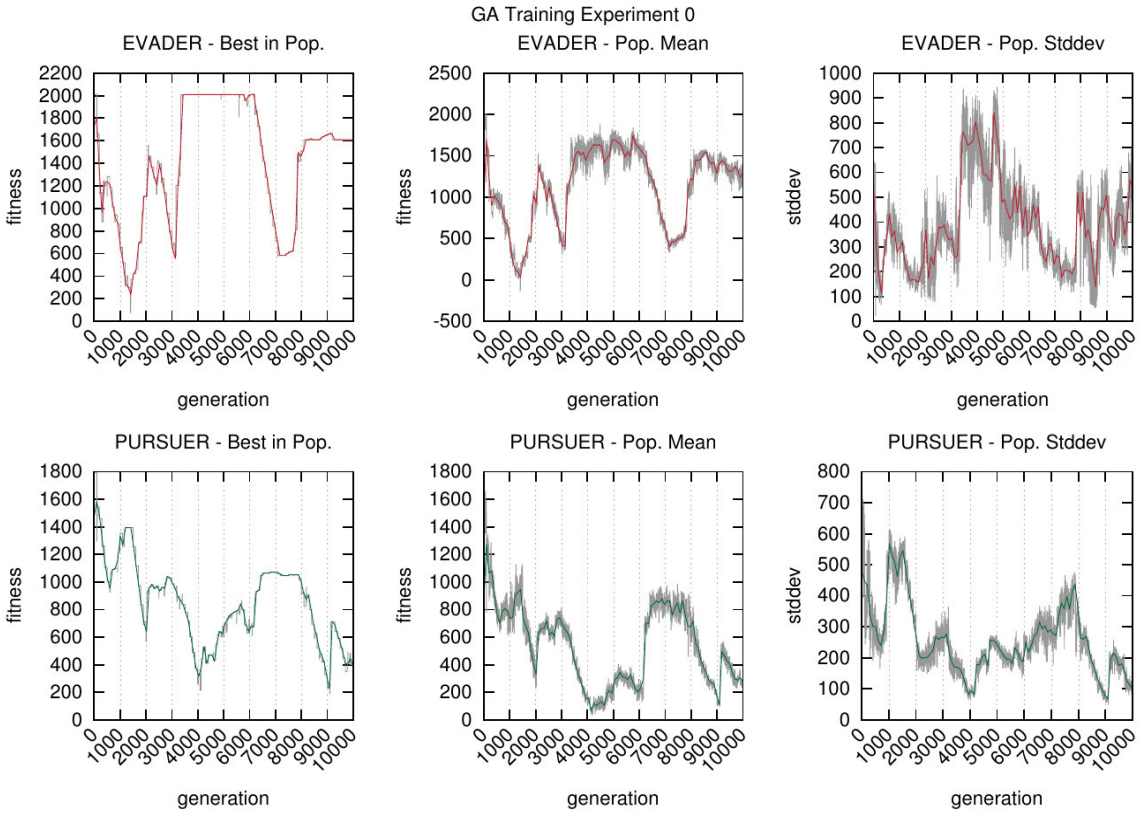


Fig. 8. A long training run 2.

teams of multiple agents with incomplete information," in *Decision and Control,* 2003. Proceedings. 42nd IEEE Conference on, vol. 1, Dec 2003. doi: 10.1109/CDC.2003.1272656. ISSN 0191-2216 pp. 756–761.

[15] T. H. Chung, G. A. Hollinger, and V. Isler, "Search and pursuit-evasion in mobile robotics," *Auton. Robots,* vol. 31, no. 4, pp. 299–316, Nov. 2011. doi: 10.1007/s10514-011-9241-4

[16] R. Vidal, O. Shakernia, H. Kim, D. Shim, and S. Sastry, "Probabilistic pursuit-evasion games: theory, implementation, and experimental evaluation," *Robotics and Automation, IEEE Transactions on,* vol. 18, no. 5, pp. 662–669, Oct 2002. doi: 10.1109/TRA.2002.804040

[17] K.-B. Sim, D.-W. Lee, and J.-Y. Kim, "Game theory based coevolutionary algorithm: A new computational coevolutionary approach," *International Journal of Control, Automation, and Systems,* vol. 2, pp. 463–474, 2004. doi: 10.1.1.132.3826

[18] H. Sally and M. Rafie, "A survey of game theory using evolutionary algorithms," in *Information Technology (ITSim),* 2010 International Symposium in, vol. 3, June 2010. doi: 10.1109/ITSIM.2010.5561648. ISSN 2155-897 pp. 1319–1325.

[19] S. Kemna, M. J. Hamilton, D. T. Hughes, and K. LePage, "Adaptive autonomous underwater vehicles for littoral surveillance: the GLINT10 field trial results," *Intelligent Service Robotics,* vol. 4, no. 4, pp. 245–258, 2011. doi: 10.1007/s11370-011-0097-4

[20] R. Isaacs, *Differential Games: A Mathematical Theory with Applications to Warfare and Pursuit, Control and Optimization.* Courier Dover Publications, 1999. ISBN 0-486-40682-2

[21] J. Pearl, "The solution for the branching factor of the Alpha-beta pruning algorithm and its optimality," *Commun. ACM,* vol. 25, no. 8, pp. 559–564, Aug. 1982. doi: 10.1145/358589.358616. [Online]. Available: http://doi.acm.org/10.1145/358589.358616

[22] K. Chellapilla and D. Fogel, "Evolution, neural networks, games, and intelligence," *Proceedings of the IEEE,* vol. 87, no. 9, pp. 1471–1496, Sep 1999. doi: 10.1109/5.784222

[23] E. Alba and J. M. Troya, "A survey of parallel distributed genetic algorithms," *Complexity,* vol. 4, no. 4, pp. 31–52, 1999. doi:10.1002/(SICI)1099-0526(199903/04)4:4<31::AID-CPLX5>3.0.CO;2-4

[24] J. Sanders. (2010) *Introduction to CUDA C. GPU Technology Conference,* NVIDIA. [Online]. Available: http://www.nvidia.com/content/GTC-2010/pdfs/2131_GTC2010.pdf

[25] S. Rennich. (2011) *CUDA C/C++ streams and concurrency.* NVIDIA. [Online]. Available: http://on-demand.gputechconf.com/gtc-express/2011/presentations/StreamsAndConcurrencyWebinar.pdf

[26] H. Baier and M. H. M. Winands, "Monte-carlo tree search and minimax hybrids," in *Computational Intelligence in Games (CIG)*, 2013 IEEE Conference on, Aug 2013. doi: 10.1109/CIG.2013.6633630. ISSN 2325-4270 pp. 1–8.

[27] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Networks,* vol. 61, pp. 85–117, 2015. doi: 10.1016/j.neunet.2014.09.003

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 41–46
DOI: 10.15439/2016F453
ACSIS, Vol. 8. ISSN 2300-5963

A new way for the exploration of a dataset based on a social choice inspired approach

Michel HERBIN, Amine AÏT YOUNES, Frédéric BLANCHARD
Université de Reims Champagne-Ardenne
CReSTIC, France
Email: {michel.herbin, amine.ait-younes, frederic.blanchard }@univ-reims.fr

***Abstract*—The exploration of a data set consists in grouping similar data. The classical statistical methods often fail when there is is no minimal assumption on the clusters. Our approach is based on the links between data, but the pairwise comparison between data and the importance of the links depend heavily on context where data lies. We propose to analyze a dataset through methods of the social choice theory where data plays both the role of a candidate and the role of a voter. The candidates are ranked by the voters and each voter gives a score to each candidate according to his ranking. We propose one specific election for each voter based on his preferences. The voters of these elections have weights computed according to their respective behaviors. In this approach, the conventional similarity indices between data are used to define the electoral behavior of each data.**

***Index Terms*—exploratory data analysis, social choice theory, representatives, vote, data reduction**

I. Introduction

ONE OF the first steps in the exploration of a data set consists in grouping similar data. For this purpose, lot of clustering methods are proposed in literature to detect clusters within a dataset. The methods often fail when there is neither a minimal assumption on the clusters nor a minimal model of the clusters. For instance the classical k-means method [8] assumes both that data could be grouped around mean values or mean vectors and that the number of clusters is known. Unfortunately the first assumption leads to important constraints on the shape of the clusters in the data space and this condition is seldom corroborated. Other approaches of clustering are based on links between data. The hierarchical agglomerative clustering methods are probably the most known methods for exploring the datasets using such links. The links are usually drawn from pairwise comparisons between data and they are based on distances or pseudo-distances [4]. But the pairwise comparison between data and the importance of the links depend heavily on context where data lies. Indeed the ranges of values of a comparison index could change when data are not in the same clusters. In other words, the links could be well suited to connect two data in one cluster and they are not adapted for the other clusters. Thus this paper proposes a new way to define the links between data through the ranks to overcome this constraint of cluster context.

We propose to analyze a dataset through methods of the social choice theory where data plays both the role of a candidate and the role of a voter [5]. The social choice inspired approach brings a metaphorical meaning that help to understand the concepts (as in bioinspired or human-inspired algorithms [10]).

The candidates are ranked by the voters and each voter gives a score to each candidate according to his ranking. Then the scores of the voters are aggregated using generally the sum of scores obtained by the candidates. In the classical procedure of election, each voter has the same weight in the aggregation. Thus this procedure is the same for all clusters. In this paper the election procedures differ from one cluster to another. We propose one specific election for each voter based on his preferences (i.e. one election per voter). The voters of these elections have weights computed by comparison of their respective behaviors. The weights differ from one election to another. The links between data are defined using these elections where each voter selects one candidate for representing itself within the dataset. The chainings between the voters and their representatives define data communities. Thus the partitions of the dataset with these communities give us a new way to explore the dataset.

The following section describes the procedure of election that we propose in this paper. It leads to a graph that permit us to structure the dataset. Then we study and we assess this method for structuring a dataset. Finally we discuss and we conclude this work.

II. Dataset and Voters

A. Collective preference

Let Ω be a dataset with n elements:

$$\Omega = \{X_1, X_2, ...X_n\}$$

In the framework of the social choice theory [9] [3], Ω is both a set of n voters and a set of n candidates. Thus each data is a voter of Ω and it also becomes an alternative that the other voters could prefer as a representative in Ω (i.e. an elected candidate of Ω).

The dataset is provided with a pairwise comparison index between data. We call D this index. In this paper, we use Euclidean distance as pairwise comparison index. But we need only two properties of D. When X_i, X_j, and X_k are three data in Ω, we should have:

- $D(X_i, X_i) \leq D(X_i, X_j)$,
- $D(X_i, X_j) \leq D(X_i, X_k)$ if X_j is more similar to X_i than X_k is (in other terms : X_j is prefered to X_k by X_i)

978-83-60810-90-3/$25.00

In the following, any pairwise comparison index should respect these two properties.

With using such a pairwise comparator, each data X_i is considered as a voter which can rank the other data. The ranks of X_i are defined between 1 and n. The ranking function is called R_{X_i} and we have:

- $R_{X_i}(X_i) = 1$,
- $R_{X_i}(X_j) \leq R_{X_i}(X_k)$ if $D(X_i, X_j) \leq D(X_i, X_k)$.

The data X_i is a voter that selects the candidates using R_{X_i} as preference indicator. The vote of X_i is realized with a score of Borda which is a classical method of social choice theory [6] [1]. In this paper, the score of Borda given by the voter X_i to the candidate X_j is defined as:

$$S_{X_i}(X_j) = \frac{n - R_{X_i}(X_j)}{(n-1)}$$

where X_j is a candidate and X_i is a voter.

The classical election procedure attributes the sum of the scores of the voters for each candidate. Thus the candidate X_j obtains the global score $S(X_j)$ defined by:

$$S(X_j) = \sum_{i=1}^{n} S_{X_i}(X_j)$$

This procedure leads to nominate the best candidate as the one with the highest score. But each voter has the same weight in this overall vote. This overall election does not take into account that two voters could belong to two different clusters.

In the following, we will change the paradigm. We consider that each voter has his own election procedure that is adapted to itself. The following describes the specific procedure for each voter.

B. Individual preference

Each voter will choose its candidate with its own election procedure. Let X_i be a voter that chooses the candidates. Each data X_j is also a voter of the election that X_i proposes. All the voters of Ω have weights that are specific of the election procedure of X_i. The weight of X_i itself is equal to one. The more similar to X_i a voter X_j is, the higher the weight of X_j is in this election. The weights are based on the similarity between the voters and the similarities with X_i are used for the election that X_i proposes.

Let us describe the similarity of the behaviors of two voters. We consider that two voters X_i and X_j are similar when their respective ranking function R_{X_i} and R_{X_j} are similar. The correlation of Spearman [11] is classically used to evaluate the correlation between ranks. The higher the correlation is close to 1, the more ranks are correlated. In this paper the correlation gives us an index of the similarity of the behavior of two voters. Spearman correlation between X_i and X_j is defined by:

$$Cor(X_i, X_j) = 1 - \frac{6 * \sum_{k=1}^{n} (R_{X_j}(X_k) - R_{X_i}(X_k))^2}{n^3 - n}$$

$Cor(X_i, X_j)$ lies between -1 and 1. We consider that X_i and X_j have similar behavior when the Spearman correlation is greater then a positive threshold which is a significance level. If we call t this level, then X_i and X_j become similar when $Cor(X_i, X_j) \geq t$.

Let $w_{X_i}(X_j)$ be the weight given to the voter X_j for the election based on the preferences of X_i.

We define this weight by:

$$w_{X_i}(X_j) = max(0, \frac{Cor(X_i, X_j) - t}{1 - t})$$

The weight lies between 0 and 1. It is equal to zero when X_i and X_j are not similar.

In the election based on the preferences of X_i, each candidate X_j obtains a score $Score_{X_i}(X_j)$ defined by:

$$Score_{X_i}(X_j) = \sum_{k=1}^{n} w_{X_i}(X_k) \times S_{X_k}(X_j)$$

Thus this election is based on a sum of scores weighted by the similarity of the voters with X_i. Other voters similar to X_i participate in the election of the representative of X_i.

C. Communities of voters

The representative of X_i becomes the one which have the highest score within Ω for the election based on the preferences of X_i. So each voter X_i has one representative in Ω elected by the specific election of X_i : $Rep_{Score}(X_i)$.

$$Score(Rep_{Score}(X_i)) = \max_{k=1}^{n}(Score_{X_i}(X_k))$$

We define a graph in Ω where the vertices are the voters and the edges are the links between the voters and their representatives. Each connected components of this graph defines a community of voters. The more we claim a high correlation between voters, the more the size of communities is reduced. In other words, the higher the threshold t is close to 1, the more the communities are small and the number of communities increases within Ω. These communities give a data structuration to study a dataset when we have neither assumption nor model for the clusters.

If the threshold t is close to 1, the representative of each voter X_i is based only on the preference of X_i. If this threshold decrease, other voters similar to X_i participate in the election of the representative of X_i.

Each data Xi has two representatives: the favorite candidate of Xi and the elected candidate of the local election of Xi. For each data X_i we define an individual loss indicator, wich represents the correlation loss between X_i and theses two representatives. The collective $loss$ indicator for the data is the sum of all the individual loss.

$$loss = \sum_{k=1}^{n} loss_{Ind}(X_k)$$

$$loss_{Ind}(X_i) = Cor(X_i, Rep_S(X_i)) - Cor(X_i, Rep_{Score}(X_i))$$

with

$$S(Rep_S(X_i)) = \max_{k \neq i}(S_{X_i}(X_k))$$

$$Score(Rep_{Score}(X_i)) = \max_{k \neq i}(Score_{X_i}(X_k))$$

III. Experimental Study

This section is devoted to the study of our method for structuring a dataset with a pairwise comparator. First let us present an example of the different steps of the dataset structuration with our method. In a second section, we assess the quality of this structuration using simulated data. Third the quality is assessed when using one real dataset.

TABLE I: Number of communities of voters, number of unique representatives and loss, using the simple example of Fig.1 when the threshold t of correlation varies between 0 and 1.

	t	nbcom	nbrep	loss
1	0.00	2	4	0.85
2	0.05	2	4	0.85
3	0.10	2	4	0.85
4	0.15	2	4	0.85
5	0.20	2	4	0.85
6	0.25	2	4	0.85
7	0.30	2	4	0.85
8	0.35	2	4	0.77
9	0.40	2	4	0.77
10	0.45	2	4	0.77
11	0.50	2	4	0.78
12	0.55	2	4	0.78
13	0.60	2	5	0.59
14	0.65	2	5	0.59
15	0.70	2	5	0.52
16	0.75	3	8	0.32
17	0.80	3	9	0.23
18	0.85	3	11	0.17
19	0.90	5	14	0.00
20	0.95	5	14	0.00
21	1.00	20	20	0.00

A. Workflow for structuring a dataset

We propose to explore a dataset with 20 simulated data in dimension 2 (see Fig.1-A). The pairwise comparisons are based on Euclidean distance. We conduct the overall election with a classical Borda's procedure. This overall election permits us to propose the best candidate which could be considered as the representative of the whole dataset (see Fig.1-B). Then we proceed to the elections based on the individual preferences for obtaining linking each data with another one. These links allow to define communities of voters. The procedure of the election with the individual preferences is based on a threshold of correlation. Fig.1-C shows the number of communities when the correlation threshold increases.

The higher the threshold, the higher the number of communities is. The highest threshold leads to the highest number of unique representatives. The higher the threshold, the lesser the losses are (both individual and collective). When the threshold is equal to 0.5, 0.95 and 0.99 (Fig.1-D, Fig.1-E, Fig.1-F) :

- the number of communities is 2, 5 and 6 (resp.)
- the number of unique representatives is 4, 14 and 15 (resp.)
- the collective loss is 0.78, 0 and 0 (resp.)

This number of communities is less than the number of data. That gives a new way for the exploration of a dataset.

B. Assessment of the links structuring a dataset

TABLE II: Number of communities of voters, number of unique representatives and loss, using the three classes of Fig.2 and criterion of assessment when the threshold of correlation varies between 0 and 1.

	t	nbcom	nbrep	loss
1	0.00	3	20	6.99
2	0.05	3	19	6.65
3	0.10	3	19	6.37
4	0.15	3	18	6.49
5	0.20	3	16	6.23
6	0.25	3	17	6.09
7	0.30	3	17	6.00
8	0.35	3	16	6.01
9	0.40	3	16	6.03
10	0.45	3	16	5.95
11	0.50	3	16	5.82
12	0.55	3	17	5.70
13	0.60	3	19	5.47
14	0.65	3	21	5.00
15	0.70	3	24	4.58
16	0.75	3	31	4.04
17	0.80	3	35	3.03
18	0.85	4	47	2.32
19	0.90	7	58	1.37
20	0.95	10	74	0.52
21	1.00	150	150	0.00

In this paper, we place ourselves resolutely in the context of the exploratory analysis of data without any a priori assumption on eventual classes, we only use an index of pairwise comparison. But the use of classes gives the most classical way to evaluate a structuration of a dataset. So this paper uses classes to assess only the links that we propose between data. The detection of classes (i.e. the clustering) is out of the scope of this paper.

The assessment of our method for structuring a dataset is performed using a dataset with known classes. Each data belongs to one class and it has the label of its class. Using our structuration each data is also linked to a representative in the dataset. A data is well represented when its own label is equal to the label of its representative. In such case the link between a voter and its representative remains inside a class of the dataset. We propose a structuration of the dataset with graphs. The vertices of the graph are labeled and the edges are labeled when their extremities have the same label. We compute the number of the labeled edges.

The percentage of such edges could assess the quality of the structuration through a graph. Unfortunately the classes are unknown in the first step of data exploration. Thus we propose to use the loss indicator instead of this percentage of labeled links.

The higher is this quality criterion and the lower the number of communities is, then the better the structuration is.

Table II gives the values of this criterion when the threshold of correlation lies between 0 and 1. The dataset is simulated in dimension 2 (see Fig.2) and the number of detected

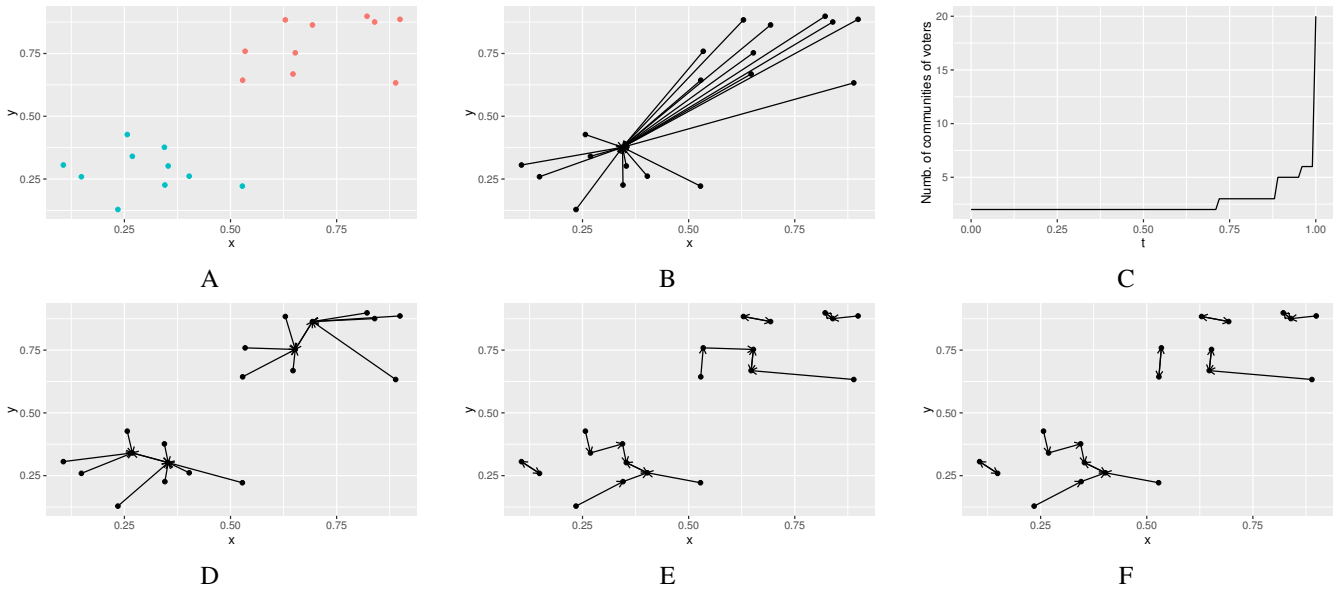


Fig. 1: Twenty simulated data in dimension 2 (A), the overal election selecting one representative (B) and elections based on individual preferences leading to several communities whose number depends on the correlation threshold (C). 2, 5, and 6 communities respectively obtained with a correlation threshold equal to 0.5, 0.95, 0.99 (D, E, F).

communities of voters is displayed when the threshold of correlation between voters increases from 0 to 1. Fig.2 gives also three examples of the communities when the threshold is respectively equal to 0, 0.5 and 0.9.

TABLE III: Number of communities of voters, number of unique representatives and loss, using the three classes of Fig.3 and criterion of assessment when the threshold of correlation varies between 0 and 1.

	t	nbcom	nbrep	loss
1	0.00	1	74	35.49
2	0.05	1	70	32.88
3	0.10	1	69	30.32
4	0.15	1	72	28.65
5	0.20	1	69	26.90
6	0.25	1	70	24.89
7	0.30	1	74	23.56
8	0.35	4	79	21.55
9	0.40	4	75	19.58
10	0.45	3	76	17.25
11	0.50	3	82	15.09
12	0.55	5	92	11.73
13	0.60	7	101	9.50
14	0.65	7	107	7.94
15	0.70	11	118	6.54
16	0.75	13	128	5.27
17	0.80	14	134	3.99
18	0.85	19	149	3.10
19	0.90	25	166	1.96
20	0.95	36	186	0.88
21	1.00	380	380	0.00

In the following we simulated a dataset with three classes that are hardly distinguishable because of their shapes and their overlapping. The dataset ($n = 380$) is simulated in dimension 2 with three uniform distributions in two rectangular crowns with 200 and 80 data and one rectangle with 100 data (see Fig.3). The number of voter communities is displayed when the threshold of correlation between voters increases from 0 to 1. Fig.3 gives also three examples of the communities when the threshold is respectively equal to 0.5, 0.8 and 0.99,

- the number of communities is 3, 14 and 71 (resp.)
- the number of unique representatives is 82, 134 and 212 (resp.)
- the collective loss is 15.09, 3.99 and 0.11 (resp.)

the number of communities are respectively equal to 8, 16 and 106. In such a case the classical clustering methods fail to detect meaning clusters. Indeed classical clustering methods are often based on statistics such as means or medoids. They use these statistics to determine the clusters and they make the assumption that data could be well represented with such statistics. Unfortunately these statistical approaches are unadapted in this case. Table III gives the values of our assessment criterion when the threshold of correlation lies between 0 and 1.

C. Assessment with real data

We use the databases from Machine Learning Repository of UCI [2] to assess our method with real data. Iris is the classical database that has 150 iris plants with 4 attributes and three clusters. Table IV gives the results we obtain with this dataset. Fig.4 displays the number of voter communities and the percentage of labeled links when the correlation threshold increases from 0 to 0.99, and the loss indicator.

IV. Discussion and Conclusion

In this paper we describe and we implement a method for exploring a data set. The main originality of this method lies in

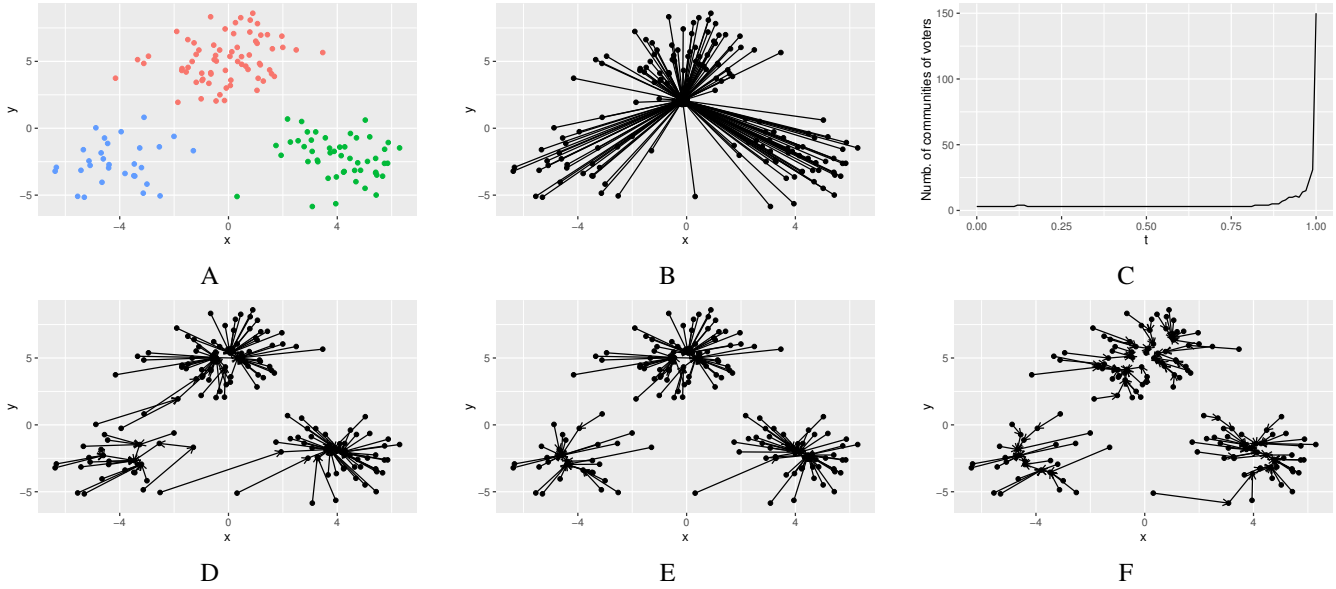


Fig. 2: Simulated data with three classes (three multinomial distributed subsamples)

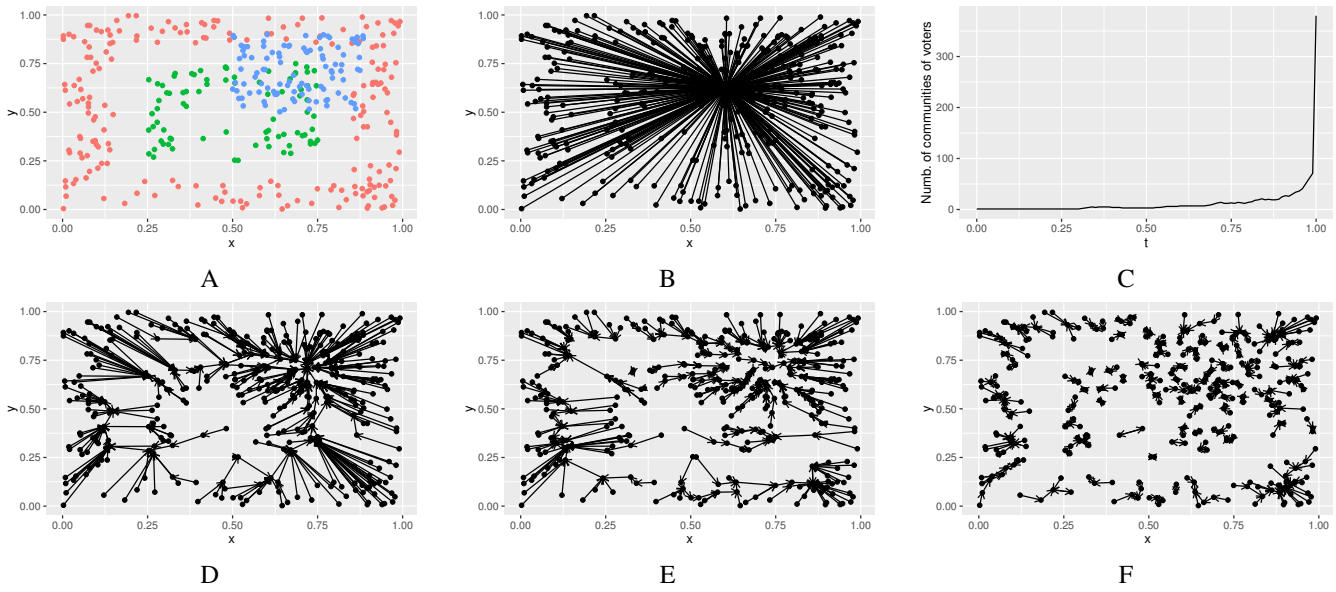


Fig. 3: Simulated data with three uniform distributions in two rectangular crowns of respectively 200 and 80 data and one rectangle of 100 data. The three classes are hardly distinguishable with classical clustering methods.

the definition of links between data. These links are based on a local election mechanism with individual preferences that connects each data to another data designated by the local election process. In this approach, the conventional similarity indices between data are used to define the electoral behavior of each data. As the preferences of the users in a recommender system, the voters then have weights corresponding to the similarity of electoral behaviors. However this approach by recommender systems is not used in this paper and the robustness of our method when data is incomplete or imperfect could be studied in future work.

Another important contribution of this work is to reduce the size of a data set from the exploration of a set of n data to a set of p communities where p is much smaller than n. This approach of dimensionality reduction has the advantage that it makes no assumption about the shape or the exact number of communities. It thus constitutes a preliminary step to a more meaningful clustering and it leads to select a more suitable method for the exploring dataset. This extension of our work could also be involved in further work.

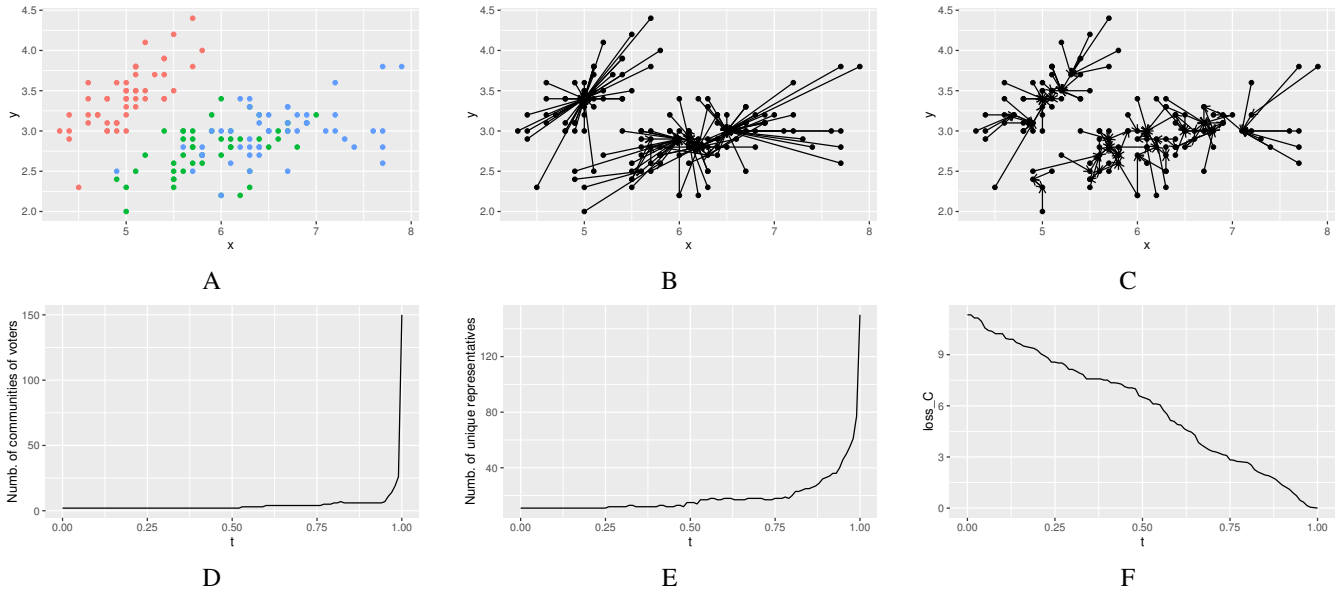


Fig. 4: Iris Data ($n = 150$) with three classes of 50 data in dimension four. Top : data projections in dimension two using sepal width and sepal length and detection of communities when the correlation threshold is equal to 0.5 and 0.95. Bottom : Number of voter communities and the percentage of labeled links when the correlation threshold increases from 0 to 0.99, and the loss indicator

TABLE IV: Number of communities of voters, number of unique representatives and loss, using the three classes of the Iris data (see Fig.4) and criterion of assessment when the threshold of correlation varies between 0 and 1.

	t	nbcom	nbrep	loss
1	0.00	2	11	11.33
2	0.05	2	11	10.57
3	0.10	2	11	10.23
4	0.15	2	11	9.64
5	0.20	2	11	9.25
6	0.25	2	11	8.56
7	0.30	2	12	8.14
8	0.35	2	12	7.58
9	0.40	2	12	7.51
10	0.45	2	12	7.11
11	0.50	2	15	6.51
12	0.55	3	17	6.06
13	0.60	4	18	4.90
14	0.65	4	17	4.27
15	0.70	4	18	3.33
16	0.75	4	18	2.83
17	0.80	6	20	2.67
18	0.85	6	25	1.94
19	0.90	6	33	1.34
20	0.95	7	46	0.46
21	1.00	150	150	0.00

We are currently working on application for sensor network data analysis.

ACKNOWLEDGMENT

This work is partially supported by the EC SCOOP project (INEA/CEF/TRAN/A2014/1042281).

REFERENCES

[1] A. Aït Younes, F. Blanchard and M. Herbin, "New similarity index based on the aggregation of membership functions through OWA operator", *Federated Conference on Computer Science and Information Systems*, FedCSIS 2015, 163–168, Lódz, Poland, 2015.

[2] K. Bache, M. Lichman, "UCI Machine learning repository", *http://archive.ics.uci.edu/ml, University of California, Irvine, School of Information and Computer Sciences*, 2013.

[3] J.P. Barthélémy and B.Montjardet, "The median procedure in cluster analysis and social choice theory", *Mathematical Social Sciences*, 1:235–267, 1981.

[4] A. Bellet, A. Habrard, M. Sebban, "A Survey on Metric Learning for Feature Vectors and Structured Data", *Technical report*, arXiv:1306.6709, 2014.

[5] F. Blanchard, C. de Runz, M. Herbin, H. Akdag, "Représentativité et graphe de représentants : une approche inspirée de la théorie du choix social pour la fouille de données relationnelles", *Atelier Fouille de Données Complexes, Conférence Extraction et Gestion des Connaissances*, EGC, 73-83, Brest, France, 2011.

[6] M. de Borda, "Memoire sur les elections au scrutin", *Academie Royale des Sciences*, Paris, 1784.

[7] A.K. Jain, M.N. Murty, P.J. Flynn, "Data Clustering: A Review", *ACM Computing Surveys*, 31(3), 264–323, 1999.

[8] A.K. Jain, "Data clustering: 50 years beyond K-means", *Pattern Recognition Letters*, 31, 651–666, 2010.

[9] J.N. Mordeson, D.S. Malik, T.D. Clark, "Application of Fuzzy Logic to Social Choice Theory", Chapman and Hall/CRC, 2015.

[10] M. Parsapoor, U. Bilstrup, "An Emotional Learning-inspired Ensemble Classifier (ELiEC)", *Proceedings of the 2013 Federated Conference on Computer Science and Information Systems*, 137-141, 2013

[11] C. Spearman, "General intelligence objectively determined and measured", *Am J Psychol*, 15, 201-293, 1904.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 47–52
DOI: 10.15439/2016F546
ACSIS, Vol. 8. ISSN 2300-5963

Identifying Fishing Activities from AIS Data with Conditional Random Fields

Baifan Hu, Xiang Jiang
Erico N de Souza, Ronald Pelot
Faculty of Computer Science
Dalhousie University Halifax, NS, Canada B3H 1W5
Email: {baifanhu, Xiang.Jiang, erico.souza, Ronald.Pelot}@dal.ca

Stan Matwin
Faculty of Computer Science,
Dalhousie University Halifax, NS, Canada B3H1W5,
and Institute of Computer Science
Polish Academy of Sciences
Email: stan@cs.dal.ca

***Abstract*—Fishing activity detection is important for fishery management to maintain abundant oceans. This paper presents a novel approach to identifying fishing activities from Automatic Identification System (AIS) data using Conditional Random Fields (CRFs). CRFs are popular for solving structured prediction problems such as sequence labeling in natural language processing. To model the conditional probability distributions that can identify fishing activities of the vessel points, we treat attributes of vessel points as observed variables and the fishing and non-fishing labels as hidden variables. We present three experiments and two comparisons to demonstrate the stability and effectiveness of the resulting models.**

I. Introduction

GLOBAL overfishing causes a dramatic decline in the fish population. Several major commercial fish species are endangered which threatens the ocean ecosystem. This affects millions of people who depend on fish for food and living. According to a 2014 report by the United Nations Food and Agriculture Organization[1], more than 90 percent of global fisheries are over exploited. Unfortunately, these fishing activities are often illegal, unreported and unregulated (IUU), which makes tracing them a challenge. Thus, in order to make fishing activities more transparent for practicing sustainable fisheries, fishing activity detection is urgently needed.

In 2000, International Maritime Organization (IMO) firstly introduced the Automatic Identification System (AIS) to enhance the security and safety of maritime navigation. Ships equipped with AIS can automatically broadcast information, including unique identification (Maritime Mobile Service Identity, MMSI), position, speed, course and further details of the vessel to its nearby ships and coastal authorities. AIS data are openly accessible and not encrypted. In 2008, satellites AIS technology was implemented, enabling the collection of massive and reliable information of vessels in global areas within seconds. Consequently, satellites AIS data could be used as an ideal source to monitor vessel movements and detect fishing activities around the world.

In this paper, we present a novel approach for identifying fishing activities using Conditional Random Fields and demonstrate its stability of performance using three different evaluation experiments and two comparisons. The remainder of the paper is organized as follows: In Section 2, we describe relevant literature; in Section 3, we explain conditional random fields (CRFs) and then elaborate on how to apply them to identify fishing activities; in Section 4, we present three experiments and two comparisons and their results on historical AIS data to demonstrate the stability and effectiveness of our models; finally, in Section 5, we discuss practices and provide direction for future work.

[1]Food and Agriculture Organization of the United Nations, 2014. http://www.fao.org

II. Background and Related Work

CRFs are prevalent in solving structured prediction problems [1]. It has been applied to many tasks in natural language processing (NLP) such as part-of-speech (POS) tagging [2], [3], shallow parsing [4] and name-entity recognition (NER) [5], [6]. In addition, Hierarchical CRFs [7] has been applied to extract human activities. We find similarity between these tasks and fishing activity detection from the following perspective. In POS tagging, the goal is to label words in sentences using word-category tags. The labels depend on both the word's meaning and context. This task involves two random variables, X and Y, where X is a sequence of words, and Y is a sequence of POS tags. Linear-chain conditional random fields can model the conditional probability distribution $p(\mathbf{y}|\mathbf{x})$ to predict POS tags. Similarly, the task of fishing activity detection involves two random variables, X and Y, where X is the observed random variable (which represents sequences of coordinates and speeds), and Y is the hidden random variable to be predicted (Y is a sequence of fishing and non-fishing labels). Consequently, it is reasonable to test whether the linear-chain conditional random fields can model the conditional probability distribution $p(\mathbf{y}|\mathbf{x})$ of fishing–non-fishing to detect fishing activities.

Most studies of fishing activity detection focus on Trawlers. For example, Mazzarella et al. identified fishing events using a clustering method [8], and Peel and Good recognized vessel activities using Hidden Markov Model [9]. It is found that trawler fishing activities are highly related with speed, meaning speed can provide useful information to aid the classification of fishing activities. However, for longliners, Souza et al. found that there is no obvious pattern to distinguish fishing activities using speed information alone [10]. Souza et al.

applied Lavielle's unsupervised trajectory segmentation algorithm, inspired by animal movements, to identify the longliner fishing behavior. Jiang et al. applied a deep learning approach using autoencoders (AE) that are pretrained with restricted Boltzmann Machines [11]. To the best of our knowledge, [10] was the first paper that we can find to apply Machine Learning to the task of fishing and non-fishing detection. Here we apply a supervised machine learning method (CRFs) to detect longliner fishing activities.

III. Linear-Chain Conditional Random Fields

A. Basic Principles of Linear-Chain Conditional Random Fields

Linear-chain conditional random fields are undirected graphical models that represent conditional probability distributions of random variables $V = X \cup Y$ that take the form

$$p(\mathbf{y}|\mathbf{x}) = \frac{1}{Z(\mathbf{x})} \prod_{j=1}^{n} \psi_j(\mathbf{x}, \mathbf{y}), \tag{1}$$

where X is a set of observed variables, Y is a set of labels that we need to predict, and $Z(\mathbf{x})$ is a normalization or partition function

$$Z(\mathbf{x}) = \sum_{\mathbf{y}'} \prod_{j=1}^{n} \psi_j(\mathbf{x}, \mathbf{y}'), \tag{2}$$

where $\psi_j(\mathbf{x}, \mathbf{y})$ are compatibility functions over a subset of random variables $A \subset V$, and n is the number of compatibility functions $\psi(\mathbf{x}, \mathbf{y})$ that factorize the probability distribution. Given compatibility functions in the form

$$\psi_j(\mathbf{x}, \mathbf{y}; \lambda) = exp(\sum_{i=1}^{m} \lambda_i f_i(y_{j-1}, y_j, \mathbf{x}, j)), \tag{3}$$

the conditional probability distribution can be written as

$$p(\mathbf{y}|\mathbf{x}; \lambda) = \frac{1}{Z(\mathbf{x})} exp(\sum_{j=1}^{t} \sum_{i=1}^{m} \lambda_i f_i(y_{j-1}, y_j, \mathbf{x}, j)), \tag{4}$$

where m is the number of feature functions, t is the length of sequence $\mathbf{y}$, and λ is a set of weight parameters that help provide weighted average over these feature functions.

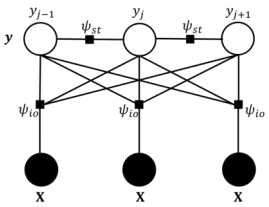


Fig. 1. Factor graph of linear-chain conditional random fields. $\mathbf{x}$ denotes a sequence of input data, $\mathbf{y}$ denotes a sequence of labels and ψ_{st} and $\psi_{\mathbf{i}o}$ denote two compatibility functions.

In (3), the parameter λ_i of compatibility functions $\psi_j(\mathbf{x}, \mathbf{y}; \lambda)$ does not depend on the index j, which means the parameters are shared along the linear chain. Fig. 1 visualizes the factor graph of linear-chain conditional random fields with two compatibility functions. In (4), the first sum runs over each position of the linear chain and the second sum runs over each feature function. The conditional probability $p(\mathbf{y}|\mathbf{x}; \lambda)$ can be represented as a mapping function from features to labels. Thus, the selection of feature functions is of great significance for the performance of a model. The parameters can be estimated using maximum-likelihood. The log-likelihood can be estimated with the Forward-Backward Algorithm. The inference of finding the most likely sequence $\mathbf{y}$ given observations $\mathbf{x}$ is performed using the Viterbi Algorithm [1]. The remainder of this section demonstrates how we adapt CRFs to identifying fishing activities.

B. Discretization

Compared with POS tagging where the input is a sequence of discrete words, AIS trajectory consists of real-valued features that are continuous by nature, such as longitudes and latitudes. Conditional random fields can model real-valued features, but they typically require proper normalization so that the value of the feature function is a linear function of the conditional probability $p(\mathbf{y}|\mathbf{x})$. However, since the relationships between AIS features and fishing activity labels are non-linear, we discretize the features to relax the normalization constraints and allow the conditional random fields to learn $p(\mathbf{y}|\mathbf{x})$ with a more flexible representation. We use a variant of equal interval binning discretization in our work. Each bin is associated with a set of parameters to fit the model.

C. Feature Functions

We use two sets of compatibility functions ψ_{st} and $\psi_{\mathbf{i}o}$ to factor $p(\mathbf{y}|\mathbf{x})$:

$$p(\mathbf{y}|\mathbf{x}) = \frac{1}{Z(\mathbf{x})} (\psi_{st}(\mathbf{x}, \mathbf{y}) \cdot \psi_{\mathbf{i}o}(\mathbf{x}, \mathbf{y})). \tag{5}$$

The first compatibility function ψ_{st} is transition compatibility function, which models the transition probability of the labels from one state to another and takes the form

$$\psi_{st}(\mathbf{x}, \mathbf{y}; \lambda) = exp(\lambda_{st} f_{st}(y_{j-1}, y_j, \mathbf{x}, j)), \tag{6}$$

where $f_{st}(y_{j-1}, y_j, \mathbf{x}, j)$ is transition feature function that takes the form

$$f_{st}(y_{j-1}, y_j, \mathbf{x}, j) = 1_{\{y_{j-1}=s\}} 1_{\{y_j=t\}}, \tag{7}$$

and st are all possible combinations of labels, y_{j-1} and y_j are the labels of the $(j-1)$-th and j-th position of the linear chain and $\mathbf{x}$ is the input sequence. More concretely, if the labels can only take two values: 0 and 1, the feature functions can be rewritten as

$$f_{00}(y_{j-1}, y_j, \mathbf{x}, j) = 1_{\{y_{j-1}=0\}} 1_{\{y_j=0\}}, \tag{8}$$

$$f_{01}(y_{j-1}, y_j, \mathbf{x}, j) = 1_{\{y_{j-1}=0\}} 1_{\{y_j=1\}}, \tag{9}$$

$$f_{10}(y_{j-1}, y_j, \mathbf{x}, j) = 1_{\{y_{j-1}=1\}} 1_{\{y_j=0\}}, \tag{10}$$

$$f_{11}(y_{j-1}, y_j, \mathbf{x}, j) = 1_{\{y_{j-1}=1\}} 1_{\{y_j=1\}}, \tag{11}$$

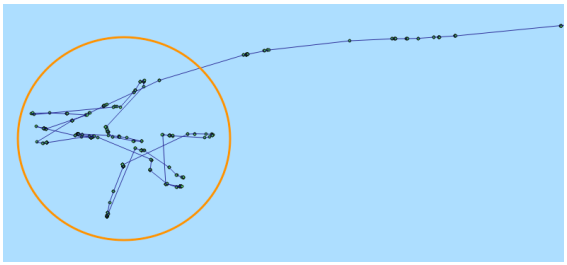


Fig. 2. Differences of fishing and non-fishing tracks. Tracks in the orange circle are fishing tracks and the rest are non-fishing tracks.

which are parameterized by $\lambda_{0,0}, \lambda_{0,1}, \lambda_{1,0}, \lambda_{1,1}$ respectively.

The second compatibility funciton $\psi_{\mathbf{i}o}$ is state-observation compatibility function, which models the probability distribution of the labels given a set of observations. The subscript $\mathbf{i}$ within $\psi_{\mathbf{i}o}$ represents a set of input observations, which is different from the index i of parameter λ in (4). The state-observation compatibility function can be written as

$$\psi_{\mathbf{i}o}(\mathbf{x}, \mathbf{y}; \lambda) = exp(\lambda_{\mathbf{i}o} f_{\mathbf{i}o}(y_{j-1}, y_j, \mathbf{x}, j)), \tag{12}$$

where $f_{\mathbf{i}o}(y_{j-1}, y_j, \mathbf{x}, j)$ is state-observation feature function that takes the form

$$f_{\mathbf{i}o}(y_{j-1}, y_j, \mathbf{x}, j) = 1_{\{y_j = o\}} 1_{\{\mathbf{x}=\mathbf{i}\}}, \tag{13}$$

where $\mathbf{i}o$ are all possible combinations of input features and its corresponding labels. The state-observation feature functions are parameterized by a set of parameters $\lambda_{\mathbf{i}o}$.

IV. Experiments

A. Data Preprocessing

AIS data contain attributes including MMSI, time, longitude, latitude, speed over ground (SOG) and course over ground (COG). In the experiments, we use historical AIS data from 14 longliners around the world collected from June 1st 2012 to Dec 31st 2013. The fishing and non-fishing activities of these data are labeled by a marine biology expert. Fig. 2 shows the differences between fishing and non-fishing tracks that are recovered from discrete AIS signals. The fishing tracks are in the form of a zigzag. The non-fishing tracks tend to follow smooth lines. To preprocess the data, we perform data cleansing to remove uninformative data, data conversion from absolute values to differential values (i.e. difference with to the previous point), data discretization to transform continuous value into nominal counterparts, and feature selection to fit the model with the most relevant features.

1) Data Cleansing: We sort the data points of each longliner in chronological order. We then remove repetitive data points as well as data points with incomplete features. We further detect and remove outliers if the speed exceeds the normal range and the location deviates from its normal trajectories. After data cleansing, we have 505893 longliners data points in total. On average, 77% of the data points are labeled as fishing. Table I shows a summary of the 14 longliners after cleansing.

TABLE I
Summary of the 14 vessels data

Track ID	Track Size	# of Fish Points	% of Fish Activity
1	21556	17148	79.6
2	8829	6326	71.7
3	30166	24422	81.0
4	6086	4226	69.4
5	28184	22153	78.6
6	37710	32977	87.4
7	24715	16857	68.2
8	2032	1755	86.4
9	12111	8470	69.9
10	17161	14277	83.2
11	2670	1761	66.0
12	90429	78765	87.1
13	108826	73005	67.1
14	115418	86256	74.7
Mean	36135	27742	76.5

2) Differential longitude and latitude: In our early experiments, we trained our predictive models using absolute value of longitudes and latitudes. However, we find the resulting models overfitting on the training data and cannot generalize to different locations. In order to generalize the model into other areas, we use differential longitude, that is the difference of absolute longitudes between the current and previous data points. Thus the absolute longitudes $L = [l_1, l_2, ..., l_n]$ are transformed to differential longitudes that takes the form $L_d = [l_2 - l_1, l_3 - l_2, ..., l_n - l_{n-1}]$. Similarly, we transform absolute latitudes to its differential form.

3) Discretization: As mentioned in Section 3, we discretize the attributes of data using a variant of equal interval binning. Early experiments indicate different features require different sizes of intervals. For differential longitudes and latitudes, the interval size m takes the form

$$m = \begin{cases} 0.05 & l \in [-1, 1] \\ 20 & l > 1, l < -1 \end{cases} \tag{14}$$

where fine intervals are selected in the range $[-1, 1]$. Because most differential longitudes and latitudes are between -1 and 1, fine intervals provide a larger number of parameters to fit the model compared to coarse intervals, when l is greater than 1 or less than -1. For SOG, we set m to 0.5. For COG, we set m to 20.

4) Feature Selection: In early experiments, we built the model using different combinations of features, such as differential longitude, differential latitude, SOG and COG. We find the model built with differential longitude, differential latitude, and SOG performed the best. We use these features in the following three experiments of this paper, shown in Fig. 3. More precisely, we use the following feature functions in our experiments: 1) pairs of longitudes and latitudes are selected to represent the positions of data points (colored in orange); 2) pairs of neighbouring longitudes are selected to represent the changes of longitudes over time (colored in red); 3) pairs of neighbouring latitudes are selected to represent the changes of latitudes over time (colored in green); 4) speed and pairs of neighbouring speed are selected to represent speed information

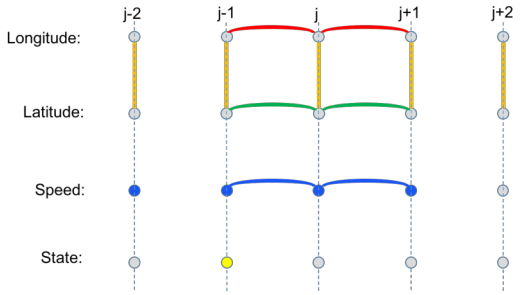


Fig. 3. Feature functions. When predicting the label of index j, the values of colored dots as well as the paired values connected by solid lines are selected as feature functions.

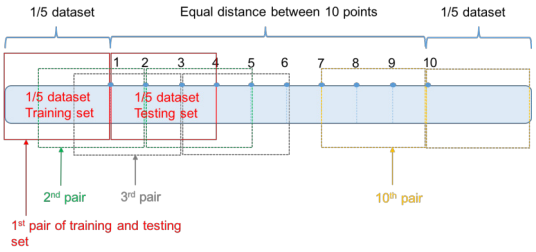


Fig. 4. Modified Monte Carlo. Choose 10 points as split point, obtain 10 pairs of training and testing sets.

(colored in blue); 5) label of the previous state is selected to help model the transition probability between states (colored in yellow). These feature functions employ spatial and temporal information to aid the classification task.

B. Results

We train CRF models using CRF++ [12]. We design three experiments to evaluate the model: Modified Monte Carlo methods, Iterative Leave One Batch Out (LOBO) and Stratified LOBO. We then compare CRFs with autoencoders and a data mining approach.

1) Modified Monte Carlo: Here, we concatenate all the data and apply Monte Carlo methods to get 10 pairs of training and testing data sets. The size of the training and testing set is $\frac{1}{5}$ of the total number of data points. To find the dividing points of the ten Monte Carlo experiments, we first set aside $\frac{1}{5}$ of the whole data set in the front and back respectively, then select 10 dividing points with identical intervals that can cover the entire dataset. For each selected point, the $\frac{1}{5}$ portion of the data set to the left of the point and the $\frac{1}{5}$ portion to the right constitute one pair of training and testing set. Fig. 4 shows the way we obtain our 10 pairs of training and testing sets.

TABLE II
EVALUATION USING MODIFIED MONTE CARLO METHODS .

Expt ID	Accuracy	Sensitivity	Specificity	PPV	NPV
1	0.868	0.481	0.959	0.730	0.888
2	0.872	0.460	0.947	0.610	0.906
3	0.959	0.826	0.985	0.916	0.966
4	0.973	0.876	0.991	0.953	0.976
5	0.860	0.622	0.941	0.785	0.879
6	0.765	0.821	0.739	0.598	0.897
7	0.851	0.743	0.908	0.809	0.871
8	0.944	0.885	0.965	0.900	0.960
9	0.939	0.735	0.981	0.892	0.947
10	0.946	0.835	0.983	0.940	0.947
Mean	0.898	0.728	0.940	0.813	0.924
SD	0.065	0.157	0.075	0.131	0.040

For each Monte Carlo experiment, the model is evaluated using accuracy, sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV), as shown in Table II, where the positive class is non-fishing. These measurements provide us with information about the overall performance of the model. Also, we present the mean and standard deviation (SD) of the metrics.

2) Iterative Leave One Batch Out: Here, we first split the 14 vessels into 2 groups, group one with 10 vessels for iterative LOBO, and group two with four vessels for stratified LOBO. To split the vessels into two groups, we first take four vessels from the 14 longliners as group two and then take the rest 10 vessels as group one. To get an accurate evaluation of the model, we need the four selected vessels to be representative of the entire data set. Therefore, we use trajectory size as a criterion to help select these vessels. Since the trajectory sizes of the 14 vessels vary from thousands to hundred thousands, as shown in Table I, we categorize the 14 trajectories into three sets based on their sizes. For the three sets, the sizes of trajectories are in the range of thousands, ten thousands and hundred thousands respectively. We then randomly select vessels that are proportional to the cardinality of each set, one vessel in set one, two vessels in set two, and one vessel in group set as the four testing vessels.
For the set of 10 vessels, in each iteration, we consider one vessel as one batch to be the test vessel and build one model on the rest of 9 vessels, and repeat this for 10 times. The results of Iterative Leave One Batch Out for the 10 test vessels are shown in Table III.

3) Stratified Leave One Batch Out: We use the selected four vessels from experiment 2 as independent test vessels and we train the model using the 10 vessels from experiment 2, and evaluate on the rest four vessels individually. The performance of the model on the four testing vessels are shown in Table IV. We further visualize the classification results in Fig. 5.

4) Comparisons with autoencoders and Data Minging approach: We reproduce the autoencoders [11] and the data mining approach [10] on the same set of data in previous ILOBO experiment. We compare the performance of CRFs with these two methods as shown in Table III. We perform paired-samples t-test to compare the classification accuracies of CRFs with autoencoders and the data mining approach. The resulting p-value from the comparison of CRFs and autoencoders is 0.057. For the comparison of CRFs and the data mining approach, the p-value is 0.032 which is less than significance level 0.05 so that we can conclude these two methods are significantly different.

TABLE III
EVALUATION USING ITERATIVE LEAVE ONE BATCH OUT.

ID	Accuracy			Sensitivity			Specificity			PPV			NPV		
	CRF	AE	DM	CRF	AE	DM	CRF	AE	DM	CRF	AE	DM	CRF	AE	DM
1	**0.86**	0.83	0.65	**0.74**	0.60	0.45	0.90	**0.93**	0.91	0.75	0.77	**0.85**	**0.90**	0.85	0.57
2	0.85	0.85	**0.89**	0.53	0.39	**0.71**	0.93	**0.96**	0.93	0.63	**0.76**	0.70	0.89	0.86	**0.93**
3	0.83	**0.86**	0.46	**0.92**	0.58	0.35	0.78	**0.94**	0.93	0.65	0.76	**0.95**	**0.96**	0.88	0.25
4	0.86	0.86	**0.87**	0.50	0.51	**0.66**	**0.96**	0.95	0.95	0.75	0.74	**0.82**	0.88	0.88	**0.89**
5	**0.96**	0.86	0.89	**0.74**	0.44	0.54	**0.99**	0.94	0.98	**0.88**	0.57	0.83	**0.96**	0.90	0.90
6	**0.82**	0.76	0.76	0.52	0.34	**0.59**	**0.96**	0.94	0.89	**0.87**	0.68	0.81	**0.81**	0.776	0.73
7	**0.92**	0.80	0.54	**0.53**	0.00	0.20	**0.98**	0.93	0.94	0.80	0.00	**0.81**	**0.93**	0.85	0.5
8	**0.90**	0.84	0.86	**0.44**	0.38	0.55	**0.99**	0.94	0.96	**0.88**	0.55	0.80	**0.90**	0.88	0.80
9	0.83	0.86	**0.88**	0.55	0.62	**0.89**	**0.97**	0.94	0.86	0.92	0.76	**0.93**	0.81	**0.89**	0.80
10	**0.91**	0.85	0.74	**0.72**	0.56	0.49	**0.98**	0.94	**0.98**	0.91	0.76	**0.96**	**0.91**	0.87	0.66
Mean	**0.87**	0.84	0.75	**0.62**	0.44	0.54	**0.94**	**0.94**	0.93	0.80	0.64	**0.85**	**0.89**	0.86	0.71
SD	0.05	0.03	0.16	0.15	0.19	0.19	0.06	0.01	0.04	0.10	0.24	0.08	0.05	0.04	0.22

TABLE IV
EVALUATION USING STRATIFIED LEAVE ONE BATCH OUT.

Expt ID	Accuracy	Sensitivity	Specificity	PPV	NPV
1	0.871	0.642	0.929	0.700	0.910
2	0.818	0.576	0.922	0.759	0.835
3	0.991	0.944	0.998	0.988	0.992
4	0.888	0.824	0.919	0.833	0.914
Mean	0.892	0.747	0.942	0.820	0.913
SD	0.072	0.168	0.038	0.125	0.064

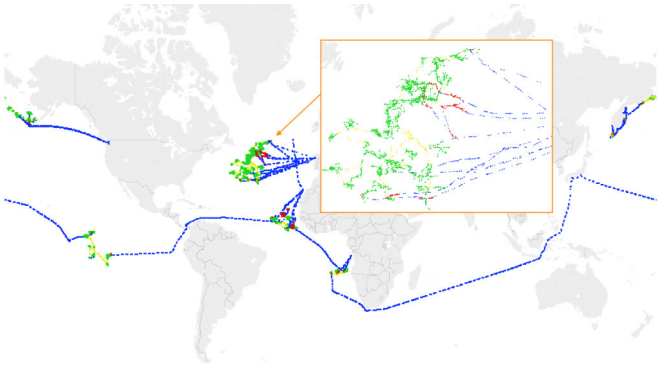


Fig. 5. The Visualization of four independent testing vessel tracks. Green points mean both the label and the prediction are fishing. Blue points mean both the label and the prediction are non-fishing. Red points mean the label is fishing while the prediction is non-fishing. Yellow points mean the label is non-fishing while the prediction is fishing. Zoomed-in region gives details of classification results in the Atlantic ocean area.

C. Discussion

The results of the three experiments are consistent. The average accuracy of the three experiments is 88.7% with 6.1% average standard deviation. By comparing the results of the second and the last experiments, we found that the models built using Iterative Leave One Batch Out perform as good as the model built in Stratified Leave One Batch Out. This further proves the stability and potential of the model in future fishing activity detection.

Through comparisons, we find CRFs can have better classification accuracy, in terms of mean and standard deviation. We also find CRFs do not suffer from imbalanced data as autoencoders: in experiment 7, autoencoders label all data points as positive class with 0% of sensitivity, whereas CRFs have 53.4% of sensitivity. According to our t-test results, despite the fact that the CRFs and autoencoders are not systematically different on significance level 0.05, CRFs can perform as well as and sometimes better than autoencoders and the data mining approach.

V. CONCLUSIONS AND FUTURE WORK

This paper presents an approach to detecting fishing activities from historical AIS data using Conditional Random Fields. Data cleaning, discretization and transformation followed by feature selection are performed to preprocess data. We then specify proper feature functions to train CRF models. The resulting models are further evaluated in three different ways. In terms of efficiency, we find CRFs can be trained efficiently than complex models such as deep learning. In terms of effectiveness, the three evaluation experiments suggest the model can generalize well in future fishing activity detection problems.

As for future work, we will investigate better ways of developing additional features, such as density and angle. We also consider systematic approaches to incorporate additional density and angle information into feature functions to aid the development of the model.

ACKNOWLEDGMENT

The authors have been supported by the Natural Sciences and Engineering Research Council of Canada. Last author's research is also supported in part by the National Research Centre of Poland (NCN) grant DEC-2013/09/B/-ST6/01549.

REFERENCES

[1] C. Sutton and A. McCallum, "An introduction to conditional random fields," *Machine Learning*, vol. 4, no. 4, pp. 267–373, 2011.
[2] J. Lafferty, A. McCallum, and F. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," in *Proceedings of the eighteenth international conference on machine learning, ICML*, vol. 1, 2001, pp. 282–289.
[3] A. PVS and G. Karthik, "Part-of-speech tagging and chunking using conditional random fields and transformation based learning," *Shallow Parsing for South Asian Languages*, vol. 21, 2007.

[4] F. Sha and F. Pereira, "Shallow parsing with conditional random fields," in *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1*. Association for Computational Linguistics, 2003, pp. 134–141.

[5] A. McCallum and W. Li, "Early results for named entity recognition with conditional random fields, feature induction and web-enhanced lexicons," in *Proceedings of the seventh conference on Natural language learning at HLT-NAACL 2003-Volume 4*. Association for Computational Linguistics, 2003, pp. 188–191.

[6] B. Settles, "Biomedical named entity recognition using conditional random fields and rich feature sets," in *Proceedings of the International Joint Workshop on Natural Language Processing in Biomedicine and its Applications*. Association for Computational Linguistics, 2004, pp. 104–107.

[7] L. Liao, D. Fox, and H. Kautz, "Extracting places and activities from gps traces using hierarchical conditional random fields," *The International Journal of Robotics Research*, vol. 26, no. 1, pp. 119–134, 2007.

[8] F. Mazzarella, M. Vespe, D. Damalas, and G. Osio, "Discovering vessel activities at sea using ais data: mapping of fishing footprints," in *Information Fusion (FUSION), 2014 17th International Conference on*. IEEE, 2014, pp. 1–7.

[9] D. Peel, N. M. Good, and T. Quinn II, "A hidden markov model approach for determining vessel activity from vessel monitoring system data," *Canadian Journal of Fisheries and Aquatic Sciences*, vol. 68, no. 7, pp. 1252–1264, 2011.

[10] E. N. de Souza, K. Boerder, S. Matwin, and B. Worm, "Improving fishing pattern detection from satellite ais using data mining and machine learning," *PLOS ONE*, vol. 11, no. 7, p. e0158248, 2016.

[11] X. Jiang, D. L. Silver, B. Hu, E. N. de Souza, and S. Matwin, "Fishing activity detection from ais data using autoencoders," in *Canadian Conference on Artificial Intelligence*. Springer, 2016, pp. 33–39.

[12] CRF++: Yet Another Toolkit . [Online]. Available: https://taku910.github.io/crfpp/

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 53–60
DOI: 10.15439/2016F309
ACSIS, Vol. 8. ISSN 2300-5963

Sparse Coding Methods for Music Induced Emotion Recognition

Jan Jakubik
Wrocław University of Science and Technology
Department of Computational Intelligence
Wrocław, Poland
Email: jan.jakubik@pwr.edu.pl

Halina Kwaśnicka
Wrocław University of Science and Technology
Department of Computational Intelligence
Wrocław, Poland
Email: halina.kwasnicka@pwr.edu.pl

***Abstract*—The paper concerns automatic recognition of emotion induced by music (MER, Music Emotion Recognition). Comparison of different sparse coding schemes in a task of MER is the main contribution of the paper. We consider a domain-specific categorization of emotions, called Geneva Emotional Music Scale (GEMS), which focuses on induced emotions rather than expressed emotions. We were able to find only one dataset, namely Emotify, in which data are annotated with GEMS categories, this set was used in our experiments. Our main goal was to compare different sparse coding approaches in a task of learning features useful for predicting musically induced emotions, taking into account categories present in the GEMS. We compared five sparse coding methods and concluded that sparse autoencoders outperform other approaches.**

I. Introduction

MUSIC information retrieval has been gaining an increased amount of attention in machine learning community over the past decade due to the increasing popularity of new musical services and a huge amount of music data available on the internet, annotated using imperfect systems such as community tagging and description supplied by producers. With the increasing amount of files, automated music analysis and content-based retrieval becomes increasingly relevant.

One particular subset of music information retrieval tasks is music emotion recognition (MER) [1]. Emotion recognition, besides the potential use in recommendation systems and search engines, poses an interesting theoretical problem for artificial intelligence researchers due to its high subjectivity and the fact human emotions are still not fully understood by psychologists. For music emotion recognition, even the set of emotions employed to annotate the dataset and use as a ground truth in machine learning is still subject to discussion. Emotional categories and scales in popular datasets are often not domain-specific.

Another important issue in automated emotion recognition is that emotion can often be related to musicological concepts such as chords and tempo [2], which from a machine learning standpoint correspond to high-level features that require specialized algorithms to extract them from the music files. However, these algorithms can be imperfect and affect the performance significantly. Meanwhile, current trends in deep learning suggest that algorithms which learn to extract more complicated features from low-level features in an unsupervised manner can achieve even better results than hand-crafted features designed for a specific domain [4]. In music information retrieval, this kind of approach has become popular in the context of genre recognition [5][6] and generalized auto-tagging task [7][8], which includes simple emotional tags, but is more focused on semantic categories such as genre, male/female vocalist, etc.

In this paper, we describe a machine learning approach based on aggregation of sparse vectors constructed from the low-level description of a sound file and apply it to Emotify, a publicly available MER dataset. It is different from usual emotion recognition datasets in that it uses Geneva Emotional Music Scale, an emotional scale designed specifically for music-induced emotion. We compare different sparse coding approaches in order to find out whether these methods can reliably learn features useful for recognition of musically induced emotion, especially the highly subjective and hard to define emotional categories present in GEMS, for which the usually employed emotional scale with Valence-Arousal dimensions does not account.

The paper is organized as follows: Section II summarizes literature related to our paper. Section III describes GEMS emotion categorization and the dataset we used in our experiments. Section IV describes different sparse coding methods which can be used interchangeably in our approach. Section V explains our representation of music files, create using sparse vector pooling, and the rationale behind it. In Section VI, we report the results of our experiments on the Emotify dataset.

II. Related work

Music emotion recognition is an important part of automated music information retrieval and as such, it has been covered extensively in existing literature. Typically, an MER task in artificial intelligence research takes one of two forms: a multi-class annotation task in which each song in a dataset is annotated with a subset of a predefined set of tags [9][10], or a regression task multiple continuous values representing different dimensions of perceived emotion are assigned to a song [11][12]. In the case of multi-class annotation, there is no single consistent system of emotional categories. The set of tags can be based on one of existing emotion categorization

978-83-60810-90-3/$25.00

systems in psychology [13] [14]. However, these are not domain-specific and can be imperfect for describing music.

A common concept in work focusing on dimensional scales is the Valence-Arousal space [15], in which only two continuous dimensions are considered. Valence differentiates between positive and negative emotions while Arousal describes the intensity of perceived emotion. E.g., music described as "energetic" or "joyful" could be placed on the V-A scale in the high valence, high arousal area while music described as "calm" will be in the low arousal, middle valence area. It should be noted that the concept of placing specific emotions on the V-A scale was not designed specifically for describing emotions concerning music. Another criticism of this scale is that its two dimensions are insufficient to describe more complex emotions and discern pairs of emotions such as fear and anger (both of which are negative emotions with high arousal). The question whether V-A scale should be extended with a third dimension is an important subject of discussion among both MIR researchers and psychologists [16].

An important aspect of music emotion is the differentiation between *induced* and *expressed* emotions. Since one of the basic purposes of music is influencing the listener's mood, that differentiation should be a subject of interest when researching emotion modelling. A domain-specific emotion categorization system called Geneva Emotional Music Scale which takes the distinction between induced and expressed emotion into account was proposed in [17]. We describe it in more detail in section III, along with Emotify, the first MER dataset annotated with GEMS categories. The dataset was researched from a machine learning viewpoint in [18], which examined the performance of available features from multiple existing MATLAB toolboxes in combination with SVR algorithm. However, more complicated approaches such as stochastic process modelling or codebook methods have not been applied to GEMS-annotated data yet.

Among existing music information retrieval approaches, we chose to examine codebook methods. Codebook methods rely on building a *dictionary* of discernible patterns appearing in sample data on a short time-scale and then expressing the contents of a music file using contents of the dictionary. In recent years, papers concerning codebook methods reported good results in music information retrieval tasks [5][8] using the Restricted Boltzmann Machine algorithm [19]. In [8], results on the state of the art level were achieved in generalized music auto-tagging using temporal pooling of sparse vectors to represent a music file. We apply the same pooling approach to music emotion modelling. The authors established that sparse RBM can learn features better than sparse coding methods used for codebook generation in the past. Another promising coding method that, unlike RBM, has not been tested extensively in this context is an autoencoder neural network. Autoencoders without sparsity constraints have been tested as an emotion recognition method in [20], using a separate network for each modelled emotion. They have been applied to chord recognition [21] with good results. Hence, in our experiments we test autoencoder with a sparsity inducing loss function as an alternative to RBM.

III. GEMS Emotion Recognition

Geneva Emotional Music Scale is a categorical model of emotion designed specifically for the music domain, based on psychological research [17]. GEMS authors propose a hierarchy with three levels, with three general categories on the top level, nine emotions in the middle and 45 specific emotions at the bottom. The emotional categories were created using surveys concerning music-induced emotion. It is important to note that the difference between emotions expressed by music and induced by music is relevant to the choice of terms. Surveys explicitly asked separate questions about emotions participants felt and emotions they perceived in music.

In [22], nine emotions from the middle level of GEMS hierarchy were chosen to annotate the Emotify dataset, consiting of 400 songs from 4 genres. These middle-level categories are: amazement, solemnity, tenderness, nostalgia, calmness, power, joyful activation, tension, and sadness. The annotations were gathered using a Facebook game Emotify, and thus can represent the task of modelling a community consensus. For each song, annotations in the form of vectors consisting of zeroes (emotion not felt) and ones (emotion felt) were gathered from multiple subjects. Overall annotation of a song can be calculated as a mean of the annotation vectors corresponding to it, resulting in a vector of 9 continuous values ranging from 0 to 1, where zero means a complete agreement that the song does not evoke a particular feeling and one complete agreement that it evokes said feeling.

The dataset was analysed from a machine learning viewpoint in [18]. Authors compared three different sets of automatically extracted features and a set of manually annotated musicological features. First analysed feature set was features designed for music description available in MIRtoolbox, a MATLAB toolbox for music information retrieval. The second one used Mel-frequency Cepstral Coefficients (MFCC) [23] and statistical functionals applied to them such as mean, variance, skewness, etc. Third feature set consisted of harmonic features proposed by the authors used in combination with MIRtoolbox features. The authors concluded that while certain emotions can be modelled with decent accuracy, the performance of machine learning is heavily limited by the issue of subjectivity. The most subjective emotional categories identified by the authors were amazement, solemnity, and tension.

IV. Sparse Representation of Data

Sparse coding is the idea of approximating data drawn from a multidimensional space by representing it in another space in a way that encourages sparsity, i.e. a vector in the output space should consist mostly of zeroes. In its simplest form, the problem of approximating a vector y with its sparse representation x in a base D (called a dictionary), can be defined as:

$$x^* = \arg\min_x \|x\|_0 \quad s.t. \quad \|y - Dx\| < \lambda \tag{1}$$

where λ is a parameter dictating the desired accuracy of reconstruction and $\|x\|_0$ is the number of non-zero elements of vector x. However, this optimization problem is NP-hard and assumes a linear transformation between two spaces, which can sometimes be insufficient to create sparse representations for complex data. Below we describe approaches practically applicable to the problem of sparse representation for any given set of data.

A. K-means Clustering

K-means clustering [24] is a well-known unsupervised learning algorithm in which a dataset is divided into clusters, and each cluster is represented by a single point in the data space, which is the centroid of the cluster. The algorithm assigns any vector in the dataset to the cluster represented by a point closest to it. This enables us to treat K-means clustering as a very restrictive form of sparse coding. A vector y in the original data space assigned to n-th cluster can be represented by a vector x in which every component except n-th is 0, and the n-th component is 1. The dimensionality of x is equal to the number of clusters. Note that this is equivalent to minimizing $\|y - Dx\|$ with constraints $\|x\|_0 = 1$ and $\|x\|_1 = 1$, where the dictionary D is a matrix which n-th column is the centroid of the n-th cluster.

For any vector which is not present in the dataset used for dictionary training, its sparse representation can be calculated by simply choosing the centroid closest to it, and creating a vector of zeros with n-th component being one, where n is the closest centroid. This allows us to express the popular Bag-of-Words model [3] as a specific case of our approach, which will be explained in section IV. This model is known for its simplicity and efficiency regarding the dictionary building process.

B. L1 Regularized Least Squares

A basic way of solving the sparse representation problem in polynomial time is to use L1 norm regularization:

$$x^* = \arg\min_x \|y - Dx\| + \lambda\|x\|_1 \tag{2}$$

L1 norm is known to encourage sparsity [25] and there has been a significant amount of research dedicated to fast solvers for both L1 regularized least squares and its nonnegative variant. Nonlinearity is possible to achieve using the kernel trick [26], the problem then becomes:

$$x^* = \arg\min_x \|\Phi(y) - \Phi(D)x\| + \lambda\|x\|_1 \tag{3}$$

where Φ denotes a mapping function from the original data space to a space of larger dimensionality, implicitly defined by a kernel function $K(a, b)$ which replaces dot product.

Variants of regularized least squares are commonly used in sparse coding problems. They can achieve better reconstruction than a K-means based dictionary approach, at the cost of more complicated dictionary building process. Another issue is that the encoding of a vector is not explicitly given and requires solving (2), which is the main disadvantage of this approach.

C. Sparse Autoencoder

Autoencoders [27] are a type of neural networks developed mostly for use in deep learning. In its basic form, an autoencoder is simply a standard feedforward neural network in which the number of input neurons is equal to the number of output neurons. The network is trained using a backpropagation algorithm [28], in which the network is given a matrix of training vectors X and a matrix desired outputs Y. In the case of autoencoders $Y = X$, meaning the objective of learning is to create an encoding in hidden layer that enables the network to reconstruct later input vectors with maximum possible accuracy.

Without additional modifications to a standard loss function used in backpropagation (e.g. mean squared error), data compression is achieved by using a hidden layer with a low number of neurons, relatively to the number of input neurons. However, it is possible to encourage sparsity [29] by modifying the loss function for hidden layer. Denoting a loss function in relation to data matrix X and desired output matrix Y as $L(X, Y)$, the new loss function minimized by the backpropagation algorithm becomes:

$$L'(X, Y) = L(X, Y) + \lambda \sum_i (\rho \log \frac{\rho}{\hat{\rho}_i} + (1 - \rho) \log \frac{1 - \rho}{1 - \hat{\rho}_i}) \tag{4}$$

where ρ is a parameter indicating the desired average activations in the hidden layer, and $\hat{\rho}_i$ is the average activation of i-th neuron in that layer over the whole dataset. This encourages neuron activations in hidden layer to be 'sparse' in the sense that only a few neurons relevant to recognizing a particular data pattern are 'active' in response to a given input, i.e. they should exhibit significantly higher activations than all the others. However, the vector of hidden layer outputs is not sparse in a strict sense, as most of the 'inactive' neuron outputs are not equal to 0, and it is nearly impossible to achieve zero activations through backpropagation. We will explain why this is not an issue in our model when discussing the representation of music by pooling sparse vectors.

D. Sparse RBM

Restricted Boltzmann Machine [19] is a stochastic neural network model consisting of a set of visible units and a set of hidden units. These layers are fully connected to each other, however, there are no connections inside a layer. Assuming Gaussian visible nodes and binary hidden nodes (which is sufficient to model real-valued inputs) a configuration of network (x, h), where x is the vector of visible values and y is the vector of hidden values, is associated with an energy function:

$$E(x, h) = \frac{1}{2\sigma^2} x^T x - \frac{1}{\sigma^2} x^T W h - a^T x - b^T h \tag{5}$$

where W is the weight matrix, a and b are bias terms. σ is a scaling parameter. This energy function is inversely proportional to the log-likelihood of observing a particular

configuration. Probability distribution of $x^{(i)}$, the value of i-th visible neuron, assuming a hidden layer configuration h is Gaussian with mean $a_i + w_i^T h$ and standard deviation σ:

$$p(x^{(i)}|h) = \mathcal{N}(a_i + w_i^T h, \sigma^2) \tag{6}$$

where w_i denotes the i-th row of W and a_i the i-th component of bias vector a. Probability for j-th hidden layer unit being active is given by:

$$p(h^{(j)}|x) = sig(\frac{1}{\sigma^2}(b_j + w_j x)) \tag{7}$$

where b_j is the j-th component of bias vector b, w_j is the j-th column of W and sig is the logistic sigmoid function. Given these definitions, a model of (W, a, b) can be learned by maximizing the log-likelihood of visible values x using the contrastive divergence learning [30].

A penalty term can be added to the objective function to encourage sparsity. Given a sequence of training vectors in which i-th vector is x_i, the penalty term is:

$$penalty = \lambda \sum_j (\rho - \frac{1}{m} \sum_i \mathbf{E}(h^{(j)}|x_i)) \tag{8}$$

where ρ is the desired average activation of hidden layer and λ is a scaling parameter.

Encoding of a given input vector can be calculated using equation (7), which is similar to coding in autoencoder networks. Restricted Boltzmann Machines are known for their efficiency in building deep neural network architectures, however, these deep networks are usually fine-tuned after the initial learning, using backpropagation. In our approach, this fine-tuning in a supervised manner is not possible.

V. Music Representation by Pooling Sparse Vectors

We apply a method of representing music files described in [8], in which a machine learning scheme based on the sparse representation of data was employed to great success. Proposed approach achieved state-of-art results in the generalized music annotation task, in which the goal is to select a subset from a set of predefined tags for each music file in the dataset. These tags included genre tags, simple categorical emotion tags and other descriptive tags such as 'female vocalist'. It is important to note that emotion recognition does not necessarily rely on the same features of music as genre recognition. For emotion recognition, it is significantly more important to take rhythmic, tempo and harmonic features into account. For example, the differentiation between major and minor chords, while not very important when differentiating between classical and rock music, becomes crucial when attempting to differentiate sad and happy songs due to strong cultural associations between major-minor scales and emotions of happiness-sadness.

In our approach, firstly we calculate the spectrogram of a music file, using log scale for frequency. A sequence of vectors is built in which every vector represents a spectrogram patch of l consecutive frames of the spectrogram, with maximum overlap between patches (i.e. for a window length l a vector x_i in the sequence contains frames from i to $i+l-1$, then the next vector x_{i+1} contains frames from $i+1$ to $i+l$). This way, each vector represents a $f \times l$ patch of the spectrogram, where f is the number of frequency bins. Overall, a spectrogram of t time-frames with f bins results in a sequence of $t - l + 1$ vectors with fl components each.

For each vector in the resulting sequence, a sparse representation is calculated. We aggregate the information from the sequence of sparse vectors via pooling, two methods of pooling are considered: max pooling and average pooling.

In max pooling, a sequence of n-dimensional vectors is aggregated to a single n-dimensional vector by choosing the maximum value of i-th dimension among all vectors in the sequence as the i-th component of the resulting vector. In average pooling, we simply sum vectors in the sequence and divide the result by the number of vectors. For both pooling methods, the length of resulting vector representing the music file will be equal to the size of sparse representation vector. It is possible to use one of the methods over the entire file, or to mix them by first splitting the sequence of vectors into fragments of length m, use one type of pooling over the fragment, resulting in a sequence that is m times shorter, and then aggregate that sequence using the second type of pooling. The process is shown in Fig. 1.

The intuitive rationale behind using pooling for a sequence of sparse vectors is that every dimension in sparse representation corresponds to a particular pattern which can be found in the data used in the process of dictionary building or learning the encoding neural network. With this interpretation, we should understand max pooling as a method to find out whether a particular pattern appeared in a given part of a song at all. If maximal value of i-th element over multiple vectors in a sequence is low, there was no vector that could be reconstructed using i-th dictionary element or a vector that would result in a strong activation of i-th neuron. Similarly, average pooling gives the information of a particular dictionary element (or a pattern detected by the neural network) appearing consistently over the course of the entire song.

One useful property of max pooling is that it eliminates the drawback of autoencoders possibly creating non-sparse representation of data in the strict sense. The notion of 'sparse' vectors in which most neurons are 'inactive', while not precise, is not a problem here because the low activation values of 'inactive' neurons are mostly lost in the process of pooling regardless of whether or not they equal 0.

VI. Experiments on Emotify Dataset

The main goal of our experiments was to compare the performance of different sparse coding methods in creating features for emotion regression task (subsection VI-B). Additionally we tested the influence of the pooling window size on the result (subsection VI-C. Our third experiment shows the spectrogram patterns recognized by sparse coding methods (subsection VI-D. At the beginning we present what is common across all performed experiments. The used tools and

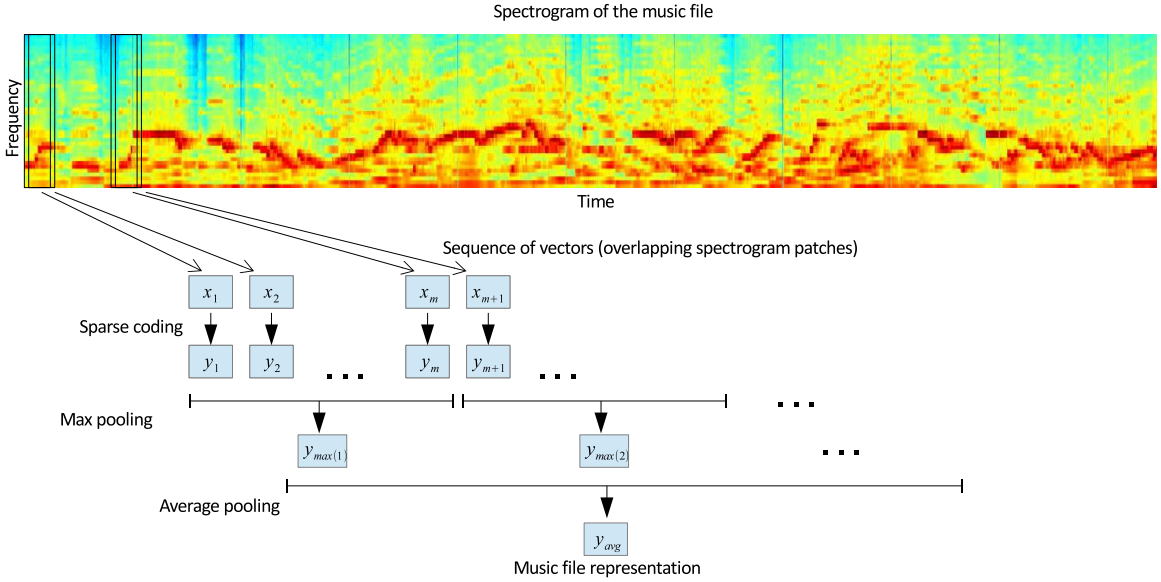


Fig. 1. Representation of a music file by pooling sparse vectors

the learning of sparse representation as well as the regression algorithm is presented in subsection VI-A.

A. Common conditions of the experiments

All performed experiments were performed using MATLAB/Octave toolboxes: MIR Toolbox [31] for spectrogram extraction from music files, Sparse Representation Toolbox [32] for the L1 regularized least squares implementation and DeeBNet [33] for the implementation of autoencoder and sparse RBM.

Sparse Representation Learning

Representation of music files we used in our experiments was based on spectrograms. Spectrogram of each file in the dataset was extracted using 50 ms frames with 25 ms overlap, using mel scale with $f = 40$ bins for frequency axis. The resulting values were logarithmized. Spectrogram patches were $l = 10$ frames long. This means vectors of length $fl = 400$ were used as an input for sparse coding algorithms.

We report the results for sparse representation size of $k = 200$ to emphasize tested methods' capabilities to build compact representations of musical files. In comparison, the best-performing feature set in previous research on the Emotify dataset consisted of 6535 features [18]. Testing sparse representation sizes 200, 500 and 1000 we found that increasing the vector size beyond 200 does not improve the performance significantly. For efficiency, both dictionary building and neural network training were performed on a randomly selected set of 40000 spectrogram patches from the whole dataset.

For kernelized version of L1-regularized least squares algorithm, we use radial basis function kernel defined as:

$$K(x, y) = \exp(-\gamma \|x - y\|^2) \tag{9}$$

Scaling parameter λ, which governs the weight of sparsity constraints relatively to reconstruction accuracy, is present in all methods except K-means. For this parameter we considered values of $\{0.01, 0.1, 1, 10, 100\}$. For parameter ρ, the desired average activation of neurons in sparse autoencoder and sparse RBM, values $\{0.01, 0.05, 0.1, 0.2, 0.5\}$ were considered. These parameters were optimized using grid search, and we report the best results.

Regression algorithm

We approached music emotion recognition as a regression problem: the input was the representation of the file calculated by pooling sparse vectors, i.e. a vector with 200 elements, and the value we were attempting to predict was a real number within [0,1] range, corresponding to the level of agreement that the music file evokes a certain emotion in the listener, as described in section III. A separate Support Vector Regression (SVR) [34] model was trained for each emotion. SVR is a model in which data is fit by a hyperplane, however, due to the possibility of using kernel trick one can accommodate for non-linear data. In our experiments we found that radial basis function kernel, defined as in (9), performs the best.

B. Experiment 1: Performance of different sparse coding methods

Obtained results are collected in Table 1. In order to compare them with [18], which is the only MER paper that reported results of research on Emotify dataset so far, we use the same performance measure, namely Pearson's correlation coefficient R. We compare proposed sparse coding methods to the results achieved using a feature set consisting of features available in MIRToolbox and harmonic features based on chord detection and interval detection.

TABLE I
SPARSE CODING METHODS COMPARED WITH A HIGHER-LEVEL FEATURE SET (PERFORMANCE MEASURED BY PEARSON'S R)

	K-means	L1 Least Squares	Kernel L1LS	Autoencoder	Sparse RBM	MIRtoolbox+Harmonic [18]
Amazement	0.27	0.20	0.18	**0.29**	0.21	0.16
Solemnity	0.41	0.43	0.36	**0.50**	0.48	0.43
Tenderness	0.30	0.46	0.38	0.54	0.49	**0.57**
Nostalgia	0.47	0.40	0.37	**0.50**	0.40	0.45
Calmness	0.47	0.47	0.45	**0.56**	0.53	0.50
Power	0.47	0.46	0.44	0.53	0.50	**0.56**
Joyful activation	0.48	0.50	0.50	0.53	0.54	**0.66**
Tension	0.32	0.47	0.26	**0.48**	0.43	0.46
Sadness	0.30	0.32	0.27	0.33	0.27	**0.42**
Average	0.39	0.43	0.37	**0.47**	0.43	**0.47**

For pooling, we first use max pooling over fragments of music file corresponding to 100 spectrogram patches (around 3 seconds), and then average pooling over the resulting sequence. Results were measured using 10-fold cross-validation.

It can be seen that on average, a sparse coding based approach can achieve performance comparable to that of hand-crafted features. However, autoencoder-based music representation outperforms hand-crafted features features in modeling highly subjective emotions, such as amazement and solemnity. At the same time, the results are significantly worse for "simple" emotions highly correlated with occurrence of major and minor chords, i.e. joyful activation and sadness.

Out of all examined methods, only autoencoders achieved this level of performance. The addition of radial basis function kernel to the L1 least squares not only did not improve the results, but worsened them. Sparse RBM performs better than most approaches, but does not achieve the performance of a sparse autoencoder. Interestingly, K-means based approach, equivalent to the simple bag of words model, achieves performance comparable to autoencoder. However, it does so only for specific emotions, while being significantly worse for others.

Under-performing in terms of recognizing joyful activation and sadness can be interpreted as a sign of difficulty in learning harmonic features. While an unsupervised dictionary learning method can recognize a pattern corresponding to a particular chord, it may be hard to generalize them to the concept of major and minor chords, and further to key detection for the entire song.

C. Experiment 2: Influence of pooling window size on the results

Using previous parameters and choosing autoencoder as the best performing method of sparse coding, we performed a second experiments with the goal of estimating the influence of selected pooling window size on the prediction of specific emotions. Usually, in music information retrieval, model parameters such as window sizes for any type of operation are constant and tuned for best overall performance. However, with a specific set of differing emotions, it is interesting to see how recognition of each emotion is affected by the time-frame considered.

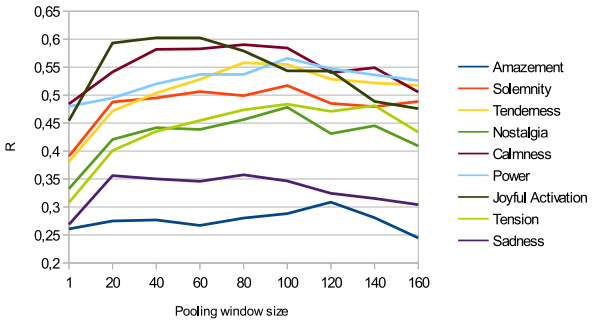


Fig. 2. Performance depending on pooling window size

As in the experiment 1, we use max pooling over small windows first, and average pooling over the resulting sequence of vectors afterwards. We consider pooling window sizes of $\{1, 20, 40, 60, 80, 100, 120, 140, 160\}$. Note that window size equal to one is equivalent to average pooling over the entire file, which is commonly used in dictionary learning. Similarly to the main experiment, results are measured using 10-fold cross-validation.

Fig. 2 shows the changes in performance for different window sizes of max pooling. It can be seen that peaks in performance appear at different window sizes for different emotions. For detection of joyful activation and sadness small window sizes seem optimal, while detection of tension can be improved with sizes close to 140 vectors.

D. Experiment 3: Encoding patterns

Since our sparse representations are built from spectrogram patches and spectrogram is a visual representation of how sound evolves, it is possible to visualize spectrogram patterns which sparse coding methods recognize. In case of autoencoder neural network, a response of a hidden layer neuron depends on the vector of input-hidden weights associated

with that particular neuron. For each neuron in the hidden layer, there are 400 weights, each corresponding to one input neuron. Since a vector given as an input to the neural network represents a 40×10 spectrogram patch, we use the same 40×10 dimensions to visualize the input weights. We want to see if the response patterns are possible to interpret by a human observer and relate to concepts existing in music. For visualization of encoding patterns, we learned an autoencoder network with the best-performing set of parameters found in previous experiments: $\lambda = 10$, $\rho = 0.1$.

Fig. 3 shows sample response patterns of single neurons in the encoding layer, with the highest weight values coloured red and lowest coloured blue. Neuron should respond with a high activation to spectrogram patches similar to these patterns.

We found certain types of response patterns appearing consistently in unsupervised training of autoencoders. In Fig. 3, leftmost pattern shows high weights in distinct narrow frequency bands, responding in high activation when the amplitude is high in these bands. Patterns like this respond well to specific music intervals, which makes them useful in extracting harmonic features. A combination of them could approximate chord detection. Second leftmost pattern shows an example of high weights in a wider frequency band, in this case responding to high amplitude in low frequency bins. Neurons responding to patterns like this for different wide bands can enable detection of overall amount high, mid and low tones, which in turn could be used e.g. in recognizing songs with a heavy bass line. Third pattern shows small weights for inputs corresponding to initial frames and high weights for inputs corresponding to latter frames in the low frequency band, indicating the neuron will respond with high activation to an onset of a musical phrase. Rightmost pattern can detect increases in amplitude separated by a specific interval of time, which gives the possibility of beat frequency detection.

We can see that through unsupervised training, autoencoders could learn features that are interpretable using concepts related to high-level features of music.

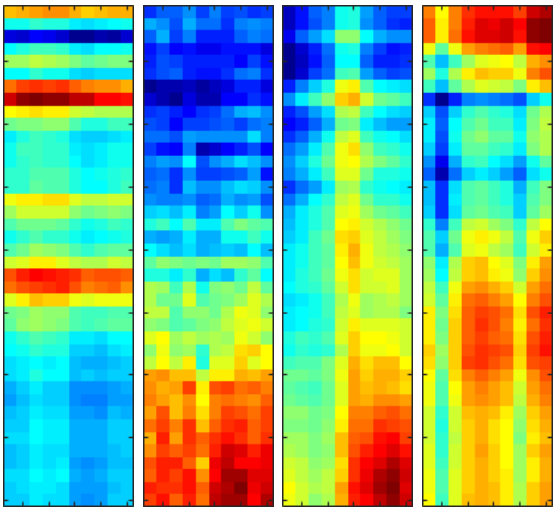


Fig. 3. Encoding patterns: input weights of four sample encoding layer neurons

VII. Conclusions and future work

We applied a machine learning approach based on pooling of sparse vectors built from spectrograms of sound files to Emotify, first publicly available music emotion recognition dataset annotated with Geneva Emotional Music Scale categories. We compared 5 sparse coding methods and measured their performance in the task of predicting a community consensus concerning emotions induced by a music piece. We found out that sparse autoencoders significantly outperform other approaches.

The results show that a sparse coding approach based on autoencoders can achieve satisfying performance in recognizing certain very subjective emotions hard to detect using higher level features. On the other hand even the best sparse coding method applied to this problem cannot outperform harmonic features in recognizing joyful activation and sadness. However, the results show that performance is significantly affected by the choice of the size of pooling window and it appears that different window sizes are optimal for detecting different emotions. These results suggest that analysing sound using fixed and constant time scales is a significant limiting factor in predicting music-induced emotions.

The main limitation of our research was the dataset, which unfortunately is the only one so far using the GEMS system for emotion classification. Generalizations based on our results should be confirmed on other datasets, possibly including more varied music types.

Future research in the area of induced emotion modelling can focus on a number subjects. First, a thorough examination of properties of the sparse autoencoder in the context of music emotion recognition can be useful. Second possible area of research is developing methods that can simultaneously consider multiple time-scales since there is no optimal selection of time windows for pooling, and maximizing average prediction performance does not guarantee maximizing performance for every emotion. Finally, building new datasets annotated with domain-specific emotional models such as GEMS should be a subject of interest.

References

[1] Y. E. Kim, E. M. Schmidt, R. Migneco, B. G. Morton, P. Richardson, J. Scott, J. A. Speck, and D. Turnbul, "Music emotion recognition: a state of the art review," *Proceedings of the 11th International Society for Music Information Retrieval Conference (ISMIR 2010)*, 2010.

[2] K. R. Scherer, M. Zentner, "Emotion effects of music: Production rules", in: *Music and emotion: Theory and research*, pp. 361-392, Oxford University Press, 2001.

[3] Qin, Zengchang et all, "A Bag-of-Tones Model with MFCC Features for Musical Genre Classification," in: Advanced Data Mining and Applications: 9th International Conference Proceedings, Part I. ADMA 2013, p. 564–575, Springer Berlin Heidelberg, 2013.

[4] E. J. Humphrey, J. P. Bello, and Y. LeCun, "Moving beyond feature design: Deep architectures and automatic feature learning in music informatics," in *Proceedings of the 13th International Conference on Music Information Retrieval (ISMIR)*, 2012.

[5] M. Henaff, K. Jarrett, K. Kavukcuoglu, and Y. LeCun, "Unsupervised learning of sparse features for scalable audio classification," in *Proceedings of the 12th International Conference on Music Information Retrieval (ISMIR)*, 2011.
[6] S. Sigtia and S. Dixon, "Improved music feature learning with deep neural networks," in *Proceedings of the 38th International Conference on Acoustics, Speech, and Signal Processing* (ICASSP), 2014.
[7] Y. Vaizman, B. McFee, and G. Lanckriet, "Codebook based audio feature representation for music information retrieval," *IEEE Transactions on Acoustics, Speech and Signal Processing*, 2014.
[8] J. Nam, J. Herrera, M. Slaney, and J. Smith, "Learning sparse feature representations for music annotation and retrieval," in *Proc. ISMIR,* 2012.
[9] T. Li and M. Ogihara, "Detecting emotion in music," in *Proc.of the Intl. Conf. on Music Information Retrieval*, Baltimore, MD, October 2003.
[10] J. Skowronek, M. McKinney, and S. van de Par, "A demonstrator for automatic music mood estimation," in *Proc. Intl. Conf. on Music Information Retrieval*, Vienna, Austria, 2007.
[11] C. Laurier, O. Lartillot, T. Eerola, and P. Toiviainen: "Exploring Relationships between Audio Features and Emotion in Music," *Conference of European Society for the Cognitive Sciences of Music*, 2009.
[12] Y. H. Yang, Y. C. Lin, Y. F. Su, and H. H. Chen, "A Regression Approach to Music Emotion Recognition," *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 16, No. 2, pp. 448-457, 2008.
[13] P. Ekman, "An argument for basic emotions," *Cognition Emotion* vol. 6, no. 3, pp. 169-200, 2001.
[14] K. Hevner, "Experimental studies of the elements of expression in music," *American Journal of Psychology*, vol. 48, pp. 248 -268, 1936.
[15] Y. E. Kim, E. M. Schmidt, R. Migneco, B. G. Morton, P. Richardson, J. Scott, J. A. Speck, and D. Turnbul, "Music emotion recognition: a state of the art review," *11th International Society for Music Information Retrieval Conference (ISMIR 2010)*, 2010.
[16] U. Schimmack and R. Reisenzein, "Experiencing activation: energetic arousal and tense arousal are not mixtures of valence and activation," *Emotion*, vol. 2, no. 4, p. 412, 2002.
[17] M. Zentner, D. Grandjean, and K. R. Scherer: "Emotions evoked by the sound of music: characterization, classification, and measurement," *Emotion*, vol. 8, no. 4, pp. 494-521, 2008.
[18] A. Aljanaki, F. Wiering, and R. Veltkamp, "Computational modeling of induced emotion using GEMS," *Proceedings of the 15th Conference of the International Society for Music Information Retrieval (ISMIR 2014)*, pp. 373-378, 2014.
[19] G. E. Hinton, "A practical guide to training restricted Boltzmann machines," Tech. Rep. UTML TR 2010-003, Dept. Comput. Sci., Univ. Toronto, 2010.
[20] N.J. Nalini, S. Palanivel, "Emotion Recognition in Music Signal using AANN and SVM," *International Journal of Computer Applications* vol. 77, no.2, 2013.
[21] N. Glazyrin , "Mid-level features for audio chord recognition using a deep neural network," *Uchenye Zapiski Kazanskogo Universiteta. Seriya Fiziko-Matematicheskie Nauki*, vol. 155, no. 4, pp. 109–117, 2013.
[22] A. Aljanaki, D. Bountouridis, J.A. Burgoyne, J. van Balen, F. Wiering, H. Honing, and R. C. Veltkamp, "Designing Games with a Purpose for Data Collection in Music Research. Emotify and Hooked: Two Case Studies," *Proceedings of Games and Learning Alliance Conference,* 2013.
[23] B. Logan, "Mel frequency cepstral coefficients for music modeling," in *Proc. of the Intl. Symposium on Music Information Retrieval*, Plymouth, MA, 2000.
[24] J. B. MacQueen, "Some Methods for classification and Analysis of Multivariate Observations," *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 1967.
[25] F. Bach, R. Jenatton, J. Mairal, G. Obozinski, "Optimization with Sparsity-Inducing Penalties," *Foundations and Trends in Machine Learning* vol. 4, no. 1, 2012.
[26] S. Gao, I. W. H. Tsang, L. T. Chia, "Sparse representation with kernels," in *IEEE Transactions on Image Processing*, vol. 22, no. 2, pp. 423-434, 2012.
[27] Y. Bengio, "Learning Deep Architectures for AI", *Foundations and Trends in Machine Learning,* vol. 2, no. 1, 2009.
[28] D. E. Rumelhart, G. E. Hinton, Williams, Ronald J. . "Learning representations by back-propagating errors", *Nature*, vol. 323, pp. 533-536, 1986.
[29] Q. V. Le, J. Ngiam, A. Coates, A. Lahiri, B. Prochnow, and A. Y. Ng, "On optimization methods for deep learning," in *Proc. 28th Int. Conf. Machine Learning,* pp. 265–272, 2011.
[30] G. E. Hinton, "Training products of experts by minimizing contrastive divergence," *Neural Computing*, vol. 14, pp. 1771-1800, 2002.
[31] Olivier Lartillot, Petri Toiviainen, "A Matlab Toolbox for Musical Feature Extraction From Audio," *International Conference on Digital Audio Effects,* Bordeaux, 2007.
[32] Y. Li "Sparse representation for high-dimensional data analysis," in *Sparse Machine Learning Models in Bioinformatics,* PhD Thesis, School of Computer Science, University of Windsor, Canada, 2013.
[33] M. A. Keyvanrad and M. M. Homayounpour, "A brief survey on deep belief networks and introducing a new object oriented toolbox (DeeBNet)," arXiv:1408.3264 [cs], Aug. 2014.
[34] A. J. Smola, B. Scholkopf, "A tutorial on support vector regression," *Statistics and Computing,* vol. 14, no. 3 , pp 199-222, 2004.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 61–70
DOI: 10.15439/2016F351
ACSIS, Vol. 8. ISSN 2300-5963

A General Method of the Hybrid Controller Construction for Temporal Planning with Preferences

Krystian Adam Jobczyk
University of Caen
Laboratoire de Automatique, Informatique, Instrumentation et Image
Marechal Juin 6, 14000 Caen, cedex 5
E-mail: krystian.jobczyk@unicaen.fr

Antoni Ligęza
AGH University of Science and Technology
Department of Applied Computer Science
al. Mickiewicza 30, 30-059 Krakow
E-mail: ligeza@agh.edu.pl

***Abstract*—This paper is aimed at presenting some general construction method of the hybrid plan controller for some task of temporal planning with preferences. This construction is multi-stage and it begins with a description of a chosen robot environment and its plan in some extended version of Linear Temporal Logic. This description is later transformed to the appropriate preferential Büchi automaton. In the same way, the real plan performing by the robot is encoded by the similar automaton. Finally, both automata are exploited to construct its product automaton, which is later described in** PROLOG**.**

***Index Terms*—the hybrid plan controller, temporal planning, preferences, the robot motion environment, PROLOG, automata, Linear Temporal Logic, Halpern-Shoham logic**

I. Introduction

A *Plan Controller* constitutes a machine (sometimes only abstract) suitable for controlling how different tasks are performed in comparison with the initially developed plans or schedules. The plan controller construction is often a multi-stage activity and it begins with a description of the robot environment and tasks in the appropriate formal language. This description should be later translated into the appropriate automata which encode the initial part of information in terms of automaton states and admitted transitions between them. Such controllers[1] satisfy many different features, and – depending on the proposed approach to the robot's motion representation – are usually rendered in terms of Linear Temporal Logic (LTL), but may be also rendered in terms of the motion description language [9] or of the control and computation language [14].

An interesting approach to the plan controller construction was presented in [4], [20], which is based, however, on the known idea of the environment triangulation. The classical approach to the controller construction leads just from a triangulation of the robot's/agent's environment and the appropriate finite transition system by a representation of this environment in terms of LTL. In the next step, the LTL-formulas are represented by the appropriate Büchi automaton[2]. Its construction is later complemented by the construction of the appropriate product automaton for LTL-specification and for the considered transition system – suitable for a grasping of the basic dynamism in the robot's environment. This product automaton ensures acceptance of certain desired transitions only, see: [4], and it forms a 'core' of the hybrid plan controller representation.

The method of the Büchi automaton construction alone for a use of different planning situations was presented in [6] and – w.r.t. formulas of some extended LTL — earlier in [21], [22]. The applicability of the formalism of temporal logic for specification of the robot's motion environment is a commonly known and widely discussed fact; see for example: [1], [2]. Indeed, such systems as Linear Temporal Logic (LTL) and Computation Tree Logic (CTL) have a satisfying expressive power to give a relatively detailed description of components of robot's motion planning such as action sequencing or objectives, the properties of the robot's environments. This utility of LTL was suggestively demonstrated in [2]. The robot's environments specification in terms of temporal logic constitutes the first fundamental step for planner and plan controller construction.

We intend to extend these recent approaches in two ways: we introduce an additional preferential component to the considered transition system and we extend the language for the robot's environment specification by introducing Halpern-Shoham logic (HS) – introduced in [7] – with single operators $\langle D \rangle$ and $\langle L \rangle$ for relations: 'during' and 'later' (*resp.*). Such a radical restriction of HS is dictated by formal requirements of the conversion to the automaton, which can be realized in this case.

Motivation. Majority of approaches to plan controlling of the robot behavior in polygonal environment – such as the presented in [1], [2], [4], [20] – discuss this issue rather in a form of extended outlines and without (often needed) technical details. By contrast, some formal papers such as [21], [22],

[1]They are often called "hybrid controllers" as they join different automaton types and they refer to different features and 'entities' such as plans and the motion environements.

[2]This useful formal tool was invented in 1962 and described in [**?**].

[6] are aimed at presenting a purely meta-logical face of the construction of automata in a language of temporal logic. In result, there is no real coincidence between these approaches. Moreover, no of them take into account preferential aspects of planning, which introduce a portion of 'rationality' to performing tasks and, its management and controlling.

These lacks constitute a main motivation factor of this paper investigations. The next motivation factor stems from an additional lack of some correlation between research on the real expressive power of temporal logic systems – such as Halpern-Shoham logic in [18], [19] – and some engineering tendency to exploit temporal logic systems in (almost) all possible contexts and without restrictions.

Against this state of the art, we intend to propose such a new preferential extension of the current approaches to the plan controller construction for a robot in polygonal environment which respects last arrangements about temporal logic and its expressive power. This intention determines a choice of the appropriate subsystems of Halpern-Shoham logic: its restriction to the modal operator $\langle D\rangle$ and $\langle L\rangle$ for representation "during' and 'later'-relations (*resp.*). In fact, these operators can be effectively transformed to the appropriate automata in the light of recent observations from [19], [17]. Some conceptual background of the automaton construction for combined formulas of some multi-valued extension of HS-logic has been presented in [13], [10].

Moreover, the authors of this paper give venture of their enthusiasm with respect to some utility of the proposed construction in different areas of utility of temporal logic systems: in engineering or – for example – in business processes and their management. Some applicability of temporal logic systems for engineering has been recently discussed in [12], [11], for a use of business management – in [15], [16]. Last, but not least, expected future implementation of the automaton construction in languages of a declarative paradigm such as PROLOG or ASP forms some additional motivating factor of this paper's analysis.

Objectives of the paper. According to these motivation factors, rendered above, objectives of this paper – in a chronological order – are the following :

1) proposing a new preferential extension of the concept of the hybrid plan controller– based on product automata,
2) construction of hybrid plan controller for a robot performing task in a polygonal environment in the block world,
3) an outline of the PROLOG-representation for some fragments of the constructed plan controller.

We also associate to these main objective some additional goal to extend the used specification language from LTL to LTL extended by Halpern-Shoham logic with $\langle D\rangle$ and $\langle L\rangle$-operators, symb. HS^D.

Organization of the paper. This paper is organized as follows. In section 2 we present a terminological background of the analysis. In section 3 we present the main problem of the paper analysis and we give a general algorithm of our hybrid controller specification. Section 4 forms the main conceptual part of the paper, where we describe in details the steps of the controller construction *via* the Büchi automaton for LTL and for the considered transition system. In section 5 we briefly describe the implementation area of this construction and HS logic with operators: $AB\bar{A}\bar{B}$. In section 6 we formulate conclusions and some remarks on the future research direction.

II. Preliminaries

Before moving to main paper body, we present a terminological framework of this paper analysis, introducing a new concept of *preferential automata* and *preferential transition system*. The recalled definitions of a (finite) transition system and a Büchi automaton are incorporated from [7].

Definition 1. . *A (finite) transition system FTS is a n-tuple $FTS = (W, W_0, Act, Tran, \Pi, Obs)$, where:*

1) *W is the finite set of states (worlds),*
2) *$W_0 \subseteq W$ is the distinguished set of initial states,*
3) *Act denotes the set of possible actions,*
4) *$Tran : W \times Act \mapsto W$ is a transition function, i.e such a total function that returns the next stage for a given state an an action,*
5) *Π denotes the set of possible observations,*
6) *$Obs : W \mapsto \Pi$ is the observability function, which returns the observable part of the current state.*

We define an *execution* on FTS as an infinite sequence of states $w_0, w_1, \ldots$, such that $w_0 \in W_0$ and $w_{k+1} = Tran(w_k, a)$ for some action $a \in Act$. The observable part of the execution will be called a *trace*.

Definition 2. *A Büchi automaton is a tuple $A = (\Sigma, S, S_0, \rightarrow, \rho, \mathcal{F})$, where:*

1) *Σ is the alphabet of the automaton,*
2) *S is the set of states of the automaton,*
3) *$S_0 \subseteq S$ is the set of initial states of the automaton,*
4) *$\rho : S \times \Sigma \mapsto 2^S$ is the transition function of the automaton and*
5) *$\mathcal{F}$ is the set of accepting words.*

We expand this definition to a definition of preferential Büchi automaton by specification of the set of accepting words by introducing some degrees/parameters α's from an interval $[0, 1]$. The role of them is to measure a *degree of a preference* of the accepting words from F, indexed by such an α.

Definition 3. *A A Preferential Büchi automaton is a tuple $A = (\Sigma, S, S_0, \rightarrow, \rho, \mathcal{F}, \alpha_1, \alpha_2 \ldots)$, where:*

1) *Σ is the alphabet of automaton,*
2) *S is the set of states of automaton,*
3) *$S_0 \subseteq S$ is the set of initial states of automaton,*
4) *$\rho : S \times \Sigma \mapsto 2^S$ is the transition function of automaton and*
5) *$\mathcal{F}^\alpha$ is the set of accepting words with associated α-degree preferences.*

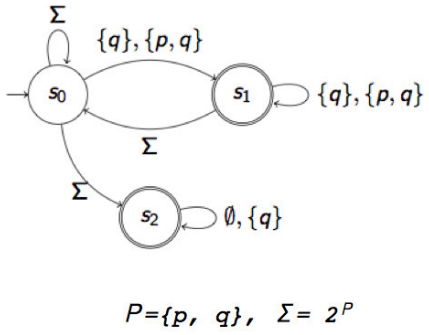


Fig. 1. Fragment of the Büchi automaton with states s_0, s_1 and s_2 over an alphabet Σ.

6) $\alpha_1, \alpha_2, \ldots$ *are values of functions* $f :\mapsto \mathcal{F}^\alpha \rightarrow [0,1]$ *called preferences.*

We assume that a set of α degrees is finite as they index accepting words from set F^α. Naturally, F^α should be finite as a a set of accepting words of a finite automaton.

For such a defined automaton we define a *run* of $\mathcal{A}$ (on an infinite words $a_0, a_1, a_2, \ldots$) is an infinite sequence of such states $s_0, s_1, \ldots \in S^\omega$ that $s_0 \in S_0$ and $s_{i+1} \in \rho(s_i, a)$. We say that a run r is *accepting* iff a set $\{s|s$ occurs in r infinitely often $\} \cap F \neq \emptyset$. If F is finite, this general condition means that there exists at least one state s that occurs in a run r infinitely often.

Linear Temporal Logic (LTL).

Syntax. Bi-modal language of LTL is obtained from standard propositional language (with the Boolean constant $\top$) by adding temporal-modal operators such as: *always in a past* (H), *always in a future* (G), *eventually in the past* (P), *eventually in the future* (F), *next and until* ($\mathcal{U}$) and *since* ($\mathcal{S}$) – co-definable with "until". The set FOR of LTL-formulas is given as follows:

$$\phi := \phi|\neg\phi|\phi \vee \psi|\phi\mathcal{U}\psi|\phi\mathcal{S}\psi|H\phi|P\phi|F\phi|Next(\phi) \qquad (1)$$

Some of the above operators of temporal-modal types are together co-definable as follows: $F\phi = \top\mathcal{U}$, $P\phi = \top\mathcal{S}\phi$ and classically: $F\phi = \neg G\phi$ and $P\phi = \neg H\phi$.

Semantics. LTL is traditionally interpreted in models based on the point-wise time-flow frames $\mathcal{F} = \langle T, <\rangle$ and dependently on a set of states S. In result, we consider pairs (t, s) (for $t \in T$ representing a time point and $s \in S$) as states of LTL-models. Anyhow, we often consider a function $f : T \mapsto S$ that associates a time-point $t \in T$ with some state $s \in S$ and we deal with pairs (t, f) instead of (t, s). Hence the satisfaction relation $\models$ is defined as follows:

1) $(t,f) \models G\phi \iff (\forall t^{'} > t)t^{'} \models \phi$, $(t,f) \models H\phi \iff (\forall t < t^{'})t^{'} \models \phi$.
2) $(t,f) \models F\phi \iff (\exists t^{'} > t)t^{'} \models \phi$, $(t,f) \models P\phi \iff (\exists t < t^{'})t^{'} \models \phi$.
3) $(t_1,f) \models \phi\mathcal{S}\psi \iff$ there is $t_2 < t_1$ such that $t_2, f \models \psi$ and $t, f \models \phi$ for all $t \in (t_1, t_2)$

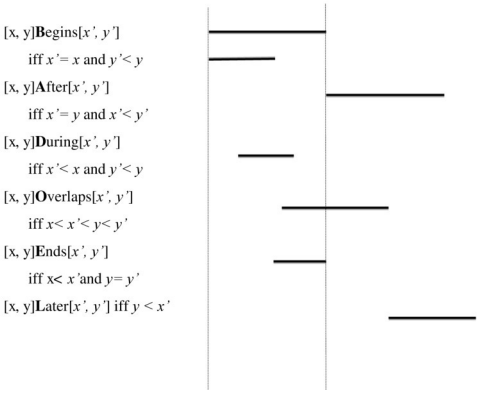


Fig. 2. Visual presentation of temporal interval relations of Allen

4) $(t_1,f) \models \phi\mathcal{U}\psi \iff$ there is $t_2 > t_1$ such that $t_2, f \models \psi$ and $t, f \models \phi$ for all $t \in (t_1, t_2)$
5) $(t_k,f) \models Next(\phi) \iff (t_{k+1}, f) \models \phi, k \in \mathcal{N}$.

Halpern-Shoham logic. HS forms a modal representation of the following temporal relation between intervals, defined by [7]: **a**fter" (or **m**eets"), (**l**ater"), **b**egins" (or **s**tart"), **d**uring", **e**nd" and **o**verlap". These relations are intuitive and their visualization can be easily found in many papers, so we omit their visual presentation, they correspond to the modal operators: $\langle A\rangle$ for **a**fter", $\langle B\rangle$ for **b**egins", $\langle D\rangle$ for **d**uring", etc. The syntax of HS entities ϕ is defined by:

$$\phi := p|\neg\phi|\phi \wedge \phi|\langle X\rangle|\langle \bar{X}\rangle\,, \qquad (2)$$

where p is a propositional variable and $\langle \bar{X}\rangle$ denotes a modal operator for the inverse relation with respect to $X \in \{A, B, D, E, O, L\}$ If $\phi \in \mathcal{L}$(HS), M is a model, and I is an interval in the M-domain, then the satisfaction for the HS-operators looks as follows:

$$M, I \models \langle X\rangle\phi \iff \exists I^{'} that\ IXI'\ and\ M, I^{'} \models \phi. \qquad (3)$$

Example 1. *Some simple spatial-temporal requirements imposed on the robot's environment E can be expressed by formulas:*

- *Always if you take a block A, take B, as well:* $G(take(A) \rightarrow take(B))$,
- *For any intervals: if you take A, than you also put B:* $[D](HOLDS(take(A)) \rightarrow HOLDS(put(B)))$.

III. Problem formulation and a general algorithm of the controller construction

Assume that E is a polygonal environment of robot motion operations. All possible admitted holes of E have to be enclosed by a single polygonal chain. The motion of robot is expressed as follows:

$$x(t) \in E \subseteq R^2, u(t) \in U \subseteq R^2, u(t) \cap x(t) \neq \emptyset \qquad (4)$$

where $x(t)$ is a trajectory of robot's motion (position of a robot in a time t) in E and $u(t)$ is a control input. Non-emptiness of the above intersection $u(t)$ and $x(t)$ ensures that a controller detects the robot's trajectory. In such a framework, the goal of the paper is to give an outline of a construction of a hybrid controller that generates controllers inputs $u(t)$ for a trajectory $x(t)$ and environment with a specification given by formulas of LTL and – partially – by HS restricted to D-operator.

A general path of our controller construction looks a follows. We begin with the environment E and its triangulation. Secondly, we consider some transition system FTS to describe a basic dynamism of E. The next, we specify E in terms of LTL (ϕ-formula) and of some subsystem of HS logic. In the next construction step, we transform FTS to the appropriate Büchi automaton $\mathcal{A}_{FTS}$ for it. The similar automaton $\mathcal{A}_{LTL,HS}$ is constructed for representation of a specification of E (with a chosen point x_0) in terms of the considered temporal logic. Having these automaton, we construct some product automaton $\mathcal{A}$ to 'reconcile' the activity of both automata. Assume that some environment E of a robot and a formula $\phi \in \mathcal{L}(LTL \cup HS^{D,L})$ – describing this environment or the robot motion are given. Thus, the algorithm of the hybrid controller construction could be given as follow – as a specified version of algorithm from [4]:

Algorithm: The Hybrid Controller Construction

Procedure: CONTROLLER(E, ϕ)

1) $\triangle \leftarrow Triangulate(E)$
2) $FTS \leftarrow TriangulationToFTS(\triangle)$
3) $\mathcal{A}_{FTS} \leftarrow FTS\ to\ Bchi\ Automaton$
4) $\mathcal{A}_{LTL,HS^D} \leftarrow LTL \cup HS^D to\ Bchi\ Automaton$
5) $\mathcal{A} \leftarrow Product \mathcal{A}_{FTS}, \mathcal{A}_{LTL,HS^D}$
6) **return**: $Controller(\mathcal{A}, \triangle, \phi)$

End procedure

IV. Controller construction for path temporal planning with observability

A. From triangulation to the FTS system

In order to propose an exact construction of our path temporal planning controller we will represent a polygonal environment E as a finite set of partitions. One can use many methods of the initial polygonal environment's decomposition, presented in [4], [5].

The main idea of such a triangulation consists in a mapping of each point $x \in E$ to a one of the disjoint equivalence classes determined by an equivalence relation $\sim$. The natural way is to define $\sim$ as follows: $\forall x, y \in E : x \sim y \iff x, y \in \triangle$, i.e each of such an equivalence class forms a triangle, what allows us to represent a quotient set $E| \sim$ as a sum of triangles. Assume now that $T : E \to Q$ for $Q = \{\triangle_1, \ldots \triangle_k\}$. Than each $T^{-1}(\triangle_i)$, for $i \in \{1, 2, \ldots, n\}$ contains some states $x \in E$ and a set $\{T^{-1}(\triangle_i) | \triangle_i \in Q$, for $i \in \{1, 2, \ldots, n\}\}$ of all such triangle anti-images is a desired partition of the initial motion environment E. In order to preserve a consideration generality, we do not impose any special requirements on E (concerning the temporal or spatial coverage etc.)

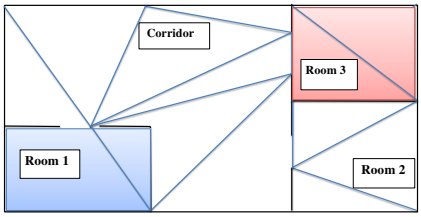


Fig. 3. An example of a triangulation of some polygonal robot environment

In this framework we can introduce a *finite transition system with preferences*– FTS, modifying a definition of of FST with observability from [4], [20] as follows:

Definition 4. *We define the finite transition system FTS with preferences as a tuple* $FTS = \langle Q, Q_0, Act, tran_{FTS}, Pref, P \rangle$, *where:*

1) Q *is the finite set of states (triangles),*
2) $Q_0 \subseteq Q$ *is the set of initial robot states,*
3) Act *is the set of actions,*
4) $tran_{FTS} : Q \times Act \mapsto Q$ *is defined as a 'move' from* $\triangle_i \to \triangle_j$ *iff the cells* $T^{-1}(\triangle_i)$ *and* $T^{-1}(\triangle_j)$ *share a common edge,*
5) P *is the set of objects called preferences,* $Pref : Q \to P$ *is a preferential map which associates to a* $\triangle_1 \in Q$ *some preference* $pref$ *such that* $pref(\triangle_1) \in P$ *if* $P \subset [0, 1]$ *and* $Pref$ *is a function.*

Example 2. *One can consider such a transition system with preferences as a system –depicted on a Fig. 2 with a function PREF, which satisfies a condition:* $PREF(\triangle(room3)) = \frac{1}{2}$. *(All triangles from a room 3 are preferable (in a sense of a robot task as places to visit) with a degree* $\frac{1}{2}$*).*

B. From FTS system to the Büchi automaton

It easy to observe that the finite transition system FTS naturally models a basic structure of the motion environment. It can play this role independently of a way of its presentation – in the standard form: $FTS = (W, W_0, Act, Tran, \Pi, Obs)$ or in the "triangle" tuple $FTS = \langle Q, Q_0, Act, tran_{FTS}, Pref, P \rangle$. Moreover, their structure – as it has been said – is similar to a structure of automata, so they seems to be a natural base of a construction of the required automata. In essence, (finite) automata may be considered as (finite) transition systems with labeling functions, connected with the appropriate language which delivers an alphabet as a required component of the automaton structure. Due to some practice in this area – expressed in [4], [20] – we will consider the standard repre-

sentation of FTS as more suitable for an expected automaton construction.

Authors of the approach from [4], [20] recommended to an immediate use a structure of such an FTS for this construction. We only partially make use of this advise. Of course, we will base the automata construction on the FTS-structure with its states as transitions as states and transitions of the newly constructed automaton. Nevertheless, we decide for a more"descriptive" solution: we firstly describe a robot environment (FTS) in terms of LTL and we construct states of desired automaton from sets of the used formulas. For that reason, one need now a mechanism of such a description. Its presentation will be given below.

C. From LTL and $HS^{L,D}$ to its Büchi automaton

The next step of our construction will consist in a constructing of the appropriate Büchi automaton for LTL- and HS-formulas expressing the robot's motion environment E and the action sequencing. This situation is, however, not so comfortable as in the FTS-case for three purposes. First, both LTL and HS form formal languages, so their translation for the Büchi automaton should be more sophisticated. Moreover, we deal with two languages: LTL and HS of a different temporal and a modal-temporal nature. Last, but not least, HS logic (in generality) cannot be expressed by the Büchi automaton because of the enormous expressive power of this logic. Fortunately – as it was already mentioned – it was proven in [17] that the subsystem of HS with an operator D can be represented by a finite state automaton. Unfortunately, it is not clear whether there exists any extension of this subsystem with the same property. Nevertheless, it is known that a subsystem $A\bar{A}B\bar{B}$ is too strong, because ω-regular languages can be embedded in this system, but not in the inverse direction [19]. According to these remarks on the Büchi automaton construction in section 2, we begin with defining of a closure of formulas of LTL and HS^D.

Definition 5. *Let assume that $\phi \in \mathcal{L}$(LTL). We define its closure $cl(\phi)$ as follows:*

1) $\phi \in cl(\phi), \neg(\phi_1) \in cl(\phi)$, *than* $\phi_1 \in cl(\phi)$,
2) $\phi_1 \wedge \phi_2 \in cl(\phi)$ *than* $\phi_1, \phi_2 \in cl(\phi)$,
3) $A(\phi_1, \ldots, \phi_k) \in cl(\phi) \rightarrow A_{sub}(\phi_1, \ldots, \phi_k) \in cl(\phi)$ *for* A_{sub} *denoting a sub-formula of A.*

In the similar way we will define a closure for $\phi \in \mathcal{L}$(HS^D). For a distinction of these languages and their formulas we will denote them as: ϕ_{LTL} and ϕ^D_{HS}.

Definition 6. *In accordance with the earlier statements we define an automaton $\mathcal{A}_{LTL,HS^D}$ as n-tuple:*

$$\mathcal{A}_{LTL,HS^D} = (2^{cl(\phi_{LTL} \cup \phi^D_{HS})}, \rho, S_0^{\phi_{LTL}}, S_0^{\phi^D_{HS}}, \mathcal{F}) \quad (5)$$

where $(2^{cl(\phi_{LTL})}$ is the set of states of the automaton as the collection of all sums of (disjoint) sets of formulas in $cl(\phi_{LTL})$ and $cl(\phi^D_{HS})$, ρ is the transition and $S_0^{\phi_{LTL}}, S_0^{\phi^D_{HS}}$ are sets of initial states of automaton containing for ϕ_{LTL} and $\phi^{D,L}_{HS}$ and $F \subseteq 2^{cl(\phi_{LTL})} \cup 2^{cl(\phi^D_{HS})}$ is a set of accepting words of automaton[3]*. (resp.).*

One can expand this definition – based on [21], [22] to the definition of a preferential automaton as follows:

Definition 7. *A preferential automaton (for words of $\mathcal{L}(LTL \cup HS^{D,L})$) is defined as n-tuple:*

$$\mathcal{A}_{LTL,HS^D} = (2^{cl(\phi_{LTL} \cup \phi^{D,L}_{HS})}, \rho, S_0^{\phi_{LTL} \cup \phi^{D,L}_{HS}}, \mathcal{F}^\alpha, \alpha_1\alpha_2, \ldots) \quad (6)$$

where $(2^{cl(\phi_{LTL} \cup \phi^{D,L}_{HS})}$ is the set of states of the automaton as the collection of all sums of (disjoint) sets of formulas in $cl(\phi_{LTL})$ and $cl(\phi^D_{HS})$, ρ is the transition,

$S_0^{\phi_{LTL} \cup \phi^{D,L}_{HS}}$ are sets of initial states of automaton containing for ϕ_{LTL} and $\phi^{D,L}_{HS}$ and $\mathcal{F}^\alpha \subseteq 2^{cl(\phi_{LTL}) \cup \phi^D_{HS})}$ is a set of accepting words of automaton, and each of $\alpha_1, \alpha_2 \ldots$ is a function $f : \mathcal{F}^\alpha \mapsto [0, 1]$.

Example 3. *Let us consider $\phi = \neg\phi_1$ for some ϕ_1 as a LTL-formula and $\psi = \langle D\rangle\psi_1$ (for some ψ_1) as our HS^D-formula.We show how to construct a Büchi automaton in a case of these formulas.*

Due to definition of a closure of the formula we obtain: $cl(\phi) = \{\phi_1, \neg\phi_1\}$ and $cl(\psi) = \{\psi, \neg\psi, \langle D\rangle\psi, \neg\langle D\rangle\psi\}$ and thus $2^{cl(\phi)} = \{\emptyset, \{\phi_1\}, \{\neq \phi_1\}, \{\phi_1, \neg\phi_1\}\}$ and $2^{cl(\psi)} = \{\emptyset, \{\psi\}, \{\neg\psi\}, \{\psi, \langle D\rangle\psi\}, \{\psi, \neg\langle D\rangle\psi\}, \{\neg\psi, \langle D\rangle\psi\}, \{\neg\psi, \neg\langle D\rangle\psi\}\}$. $N_\psi = \{\neg\phi_1, \{\phi_1, \neg\phi_1\}\}$ (because these sets contain ϕ) and $N_\psi = \{\{\psi_1, \langle D\rangle\psi_1\}, \{\neg\psi_1, \langle D\rangle\psi_1\}$ (because these sets contain $\langle D\rangle\psi_1$). Transitions ρ_{LTL} and ρ_{HS^D} are defined in such a way that sets from N_ϕ are initial in ρ_{LTL} and N_ψ are initial for ρ_{HS^D}.

D. Product automaton $\mathcal{A}_{FTS} \times \mathcal{A}_{LTL,HS^D}$

We have already defined the automaton $\mathcal{A}_{FTS}$, which describes the finite transition system and the automaton $\mathcal{A}_{LTL,HS^D}$ that represents the initial temporal logic-based specification of the motion environment. There is a need to reconcile both automata in order to construct our open-loop hybrid controller. For this purpose, it seem to be reasonable to restrict a *spectrum* of the possible transitions to these of them, which can ensure some form of observability.

For this purpose we introduce a product automaton $\mathcal{A} = A_{FTS} \times \mathcal{A}_{LTL,HS^D}$ with a new transition $\rightarrow_{\mathcal{A}}$ between pairs: $(q_i, w_i) \rightarrow_{\mathcal{A}} (q_j, w_j)$. We assume that this transition holds between such pairs if and only if $q_i \rightarrow_{FTS} q_j$ and $(w_i; \pi(q_j)) \rightarrow_{LTL} w_j$. This last condition means that the last transition has an input that contains an action $\pi(q_j)$ being an observation of the newly achieved (in sense of $\rightarrow_{FTS}$) state q_j.

Definition 8. *Let $\mathcal{A}_1 = \langle \Sigma_1, S, S_0, \rightarrow_{A_1}, \mathcal{F}^\alpha, \alpha_1, \alpha_2 \ldots\rangle$ and $\mathcal{A}_2 = \langle \Sigma_2, T, T_0, \rightarrow_{A_2}, \mathcal{F}^\beta, \beta_1, \beta_2 \ldots\rangle$ are preferential*

[3]Let us observe that defining of $\mathcal{F}$ as accepted words determines, somehow, a set of final states by this way of state definition – just by formulas, what seems to justify $\mathcal{F}$ as a subset of $2^{cl(\phi_{LTL}) \cup cl(\phi^D_{HS})}$.

automata, than their preferential product automaton is the automaton of the form:

$$\langle \Sigma_1 \times \Sigma_2, S \times T, S_0 \times T_0, \rightarrow_{A_1 \times A_2}, \mathcal{F}^{\alpha} \times \mathcal{F}^{\beta}, \alpha_1, \alpha_2, \beta_1, \beta_2 \ldots \rangle, \tag{7}$$

where $\rightarrow$ is a product transition defined such that: $(s_i, t_i) \rightarrow (s_{i+k}, t_{i+k})$ holds for natural $i \in I$ and $1 \leq k$ if and only if $s_i \rightarrow_{A_1} s_{i+k}$ and $t_i \rightarrow_{A_2} t_{i+k}$ and $\alpha_1, \alpha_2, \beta_1, \beta_2 \in [0, 1]$ are fuzzy values expressing preference degrees of accepting words from $\mathcal{F}^{\alpha}$ and $\mathcal{F}^{\beta}$ (resp.)

(For simplicity, we will shortly write $\rightarrow$ instead of $\rightarrow_{A_1 \times A_2}$, when it will not lead to any confusion.)

E. Complementation of conditions for a product automaton

After the product automaton construction the design of an open-loop hybrid controller for motion planning reduces to the finding problem the accepting execution of this automaton. Nevertheless, this construction requires a small complementation. In fact, some components of $\mathcal{A}$ being singletons can have no outgoing transitions. In order to ensure a normal work of the automaton, we add the so-called stutter extension [8] rule, which adds a self-transition on the blocking states. More formally: for all states $s \in$ domain of $\mathcal{A}$ we define a new transition $\rightarrow_{\mathcal{A}^*} = \rightarrow_{\mathcal{A}} \cup (s \rightarrow_{\mathcal{A}} s)$, where $\rightarrow_{\mathcal{A}}$ is the transition of the automaton $\mathcal{A}$ [4], earlier defined. In such a framework it holds the following:

Theorem 1. *(adopted from [6]) An execution of FTS that satisfies the specification in terms of LTL and HS-formulas exists iff the language of A is non empty.*

We omit the proof details. For a case of LTL-specification it can be found in[7]. For a case of HS^D it follows from the existence of the automaton accepting formulas of this logic from [17].

V. Part II: Implementation

In last part of the paper we gave a theoretic outline of the hybrid controller construction – based on some product automaton. In addition, a description of the robot motion environment and the robot plan have been rendered in LTL extended by some fragment of HS-logic. In this part we intend to illustrate these ideas by proposing a concrete construction of such a controller. According to the earlier arrangements – this construction will be multi-stage and it will contain the following stages:

1) a presentation of the robot motion environment,
2) a formal description of the environment and the plan of the robot in terms of $LTL \bigcup HS^{D,L}$,
3) a formal description of the environment and a real plan performing by the robot in terms of $LTL \bigcup HS^{D,L}$
4) a construction of the appropriate Büchi automata for both the cases (the first one – for a desired plan, the second one – for a real plan performing by the robot) ,
5) a construction of a product automaton built up from automata from a point (4).
6) a PROLOG-description of the product automaton in order to detect eventual discrepancies between a plan and its performing by the robot.

[4]In essence we consider a projection of $\rightarrow_{\mathcal{A}}$ for the set of s-states, because $\rightarrow_{\mathcal{A}}$ works for pairs of states.

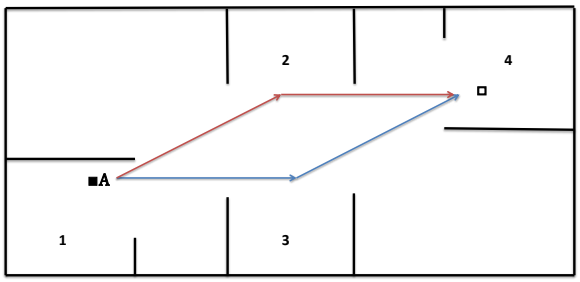


Fig. 4. The polygonal environment of the robot's motion with 4 rooms. The blue broken line illustrates the planned trajectory of the robot's move from a room no. 1 to the room no. 4. The red one illustrates the deviated trajectory of the robot's move.

A. The motion robot environment and its $LTL \bigcup HS^{D,L}$-specification

Let us consider a robot, say R, in some polygonal environment with 4 rooms as depicted on a picture below. Assume that R performs a task to dislocate a black block A from a room 1 to the room 4 and put in on a block B there and the planned (preferred) move trajectory leads from the room 1 by a neighborhood of the room 3 to the room 4 (the blue line on a picture). Let also assume that our robot exchanged this trajectory for another one (marked by a red line).
Therefore, the robot motion environment and plan specification in $LTL \cup HS^{L,D}$ may be rendered as follows:

- **Plan + Preferences**:
  1) Take a block A.
  2) Move from R_1 to R_3 (more preferable) or Move from R_1 to R_4 (less preferable).
  3) If you are in R_3, move from R_3 to R_4.
  4) Go to the room R_4).
  5) Put a block A on the block B.

We can also extract the following 'behavioral' rule for the robot as a condition of an effective plan performing.

- **Condition for the plan performing/behavioral rules**:
  1) Always in a future, if you take a block A, go to the room R_3.

Finally, we should give a short description of the polygonal robot motion environment.

- **The robot environment**:
  1) Environments consists of 4 rooms ,
  2) In a room R_1, block A is initially located ,
  3) In a room R_4, block B is initially located ,
  4) In rooms R_2, R_3 no blocks are located ,
  5) In room R_4, block A is finally located ,
  6) The robot motion area is always located on the left of a room R_1 ,
  7) The robot's motion area is always located on the right of a room R_4 .

All these conditions may be regarded now in terms of $LTL \bigcup HS^{D,L}$ in the corresponding way as follows:

- **Plan + Preferences**:
  1) $Take(A)$.
  2) $Move(R_1^A, R_3) \vee Move(R_1^A, R_4)$.
  3) $HOLDS(R_3^R) \rightarrow Move(R_3, R_4)$.
  4) $Put(A)$.
  5) $HOLDS(R_4^A)$.
- **Condition for the plan performing/behavioral rules**:
  1) $G(take(A) \rightarrow \langle L \rangle go(R_3)$.
- **The robot environment**:
  1) $R_1 \wedge R_2 \wedge R_3 \wedge R_4$,
  2) $R_1^A \wedge R_4^B$ $(HOLDS_{Init}(R_1^A))$,
  3) R_4^A $(HOLDS_{Fin}(R_1^A))$,
  4) $[D](R_1 \rightarrow Left)$
  5) $[D](R_4 \rightarrow Right)$.

Let us return now the initial assumption that the robot deviated from the planned path and has chosen a red line from $R_1 to R_4$, so via R_2. We can trace this deviation for the following juxtaposition of two formal descriptions in terms of $LTL \bigcup HS^{D,L}$ for both situations.

plan of the robot	**the real plan performing**
$Take(A)$	$Take(A)$
$Move(R_1^A, R_3) \vee Move(R_1^A, R_4)$	$Move(R_1^A, R_2)$
$HOLDS(R_3^R) \rightarrow Move(R_3, R_4)$	$Move(R_2^A, R_4)$
$Put(A)$	$Put(A)$
$HOLDS(R_4^A)$	$HOLDS(R_4^A)$
behavioral rule	**behavioral rule**
$G(take(A) \rightarrow \langle L \rangle go(R_3)$	?

B. From $LTL \cup HS^L$ to the Büchi automaton

The next stage of the plan controller construction consists in a translation of $LTL \cup HS^L$-formulas to states of an automaton which usually is not given *a priori*, but it must be constructed in the appropriate way. The automaton in our case will be constructed due to [21], [22] *via* such that states are identified with subsets of closures of the $LTL \cup HS^L$-formulas. In order to illustrate this procedure let us consider a case of a single LTL-formula Take(A). Due to definition 5 from p. 5 a closure of a formula ϕ contains all of its sub-formulas and its negations. In this case we get the following sets of formulas: $\emptyset$, $\{\neg A\}, \{A\}$, $\{TAKE(\neg A)\}$, $\{A, TAKE(\neg A)\}, \{TAKE(A)\}, \{A, TAKE(A)\}$ – as depicted on a fig. 3.

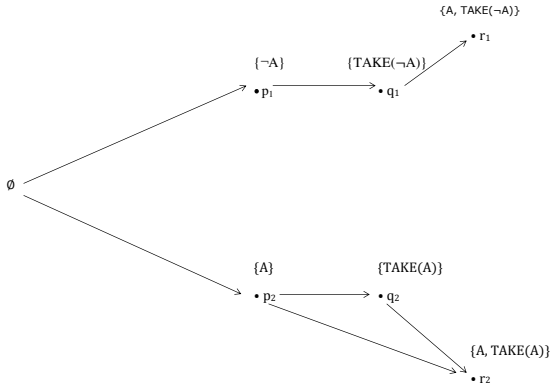


Fig. 5. Fragment of the Büchi automaton with states for a closure of the LTL-formula Take(A).

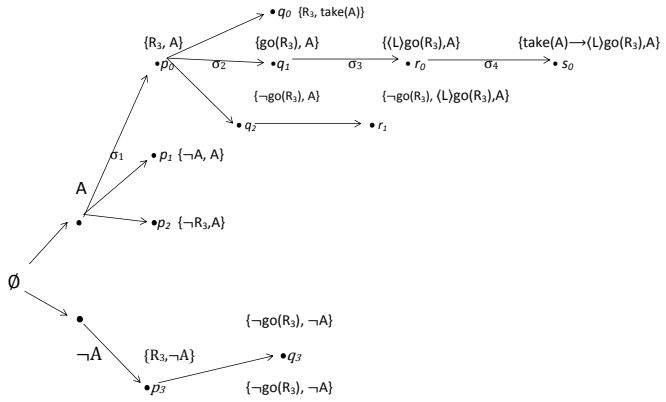


Fig. 6. Fragment of the Büchi automaton with states for a closure of the LTL-formula $Take(A) \rightarrow \langle L \rangle go(R_3)$.

The fragment of Büchi automaton for a more complicated formula $LTL \cup HS^L$-formula $take(A) \rightarrow \langle L \rangle go(R_3)$ was presented on a diagram 4.

The fragment of the Büchi automaton, say $\mathcal{A}$, for LTL-formula $Move(R_1^A, R_3)$ expressing the second step of the robot's plan is more complicated and it looks like depicted on fig.3. The same principle determines a construction way

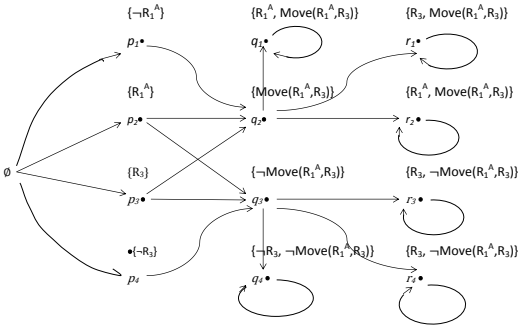


Fig. 7. Fragment of the Büchi automaton with states for a closure of the LTL-formula a$Move(R_1^A, R_3)$.

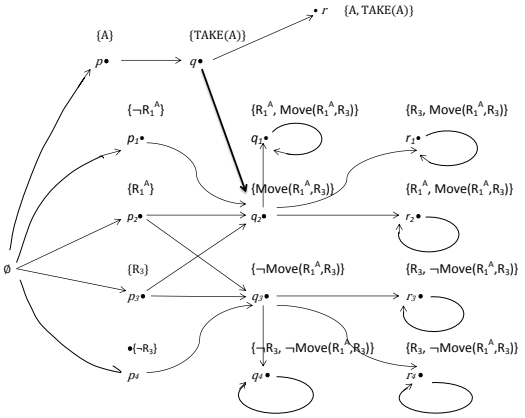


Fig. 8. Fragment of the Büchi automaton with states for a closure of the LTL-formulas $Take(A)$ and $Move(R_1^A, R_3)$.

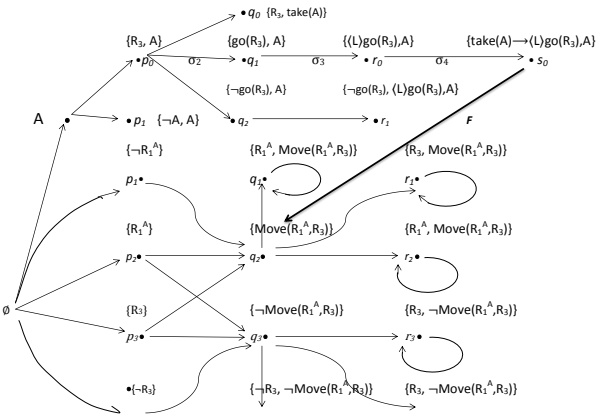


Fig. 9. Fragment of the Büchi automaton for $Move(R_1^A, R_3)$ and for a fragment of a formula $take(A) \rightarrow \langle L \rangle go(R_3)$.

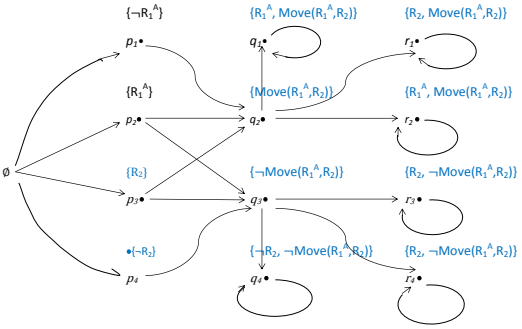


Fig. 10. Fragment of the Büchi automaton for the real task performing for the closure of LTL-formula $Move(R_1^A, R_2)$.

of the automaton fragment for $Move(R_1^A, R_3)$ – as depicted on a Fig. 5. In addition, the fragment of automaton for both formulas: $Take(A)$ and $Move(R_1^A, R_3)$ (taken together) is demonstrated on a Fig. 6. Finally, a more extended fragment of this automata for $Move(R_1^A, R_3)$ and the 'behavioral rule' of the robot $Take(A) \rightarrow \langle L \rangle go(R_3)$ – providing the plan performing – was presented on a Fig. 7.

It is not difficult to observe that a 'global size' of such automata for a complete plan of the robot and its motion environment is very large and its complete presentation would be difficult and non-suggestive. It is enough to observe that the full automaton is a composition of the appropriate fragments enriched by some 'move-lines' (the dark line on a Fig. 7) in order to connect the appropriate fragments of this automaton.

In order to detect discrepancies between the plan-automaton (for $Move(R_1^A, R_3)$ and for the 'behavioral rule' $take(A) \rightarrow \langle L \rangle go(R_3)$) with the corresponding part of automaton for the real performing of the plan, it is enough to compare both pictures. (The different states, i.e. with different formulas satisfied in them, are marked in a blue color.) Moreover, the plan-automata is larger as it also contains the transition 'branch' for the 'behavioral rule' of the robot. This branch cannot be added to the second automaton – due to observations from last section.

It has emerged that a discrepancy on the level of $LTL \cup HS^L$-formalism can be naturally reflected by its corresponding automata – as depicted on the Fig. 8. and Fig. 6.

C. The product automaton

As it has already been mentioned – each product automaton preserves some portion of information encoded by two automata: by the plan-automata, say $\mathcal{A}^{plan}$ and by its 'rival' for the real task performing, say $\mathcal{A}^{perf}$. As each product structure, the product automaton $\mathcal{A}^{plan} \times \mathcal{A}^{perf}$ is built up from pairs of states of the form: (s_1^{plan}, s_2^{perf}), where each $s_1^{plan} \in \mathcal{A}^{plan}$ and each $s_2^{perf} \in \mathcal{A}^{perf}$. [5] The similar pairs can be constructed with respect to transitions taken from $\mathcal{A}^{plan}$ and their equivalents from $\mathcal{A}^{perf}$ (if there are)[6].

The product automata – due to the earlier definition of preferential product automaton – $\mathcal{A}^{plan} \times \mathcal{A}^{perf}$ must have a general form:

$$\langle \Sigma_1 \times \Sigma_2, S^{plan} \times S^{perf}, S_0^{plan} \times S_0^{perf}, \rightarrow, \mathcal{F}^{\alpha} \times \mathcal{F}^{\beta}, \alpha_1, \alpha_2 \ldots \rangle . \tag{8}$$

Assuming that the choice of $Move(R_1, R_3)$ is preferable with a degree, say $\frac{2}{3}$ and the choice of $Move(R_1, R_4)$ – with a degree $\frac{1}{3}$ in the robot plan, the fragment of product automata $\mathcal{A} \times \mathcal{A}^{pref}$ for formulas $Move(R_1, R_3) \vee Move(R_1, R_4)$ and $Move(R_1, R_2)$ – due to definition – will have the following algebraic representation:

$$\Sigma_1 \times \Sigma_2 = \mathcal{L}(LTL \cup HS^L) \times \mathcal{L}(LTL \cup HS^L),$$
$$S^{plan} \times S^{perf} = 2^{cl(Move(R_1,R_3) \vee Move(R_1,R_4)} \times 2^{cl(Move(R_1,R_2)},$$
$$S_0^{plan} \times S_0^{perf} = \{\emptyset, \{R_1\}, \{\neg R_1\} \ldots\}^2,$$
$$\mathcal{F}^{\alpha} \times \mathcal{F}^{\beta} = \{R_1, R_3, R_4, Move(R_1, R_3)^{\frac{2}{3}}, Move(R_1, R_4)^{\frac{1}{3}}\} \times \{R_1, R_2, Move(R_1, R_2)\},$$
$$\alpha_1 = \tfrac{2}{3}, \alpha_2 = \tfrac{1}{3}.$$

The diagram presentation of the product automaton $\mathcal{A} \times \mathcal{A}^{pref}$ is simple and much more suggestive. For example, a diagram of the product automaton fragment for formulas $Take(A), Move(R_1^A, R_3)$ 'produced' with $Move(R_1^A, R_2)$ would be a product of single automata from Fig.6 and Fig. 8. (Both these automata – so to speak – combined together). The preferences imposed on some transition paths must be only marked by the appropriate values such as $\frac{2}{3}$, $\frac{1}{3}$ [7].

D. PROLOG-description of the product automaton

The last step of our plan controller construction is to encode the product automaton – earlier described in details – in a declarative language such as PROLOG. An idea of encoding is simple: a 'status' and localization of automaton states will be described by such predicates as: final(x), initial(x), but the transitions between them may be rendered by 3-argument predicates arc(x,y,z).

In this framework the PROLOG-description of the fragment of $\mathcal{A}^{plan}$-automaton for two formulas: $Take(A) \cup Move(R_1^A, R_3)$ may be as follows:

```
initial(0).
  final(r3).    final(r1).    final(r2).
final(r3).    final(r4).    final(q4).
arc(0,p1). arc(0,p2). arc(0,p3).
arc(0,p4).
arc(p1,q2). arc(p2,p2). arc(p2,q2).
arc(p2,q3). arc(p3,q2). arc(p3,q3).
arc(p4,q3).
arc(q1,q1). arc(q2,q1). arc(q2,r1).
arc(q2,r2). arc(q3,r3). arc(q3,r4).
arc(q3,q4).
arc(r1,r1). arc(r2,r1). arc(r3,r3).
arc(r4,r4).
arc(0, p). arc(p,q). arc(q,r). arc(q,q2).
```

In the similar way one can encode the fragment of the second automaton $\mathcal{A}^{pref}$ for $Move(R_1^A, R_3)$. Since each state is determined by a set of formulas satisfied in it and most of states of this automaton is determined by formulas different from the earlier ones, one should denote the corresponding states – when needed – by capital letters: R_1, R_2, Q_1, Q_2 etc. In order to encode them in PROLOG we will represent them by $'stateR_1', 'stateR_2'$ etc.[8] Thus the PROLOG-description of the required fragment of $\mathcal{A}^{pref}$ looks as follows:

```
initial(0).
final(r3).    final(r1).    final(r2).
final(r3).    final(r4).    final(q4).
arc(0,p1). arc(0,p2). arc(0,P3).
arc(0,P4).
arc(p1,Q2). arc(p2, stateP2).
arc(p2,stateQ2). arc(p2,stateQ3).
arc(stateP3,stateQ2).
arc(stateP3,stateQ3).
arc(stateP4,stateQ3).
arc(stateQ1,stateQ1).
arc(stateQ2,stateQ1).
arc(stateQ2,stateR1).
arc(stateQ2,stateR2).
arc(stateQ3,stateR3).
arc(stateQ3,stateR4).
arc(stateQ3,stateQ4).
arc(stateR1,stateR1).
arc(stateR2,stateR1).
arc(stateR3,stateR3).
arc(stateR4,stateR4).
```

In order to better elucidate differences between both PROLOG-encodings, let us write in the lines which differentiate both codes:

[5] It is not completely correct, because we should pedantically state that s_1^{plan} belongs to some set S^{plan} of states of $\mathcal{A}^{plan}$, but we will omit this distinction for a simplicity and some suggestiveness of analysis.

[6] Of course, both automata $\mathcal{A}^{plan}$ and $\mathcal{A}^{perf}$ are defined over the same alphabet $\Sigma = \mathcal{L}(LTL \cup HS^{L,D})$.

[7] Naturally, these paths (with associated values) could by also distinguished among other ones in other way on a diagram representation of the automaton.

[8] This encoding follows from the fact that capital letters in PROLOG are variables.

```
arc(stateQ1,stateQ1).
arc(stateQ2,stateQ1).
arc(stateQ2,stateR1).
arc(stateQ2,stateR2).
arc(stateQ3,stateR3). arc(stateQ3,R4).
arc(stateQ3,statesQ4).
arc(tateR1,stateR1). arc(stateR2,stateR1).
arc(stateR3,stateR3).
arc(stateR4,stateR4).
```

In addition, the third line of this first code for $\mathcal{A}^{plan}$ is partially incompatible with the corresponding line of the second code. The lack of the last line of the first code in the second code is accidental, since it follows from the automaton construction for $Move(R_1^A, R_2)$ only, but the "branch" for $Take(A)$ could be also added to this fragment of $\mathcal{A}^{pref}$. Nevertheless, the branch for $Take(A) \rightarrow \langle L \rangle go(R_3)$ cannot be added to is as the robot did not respect this 'behavioral rule' in its real task performing what it is reflected by $\mathcal{A}^{pref}$.

It remains to enrich these PROLOG–descriptions by some preferential component – as it it has been made with respect to automata. Due to our convention – preferences are denoted by rational numbers from a fuzzy set $[0, 1]$ and they are associated to paths/arcs between automaton states. For simplicity of the PROLOG-representation, we can assume that they can be associated to final states of such paths/arcs. Assume, however, that we do not know how they are associated to concrete formulas or states, but we only known that each of formula can take one of values from the set $\{0, \frac{1}{2}, \frac{2}{3}, 1\}$. If we define the automaton branches for a $Take(A)$-formula by PROLOG lists, say X and Y, we also need to add a piece of information about possible fuzzy values that can be considered:

X ins $0, \frac{1}{2}, \frac{2}{3}$, 1, and Y ins $0, \frac{1}{2}, \frac{2}{3}$, 1.

These coding examples do not exhaust the list of possible ways of PROLOG-encoding, but they are used for illustration and can be extended and specified in many ways.

VI. Conclusions

It has just been demonstrated how the hybrid plan controller can be constructed beginning with a formal description of the robot motion environment. As it could be observed – we were mostly interested in a theoretic side of a construction of such a controller, putting aside its real robotic materialization. This issue could constitute a subject of further research direction.

Anyhow, it has emerged that the initial discrepancy between a plan and its real performing by a robot can be encoded at each stage of the controller construction without losing of any portion of information. In fact, the same discrepancy at the stage of the LTL-description can be transformed to the stage of the automaton construction and – finally – could be visible at the stage of its PROLOG-representation, too. One could venture a thesis that attempts with other languages of a declarative paradigm give the similar results.

Naturally, the preferential extension of automata and the whole construction of the plan controller that we have just proposed forms a kind of an 'external' extension. In fact, we have not introduced any explicit preferential language to LTL extended by HS-language with $\langle D \rangle$ and $\langle L \rangle$. It seems that this task could be feasible in some preferential extension of HS or LTL.

References

[1] M. Antonniotti and B. Mishra. Discrete event models+ temporal logic = supervisory controller: Automatic synthesis of locomotion controllers. *Proceedings of IEEE Intern. Conf. on Robotics and Automation*, 1999.

[2] F. Bacchus and F. Kabanza. Using temporal logic to express search control knowledge for planning. *Artificial Intelligence*, 116, 2000.

[3] R. Buchi. On a decision method in restricted second-order arithmetic. *Stanford University Press*, 1962.

[4] G. Fainekos, H Kress-gazit, and G. Pappas. Hybrid controllers for path planning: A temporal logic approach. *Proceeding of the IEEE International Conference on Decision and Control, Sevilla*, December:4885–4890, 2005.

[5] G. Fainekos, H Kress-gazit, and G. Pappas. Temporal logic moton planning for mobile robots. *Proceeding of the IEEE International Conference on Robotics and Automaton*, pages 2032–2037, 2005.

[6] D. Giacomo and M Vardi. Automata-theoretic approach to planning for temporally extended goals. *Proceeding of the 5th European Conference on Planning*, 1809:226–238, 2000.

[7] J. Halpern and Y. Shoham. A propositional modal logic of time intervals. *Journal of the ACM*, 38:935–962, 1991.

[8] G. Holzmann. *The spin model checker, primer and reference manual*. Addison-Wesley, 2004.

[9] D. Hristu-Varsakelis, M. Egersted, and S. Krishnaprasad. On the complexity of the motion description language mdle. *Proceedings of the 42 IEEE Conference on Decision and Control*, December:3360–3365, 2003.

[10] K. Jobczyk and A. Ligeza. Multi-valued halpern-shoham logic for temporal allen's relations and preferences. *Proceedings of the annual international conference of Fuzzy Systems (FuzzIEEE)*, page to appear, 2016.

[11] K. Jobczyk and A. Ligeza. Systems of temporal logic for a use of engineering. toward a more practical approach. In *Intelligent Systems for Computer Modelling*, pages 147–157, 2016.

[12] K. Jobczyk, A. Ligeza, and K. Kluza. *Proceedings of ICAISC'16*, LNAI II:to appear, 2016.

[13] K. Jobczyk and J. Ligeza, A. nad Kaczmarczuk. Fuzzy-temporal approach to the handling of temporal interval relations and preferences. *Proceedings of INISTA2015*, pages 1–8, 2015.

[14] E. Klavins. A language for modelling and programming cooperative control systems. *Proceedings of the 42 IEEE Conference on Robotics and Automaton, New Orleans*, April, 2004.

[15] M. Mach-Krol. Perspectives of using temporal logics for knowledge management. *Proceedings of FedCsIS*, 49:935–938, 2012.

[16] M. Mach-Krol. Perspects of using temporal logics for knowledge management. *ABICT*, 49:41–52, 2012.

[17] J. Marcinkowski and J. Michaliszyn. The last paper on the halpern-shoham interval temporal logic. 2010.

[18] L. Maximova. Temporal logics with operator 'the next' do not have interpolation or beth property. In *Sibirskii Matematicheskii Zhurnal*, pages 109–113, 32(6)1991.

[19] A. Montanari and P. Sala. Interval logics and omegab-regular languages. *LATA*, LNCS:431–444, 2013.

[20] P. Tabauda and G. Pappas. From discrete specification to hybrid control. *Proceedings of the 42IEEE Conference on Decision and Control*, 2003.

[21] M. Vardi and P. Wolper. An automata-theoretic approach to automatic program verification. *Proceedings of the 1st Symposium on Logic in Computer Science*, June:322–331, 1986.

[22] M. Vardi and P. Wolper. Reasoning about infinite computations. *Information and Computation*, 115(1):1–37, 1994.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 71–78
DOI: 10.15439/2016F548
ACSIS, Vol. 8. ISSN 2300-5963

Rough Sets Applied to Mood of Music Recognition

Bożena Kostek
Gdansk University of Technology, Faculty of Electronics, Telecommunications and Informatics, Audio Acoustics Laboratory, Narutowicza 11/12, 80-233 Gdańsk, Poland
Email: bokostek@audioakustyka.org

Magdalena Plewa
Gdansk University of Technology, Faculty of Electronics, Telecommunications and Informatics, Multimedia Systems Department and Audio Acoustics Laboratory, Narutowicza 11/12, 80-233 Gdańsk, PL
Email: mplewa@sound.eti.pg.gda.pl

***Abstract*—Mood of music is considered as one of the most intuitive criteria for listeners, thus this work is focused on the emotional content of music and its automatic recognition. The research study presented in this work contains an attempt to music emotion recognition including audio parameterization and rough sets. A music set consisting of 154 excerpts from 10 music genres was evaluated in the listening experiment. This may be treated as a ground truth. The results achieved indicated a strong correlation between subjective results and objective descriptors and on that basis a vector of parameters related to mood of music was created. On the other hand, rough set-based processing was applied to derive reducts containing the most promising features in the context of mood recognition, as well as confusion matrices of the mood recognition. Both approaches indicate strong relationship between objective descriptors and subjective evaluation of mood of music.**

Introduction

In music perception studies many different classifications and systems that describe music components are defined. Levitin [1] observed that from the listener's perspective there are seven major elements of music: loudness, pitch, melody, harmony, rhythm, tempo, and meter. These components are significant for discussion related to emotions included in music. The conventional approach to studying music emotion perception or to assigning music genre consists in subjective tests, in which a number of listeners evaluate a given music excerpt, and then these results are analyzed using statistical processing. This process is very lengthy and arduous, and does not always return reliable results, as mood and emotion are often treated by the listeners as interchangeable terms. Even such a notion as "music genre" may not precisely be defined, but still it plays an essential role in music appreciation and cognition. Mood plays an important role in music interpretation and annotation. Still, even though mood is an intuitive way of music describing, it is very difficult to find an exact correlation between physical features and perceived mood, which is necessary to make the annotation process automatic. This way, mood could automatically be added to music recommendation systems [2], [3].

Music as a form of art is perceived and interpreted in many different ways. It contains layers of music composition elements, emotions that may elicit different meanings, references to other pieces and other elements that are difficult to define and interpret. On the other hand, music can be treated as an audio signal and parameterized according to its temporal and spectral characteristics. The relationship between music described by music features and parameters derived from the signal is very difficult to find. The researchers' task is to identify parameters that are related to those features.

As already mentioned, with the growth of accessible digital music libraries over the past decade, there is a need for research into automated systems for searching, organizing and recommending music. Therefore, the aim of this study is to find a correlation between subjective evaluation of music mood and objective data processing. For that purpose, the rough set-based [4], [5] analysis is carried out, employing the Rough Set Exploration (RSES) system [6], [7].

The paper is organized as follows; Section II presents the mood model used in this study, list of labels and the principles of the listening tests carried out by the authors, as well as results of mood perception-based evaluation. In Section III correlation analysis is performed for data gathered in the listening tests and in feature vectors that are assigned to the set of 150 music excerpts. As a result, a table is created that contains the most significant parameters correlated to two mood-related features, namely: valence and arousal. Next, the same data are analyzed using the rough set-based processing and then results obtained compared to those presented in Section III. Finally, a short summary is included in Conclusion Section.

Subjective Evaluation

A. Model of Mood of Music

This section includes information with regard to the main experiment that aimed for subjective emotional content evaluation of larger set music. Outcomes from previous experiments [8]–[12] were taken into consideration and affected the final form of this experiment including the model of emotions, music dataset and the experimental procedure.

The test was executed to collect subjective mood evaluation results of a large set of songs using specially

This work was supported by Gdansk University of Technology

designed graphical interface (Fig. 1, for translation see Table I). The model of mood of music used in the experiment was proposed by the authors. The main assumptions were that the model has to be intuitive for users and compatible with dimensional models consisting of two dimensions as proposed by Thayer [13].

Fig. 1 Graphical interface dedicated for mood of music evaluation

Since previous stages of the research showed that 2-dimensional model is not very intuitive for listeners, an alternative solution was proposed [10]. The set of mood labels was selected from the Mood Dictionary Associated with Music proposed by the authors [10] Mood labels (originally in Polish) along with their translation can be found in Table I.

Mood descriptors were placed on a 2-dimensional plane, with regards to dimensions retrieved from Multidimensional Scaling (MDS) experiment [11]. This placement is coherent with Thayer's model [13] and Russel's [14] emotion representation. It is also consistent with findings of Brinker *et al*., [15] and Hevner [16], who placed mood descriptors on Valence/Arousal (VA) plane.

TABLE I.
LIST OF MOOD LABELS USED IN GRAPHICAL INTERFACE DESIGNED FOR MOOD OF MUSIC REPRESENTATION

No.	Mood label (Polish)	Mood label (English)
1	Agresywny	Aggressive
2	Depresyjny	Depressive
3	Ekscytujący	Exciting
4	Energiczny	Energetic
5	Neutralny	Neutral
6	Relaksujący	Relaxing
7	Smutny	Sad
8	Spokojny	Calm
9	Wesoły	Joyful

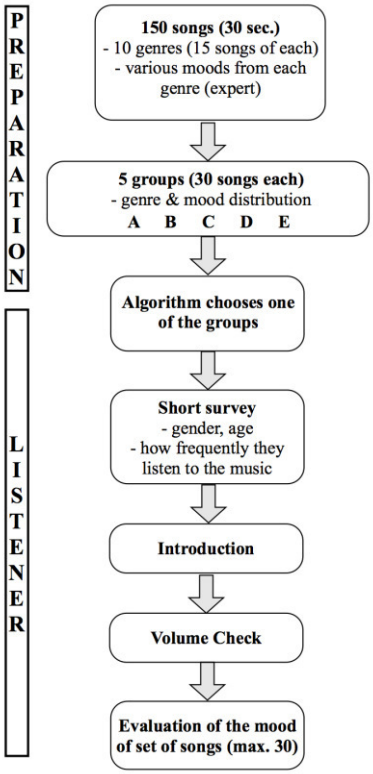


Fig. 2 Listening test arrangement related to music mood evaluation

The intensity of color corresponds to the intensity of the particular emotion contained in music. The "white" area placed in the center is considered as a neutral, where no emotional content is included. This concept is also strongly related to fuzzy logic and the concept of "degrees of truth". It includes various states of truth in between 0 (false) and 1 (truth), which is very intuitive, when it comes to evaluation of such a subtle substance as emotions.

B. Listening Procedure

The stages of the test are presented in Fig. 2. The main part of the test consisted of a series of musical excerpts presented one after another, where listeners were asked to evaluate the mood of music by clicking at the graphical mood representation. It was preceded by a short survey and the level check. The experiment was performed in Polish and its total average duration was approximately 12 minutes. The test was performed using a WEB-based interface. 112 listeners (57 women and 55 man) within age from 16 to 56 (average age 28) participated in the experiment. Majority of the audience reported that they listen to music everyday.

150 tracks were chosen from 10 different music genres, to obtain a diversified set. Chosen music styles were as follows: Blues, Classical, Country, Dance & DJ, Hard Rock & Metal, Jazz, Pop, R&B, Rap & Hip-Hop, Rock. It is worth noting, that styles chosen are easy to distinguish between each other and cover various music material. 30-sec music excerpts used in the experiment came from the SYNAT music database

[17], [18]. A detailed list of music files used in the experiments is available in [9].

C. Results

Answers provided by subjects were pre-processed and only valid entries were included in the result analysis. A submission was considered as valid, if consisted of 5 to 30 evaluated songs.

Results were analyzed with both dimensional and label approaches. In the dimensional approach answers were analyzed in the polar coordinates. To each field on the graphical interface, the number value was assigned according to the mood intensity (from 0 to 3) and the angle was assigned according to the position of the label. This allowed mapping onto 2D Energy/Arousal plane (Fig. 3). Achieved results were used as polar and Euclidean coordinates depending on the employed method. Selected results are listed in Table II.

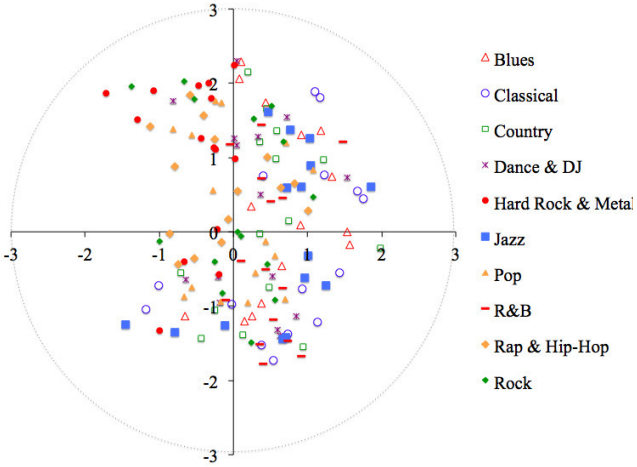


Fig. 3 *Mapping of 150 songs (divided by the genre) onto mood plane based on the listening test results*

Averaged ratings are mostly within range [0,2]. Only for aggressive mood a stronger result was obtained. In this case mood of music might be strongly related to the perceived emotion. Also distribution of music pieces sorted by genre is very interesting.

For pop and rock music, excerpts are distributed in all quadrants of the AV plane. There is no jazz songs in III quadrant. Blues was placed by listeners mostly on the right half of the VA plane (positive) as well as classical, where only two pieces were placed in quadrant III. A similar tendency is observed for country excerpts. Rap & Hip-Hop was considered mostly as music with high arousal and appeared mostly in quadrants I and II but some excerpts are placed in quadrant III.

Distribution of Blues and R&B was quite similar (mostly quadrants I and IV), which might be related to musical similarities and common roots of both genres. Although blues is often considered as very sad music, it was placed on the "positive" half of VA plane.

What is interesting, Dance & DJ excerpts are distributed among all quadrants as well as Pop and Rock. These music genres are very frequent in popular culture, therefore are not strongly related to one and only esthetics or topic but are mixtures of different trends. That is also reflected in values of the standard deviation, which are highest for these genres. Averaged results for each genre are listed in Table II. This table was used in the rough set-based analysis further on. Detailed results of the listening test are presented in [9].

TABLE II.
AVERAGED RESULTS FOR VARIOUS MUSIC GENRES

Music genre	Averaged Valence	Averaged Arousal	St. Dev. Valence	St. Dev. Arousal
Blues	0.60	0.33	1.11	1.16
Classical	0.68	-0.25	1.28	1.25
Country	0.44	0.10	1.05	1.02
Dance & DJ	0.25	0.26	1.06	1.15
Hard Rock & Metal	-0.55	1.03	0.99	1.09
Jazz	0.60	-0.10	1.19	1.33
Pop	0.04	0.22	1.04	1.20
R&B	0.49	-0.32	0.99	1.12
Rap & Hip-Hop	-0.17	0.62	1.10	1.12
Rock	0.00	0.43	1.24	1.13

All observations listed above are important cues for conducting experiments with decision systems. Although they are related to the specific music set, they might represent more general trends due to carefully selecting excerpts for the performed evaluation.

Contrarily, the label analysis was performed, where the number of occurrences of each label was calculated for every song. As a result, each song is described with a 9-element vector, where each position refers to the mood label (Table I). The value describes the percentage of occurrences of each label. Some songs are described by all listeners with mainly one label, while for other evaluations is spread among the labels. Examples of songs along with their label description are shown in Fig. 4.

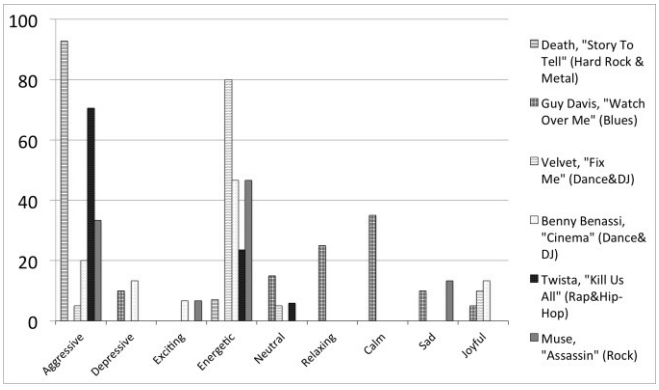


Fig. 4 *Example of results of mood labels assigned to particular songs. The vertical axis describes the percentage of occurrences of each label*

CORRELATION ANALYSIS

The starting point of the feature vector (FV) content creation for the purpose of automatic mood recognition was examination of previous studies performed in MIR by the authors and their collaborators. Resulted from them was FV applied to two databases, namely ISMIS [18] and SYNAT [9], [19]–[25], thus its content may be treated as very thoroughly analyzed. Moreover, the same FV was used in the ISMIS'2011 conference in music competition [18], in which more than 100 teams participated, thus it may also be treated as a kind of benchmarking. ISMIS is a database of approx. 1300 music excerpts of high quality audio excerpts, collected and divided into six music genres. On the other hand, the SYNAT database is a collection of 52532 pieces of music described with a set of descriptors obtained through the analysis of mp3-quality recordings. For the SYNAT database, the analysis band is limited to 8kHz. The database stores 173-feature vectors, which in majority are the MPEG-7 standard parameters (109). The vector has additionally been supplemented with 20 Mel-Frequency Cepstral Coefficients (MFCC), 20 MFCC variances and 24 time-related 'dedicated' parameters. The SYNAT database was realized by the Gdansk University of Technology (GUT) [19] and music was collected from the Internet. The vector includes parameters associated with the MPEG-7 standard, melcepstral coefficients (MFCC) and is supplemented by the so-called dedicated parameters which refer to temporal characteristic of the analyzed music excerpt. Full list of parameters was shown in the earlier study [12], [20].

In addition, parameters from the MIR Toolbox [26] were calculated. This set of parameters contains a lot of information but not necessarily related to mood of music. At previous stages of this study a selection of parameters based on correlation analysis was performed and returned good results, therefore this approach was also applied. Correlation between subjective values of Valence and Arousal and parameters was calculated and therefore a set of features strongly related to mood was created. Only parameters with correlation coefficient higher than 0.50 were included in the final feature vector. Eventually, the feature vector describing mood of music consisted of 16 parameters from SYNAT, listed in Table III and 16 parameters from MIR Toolbox Table IV. It is worth noting that correlation is slightly stronger with parameters based on music characteristic than from SYNAT, which describe general properties of an audio signal. Description of parameters included in SYNAT can be found in [9], [12] and features from MIR Toolbox in the MIR manual [26].

TABLE III.
PARAMETERS CORRELATED WITH SUBJECTIVE MOOD OF MUSIC EVALUATION SELECTED FROM 173-SYNAT FVS

Valence		
No.	**Parameter**	**Corr.**
1	ASE2	-0.72
2	MFCC7	-0.62
3	PEAK_RMS10FR_MEAN	0.51
4	ASE_M	-0.50
5	ASE26	-0.50
Arousal		
6	MFCC1	-0.79
7	MFCC2	-0.78
8	SFM13	-0.63
9	SFM12	-0.61
10	SFM14	-0.56
11	SFM15	-0.56
12	SFM10	-0.53
13	1RMS_TCD	0.52
14	SFM_M	-0.51
15	SFM11	-0.51
16	SFM16	-0.50

TABLE IV.
PARAMETERS CORRELATED WITH SUBJECTIVE MOOD OF MUSIC EVALUATION SELECTED FROM THE MIR TOOLBOX

Valence		
No.	**Parameter**	**Corr.**
1	Spectral irregularity	0.73
2	MTBF2	0.54
3	Spectral roughness	0.52
Arousal		
4	Brightness	0.83
5	Entropy of Spectrum	0.79
6	Timbre Zerocross.	0.64
7	Tonal Harmonic Change Detection	0.64
8	Harmonic Change Detection Function	0.62
9	Spectral Centroid Mean	0.57
10	Key Clarity	0.55
11	Tempo	0.55
12	Spectral Flux Period	0.55
13	Spectral Irregularity	0.53
14	Spectral Rolloff 185	0.52
15	Spectral Kurtosis	0.51
16	Spectral roughness	0.50

ROUGH SETS-BASED ANALYSIS

A. Input Data for Rough Sets

As mentioned earlier, the RSES was used in this study. The system generates interpretable results in the form of reducts,

classification measures, etc., thus it is very useful for the analysis and classification of data.

Data for the rough set-based analysis were prepared in various configurations. Table V gathers datasets in a form of decision tables, in which: "Mood label I", " Mood label II", and "Music Genre" are the decision attributes. Averaged values of Valence and Arousal were calculated, and averaged position on the Valence/Arousal plane indicated Mood label I. Mood label II was chosen according to the label analysis, where the number of occurrences of each label was calculated for every song and the label with maximum number of occurrences was chosen. The last column shows music genres. As seen from Table V, not always labels from Mood labels I and II are consistent between each other. Moreover, it is interesting that the position in VA plane does not condition music genre assigned.

Datasets from Table V were analyzed by the RSES system, the returned results are shown in a form of confusion matrices (Tables VI-VIII), containing true positives (TP), as well as coverage and accuracy measures. Data were randomly split into 60% and 40% subsets for training and testing the classifier. For every dataset a separate set of rules was created by the following procedure:

1. Every parameter value range was discretized by a local variant of Maximum Discernibility discretization method.
2. Reducts were calculated from training subset by the genetic method available in RSES software.
3. Reducts were tested against training subset to create decision rules, based on every object in the subset.
4. Testing subset was classified and accuracy was expressed by calculating a confusion matrix.

TABLE V.

LISTENING EXPERIMENT RESULTS USED IN THE RSES-BASED PROCESSING

Song No.	Valence	Arousal	Mood Label I	Mood Label II	Music Genre
1	0.09	2.06	Energetic	Energetic	Blues
2	0.43	1.74	Energetic	Energetic	Blues
...	...	...	...		...
15	0.92	1.31	Exciting	Energetic	Blues
16	0.41	0.75	Exciting	Exciting	Classical
17	1.23	0.77	Exciting	Exciting	Classical
25	-1.00	-0.73	Depressive	Sad	Classical
...	...	...	...		...
30	1.67	0.55	Joyful	Exciting	Classical
31	0.36	1.21	Energetic	Energetic	Country
32	0.19	2.15	Energetic	Energetic	Country
...	...	...	...		...
40	-0.71	-0.55	Depressive	Depressive	Country
45	1.21	0.97	Exciting	Energetic	Country
46	-0.81	1.77	Aggressive	Aggressive	Dance & DJ
47	-0.21	-0.61	Calm	Relaxing	Dance & DJ
...	...	...	...		...
60	0.72	1.54	Exciting	Energetic	Dance & DJ
61	-1.30	1.51	Aggressive	Aggressive	Hard Rock & Metal
62	-1.72	1.86	Aggressive	Aggressive	Hard Rock & Metal
...	...	...	...		...
71	-0.19	-0.57	Calm	Sad	Hard Rock & Metal
...	...	...	...		...
75	-0.26	1.14	Energetic	Aggressive	Hard Rock & Metal
76	0.77	1.37	Exciting	Energetic	Jazz
77	0.72	0.59	Exciting	Energetic	Jazz
...	...	...	...		...
90	0.92	0.61	Exciting	Energetic	Jazz
91	-0.56	1.30	Energetic	Energetic	Pop
92	-0.16	1.74	Energetic	Energetic	Pop
...	...	...	...		...
105	0.69	-0.90	Relaxing	Calm	Pop
106	-0.05	1.17	Energetic	Energetic	R&B
107	0.38	0.72	Exciting	Exciting	R&B
...	...	...	...		...
120	0.43	-0.51	Relaxing	Neutral	R&B
121	-1.12	1.42	Aggressive	Aggressive	Rap & Hip-Hop
122	-0.59	1.84	Energetic	Aggressive	Rap & Hip-Hop
...	...	...	...		...
135	-0.26	1.25	Energetic	Aggressive	Rock
136	-1.37	1.96	Aggressive	Aggressive	Rock
...	...	...	...		...
149	0.24	-1.49	Calm	Calm	Rock
150	1.08	0.47	Exciting	Joyful	Rock

B. Results

As seen from Tables VI-VIII, the rough set-based processing brought satisfying results with a high accuracy and coverage.

TABLE VI.

CONFUSION MATRIX FOR CLASSIFICATION OF MOOD LABEL I, WITH ATTRIBUTES "VALENCE", "AROUSAL", "GENRE"

	Energ.	Exc.	Joy.	Calm	Relax.	Depr.	Aggr.	Sad	No. of obj.	Acc.	Cov.
Energ.	16	0	0	0	0	0	0	0	16	1	1
Exc.	0	11	0	0	0	0	0	0	11	1	1
Joy.	0	0	4	0	0	0	0	2	6	0.667	1
Calm	1	0	0	9	0	0	0	0	10	0.9	1
Relax.	1	0	0	0	1	0	0	0	2	0.5	1
Depr.	0	0	1	0	0	7	0	0	9	0.875	0.889
Aggr.	1	0	0	0	0	0	3	0	4	0.75	1
Sad	0	0	0	0	1	0	0	0	2	0	0.5
TP	0.84	1	0.8	1	0.5	1	1	0			

Total number of tested objects: 60, Total accuracy: 0.879, Total coverage: 0.967.

TABLE VII.
CONFUSION MATRIX FOR CLASSIFICATION OF MOOD, WITH FEATURES "VALENCE", "AROUSAL", CLASSIFICATION GOAL IS (MOOD LABEL I)

	Energ.	Exc.	Joy.	Calm	Relax.	Depr.	Aggr.	Sad	No. of obj.	Acc.	Cov.
Energ.	10	0	1	1	0	0	0	0	13	0.833	0.923
Exc.	0	4	0	0	0	0	0	0	5	1	0.8
Joy.	0	0	13	1	0	0	0	0	14	0.929	1
Calm	0	0	0	2	0	0	0	0	8	1	0.25
Relax.	0	0	0	0	7	1	0	0	8	0.875	1
Depr.	0	0	0	0	0	9	0	0	10	1	0.9
Aggr.	0	0	0	0	0	0	2	0	2	1	1
Sad	0	0	0	0	0	0	0	0	0	0	0
TP	1	1	0.93	0.5	1	0.9	1	0			

Total number of tested objects: 60, Total accuracy: 0.922, Total coverage: 0.85.

TABLE VIII.
CONFUSION MATRIX FOR CLASSIFICATION OF MOOD, WITH FEATURES "VALENCE", "AROUSAL", CLASSIFICATION GOAL IS MOOD LABEL II

	Energ.	Exc.	Joy.	Calm	Relax.	Depr.	Aggr.	Sad	No. of obj.	Acc.	Cov.
Energ.	12	0	0	0	0	2	0	0	15	0.857	0.933
Exc.	0	9	0	0	0	0	0	0	9	1	1
Joy.	0	0	4	0	0	0	0	0	5	1	0.8
Calm	0	0	0	9	1	0	0	0	12	0.9	0.833
Relax.	0	0	1	0	6	0	0	0	8	0.857	0.875
Depr.	0	0	1	0	0	5	1	0	7	0.714	1
Aggr.	0	0	0	0	0	0	2	0	2	1	1
Sad	0	0	1	0	0	0	0	0	2	0	0.5
TP	1	1	0.57	1	0.86	0.71	0.67	0			

Total number of tested objects: 60, Total accuracy: 0.87, Total coverage: 0.9.

Another set of data containing FVs and assigned Mood label I, Mood label II and music genre for the same 150 music excerpts was called FVs. These data returned confusion matrix (see Table IX). This dataset was classified by the following 10 reducts, with sizes ranging from 8 features to 11 features, each with positive region of 1.0 (i.e. percentage of objects from training set accurately classified), see Table X.

It should be of interest that reducts contain rhythm-related features belonging to "dedicated" features, such e.g. 1RMS_TCD_10FR_VAR, 2RMS_TCD_10FR_VAR, etc. This confirms a notion that music genre and related mood rely heavily on rhythmic features of music. Moreover, when comparing the correlation analysis performed (Table III), it may be seen that similar parameters were derived from this analysis for Valence and Arousal, which are features corresponding to mood of music.

TABLE IX.
CONFUSION MATRIX FOR FVS-BASED SET

	Energ.	Exc.	Joy.	Calm	Relax.	Depr.	Aggr.	Sad	No. of obj.	Acc.	Cov.
Energ.	3	0	1	0	3	0	6	1	14	0.214	1
Exc.	1	2	4	0	2	1	1	2	13	0.154	1
Joy.	0	1	1	1	0	0	0	0	3	0.333	1
Calm	1	0	0	1	0	2	0	1	5	0.2	1
Relax.	0	0	2	1	7	0	1	1	12	0.583	1
Depr.	0	0	3	2	2	1	1	0	9	0.111	1
Aggr.	0	0	0	0	0	0	3	0	3	1	1
Sad	0	0	0	1	0	0	0	0	1	0	1
TP	0.6	0.67	0.09	0.17	0.5	0.25	0.25	0			

Total number of tested objects: 60, Total accuracy: 0.3, Total coverage: 1.

As mentioned before other datasets combinations were tested. Two more attempts were made to classify songs:
- mood based on "Valence", "Arousal", and music genre features,
- genre based on "Valence", "Arousal", and "Mood Label I" features.

Even though, confusion matrices did not bring very satisfying results, i.e. the total accuracy was at level of 0.25 and the total coverage was approx. 0.56, still, the classification performed, based on a majority voting, returned adequate results.

TABLE X.
REDUCTS DERIVED FROM THE RSES SYSTEM FOR FVS

No.	Parameters
1	1RMS_TCD_10FR_VAR, 2RMS_TCD_10FR_VAR, ASC_V, ASE12, ASE27, ASEV14, MFCC20, SFMV6
2	1RMS_TCD, 1RMS_TCD_10FR_VAR, 2RMS_TCD_10FR_VAR, 3RMS_TCD, ASC_V, ASE_MV, ASE10, ASE17, ASE27
3	1RMS_TCD, 1RMS_TCD_10FR_VAR, 2RMS_TCD_10FR_VAR, 3RMS_TCD, ASC_V, ASE_MV, ASE10, ASE27, SFM7
4	1RMS_TCD_10FR_VAR, 2RMS_TCD_10FR_VAR, 3RMS_TCD_10FR_MEAN, ASE_MV, ASE10, ASE12, ASE20, ASEV22, MFCC20, SFM8
5	1RMS_TCD_10FR_VAR, 2RMS_TCD_10FR_VAR, 3RMS_TCD_10FR_MEAN, ASC_V, ASE_MV, ASE10, ASE5, ASEV14, ASEV22, MFCC20

6	1RMS_TCD, 1RMS_TCD_10FR_VAR, ASE10, ASE17, ASE20, ASE5, ASEV1, MFCC20, MFCC7, SFM5, SFMV6
7	1RMS_TCD, 1RMS_TCD_10FR_VAR, 3RMS_TCD, ASE_MV, ASE10, ASE17, ASE20, MFCC7, SFM5, SFM7, SFMV6
8	1RMS_TCD, 1RMS_TCD_10FR_VAR, 2RMS_TCD_10FR_VAR, ASC_V, ASE10, ASE12, ASEV14, MFCC7, SFM15, SFM8, SFMV6
9	1RMS_TCD, 2RMS_TCD_10FR_VAR, 3RMS_TCD_10FR_MEAN, ASC_V, ASE12, ASE22, ASE27, ASEV14, ASEV17, SFM5, SFM7
10	1RMS_TCD, 2RMS_TCD_10FR_VAR, 3RMS_TCD, ASC_V, ASE_MV, ASE12, ASE17, ASE27, ASE9, ASEV1, ASEV17

Finally, the last experiment used a decision table consisted of the joint Synat and MIR Toolbox parameters (derived from the correlation analysis as the ones correlated with mood of music). The decision attribute was Mood Label I. The confusion matrix for this dataset is presented in Table XI. 44 reducts resulted from the processing, each containing 6 to 8 attributes (Fig. 5).

TABLE XI.

CONFUSION MATRIX FOR CLASSIFICATION OF MOOD

	Energ.	Depr.	Exc.	Joy.	Relax.	Calm	Aggr.	Sad	No. of obj.	Acc.	Cov.
Energ.	5	0	0	1	0	0	1	2	9	0.556	1
Depr.	0	1	0	2	1	2	0	1	7	0.143	1
Exc.	4	0	5	2	1	1	1	0	14	0.357	1
Joy.	0	0	0	3	1	1	0	0	5	0.6	1
Relax.	0	2	0	0	4	1	1	2	10	0.4	1
Calm	1	1	1	1	1	3	1	2	11	0.273	1
Aggr.	0	0	0	0	0	0	2	0	2	1	1
Sad	0	0	0	1	0	0	1	0	2	0	1
TP	0.5	0.25	0.83	0.3	0.5	0.38	0.29	0			

Total number of tested objects: 60, Total accuracy: 0.383, Total coverage: 1.

Omitted attributes:
Tonal_hcdf1_HarmonicChangeDetectionFunction_Mean, Spectral_irregularity1_Spectral irregularity_Mean2, Spectral_kurtosis1_Spectral_kurtosis_Mean.

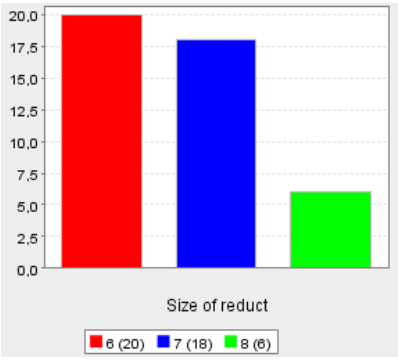


Fig. 5 *Histogram of reduct sizes*

An attribute called Spectral irregularity_mean was the most often used (see Fig. 6), also Spectral brightness is among those most often occurring, which confirm the results derived from the correlation analysis (see Table IV).

Based on these attributes and all training cases a set of rules was generated, containing 2583 rules (Fig. 7), aimed at precise classification of all 90 training objects. Such a disproportion indicates low quality of approximations, and low generalization, as the test set is classified with total accuracy of only 0.383.

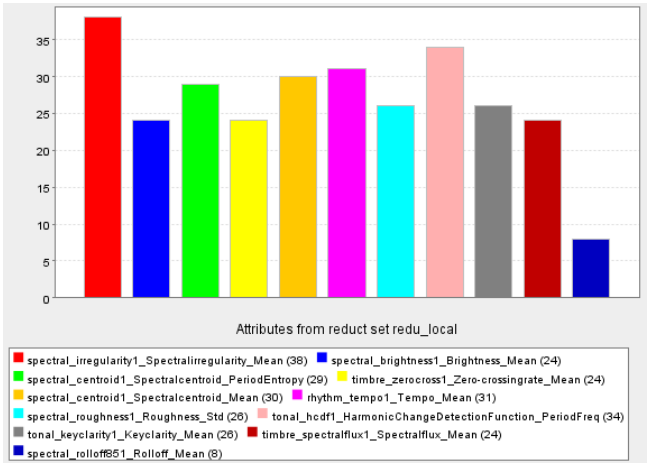


Fig. 5 *Histogram of occurrence of attributes in reducts*

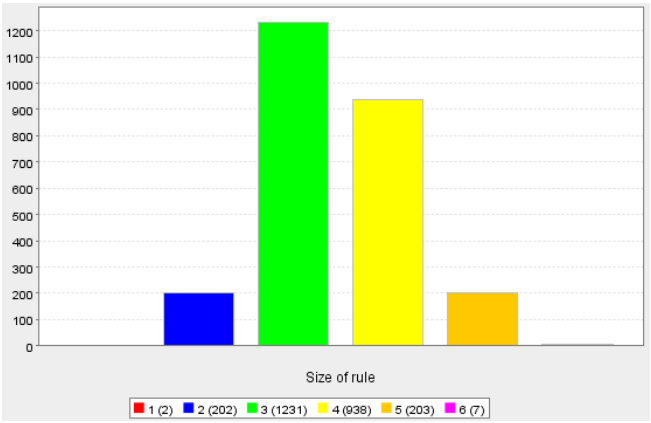


Fig. 7 *Histogram of rule lengths*

CONCLUSION

Several subsets were created for mood classification within the study presented. Gathered data contain results of subjective evaluation, which may be treated as ground truth in this type of classification. This means that we treat listeners' opinions, derived from a meaningful statistical analysis, as the starting point in such analyses. However, it should be remembered that listeners may impose their own emotions on those intended by the composer. That's why it would be preferable if the composer's perspective on the emotion content is known. However from the results obtained, distribution of music objects on the VA plane is coherent with findings available in the literature of the subject

[27], reflected by the prior concepts on musical expressiveness.

Data were carefully analyzed using correlation analysis, as well as the rough set-based processing. In the experiments conducted, the authors used the RSES application for testing the effectiveness of recognizing music genres using the rough set theory. It should be observed that reducts containing only a small number of parameters brought very similar results to those obtained in the correlation analysis, among them spectral irregularity mean and spectral brightness features may be found. The latter analysis was performed for two important features describing mood of music, namely valence and arousal. High accuracy and coverage achieved in rough sets processing confirm consistency between dimensional and label interpretation of the proposed model of mood of music.

REFERENCES

[1] D. J. Levitin, This Is Your Brain on Music: The Science of a Human Obsession, London, Grove/Atlantic, 2008.

[2] M. A. Casey, R. Veltkamp, M. Goto, M. Leman, C. Rhodes, M. Slaney, Content-Based Music Information Retrieval: Current Directions and Future Challenges, Proceedings of the IEEE, vol. 96, 4, pp. 668-696, April 2008.

[3] M. Purgina, A. Kuznetsov, E. Pyshkin, An Approach for Developing a Mobile Accessed Music Search Integration Platform, Proceedings of the 2013 Federated Conference on Computer Science and Information Systems, pp. 267–273, 2013.

[4] Z. Pawlak, "Rough sets", International Journal of Computer & Information Sciences, vol. 11, no. 5, 341-356, 1982.

[5] A. Skowron, L. Polkowski (ed.), "Rough Sets" in Knowledge Discovery vol. 1 and 2, Physica Verlag, Heidelberg, 1998.

[6] J. G. Bazan, M. S. Szczuka, J. Wróblewski, "A new version of rough set exploration system", Third International Conference on Rough Sets and Current Trends in Computing RSCTC, volume 2475, Lecture Notes in Artificial Intelligence, Malvern, PA, Springer-Verlag, pp. 397-404, October 14-16, 2002.

[7] J. Wróblewski, "Covering with Reducts – A Fast Algorithm for Rule Generation", Proceeding of RSCTC'98, LNAI 1424, Springer Verlag, Berlin, pp. 402–407, 1998.

[8] B. Kostek, M. Plewa, "Parametrization and correlation analysis applied to music mood classification", International J. of Computational Intelligence Studies, vol. 2, no. 1, pp. 4-25, 2013.

[9] M. Plewa, "Automatic mood indexing of music excerpts based on correlation between subjective evaluation and feature vector", Doctoral Thesis, Gdańsk University of Technology. Faculty of Electronics, Telecommunications and Informatics, March 2016.

[10] M. Plewa, B. Kostek, "Creating Mood Dictionary Associated with Music, 132 Audio Eng. Soc. Convention", Paper no. 8607, Budapest, April 26-2, 2012.

[11] M. Plewa, B. Kostek, "Multidimensional Scaling Analysis Applied to Music Mood Recognition", 134th Audio Eng. Soc. Convention, Paper no. 8876, Rome, May 4-7, 2013.

[12] M. Plewa, B. Kostek, "Music Mood Visualization Using Self-Organizing Maps", Archives of Acoustics, vol. 50, no. 4, 2015, http://dx.doi.org/10.1515/aoa-2015-0051

[13] R. E. Thayer, The Biopsychology of Mood and Arousal, Oxford University Press, 1989.

[14] J. A. Russel, "A circumplex model of affects", Journal of personality and Social Psychology, 39, pp. 1161-1178, 1980.

[15] B. Brinker, Dinther R., Skowronek J., "Expressed music mood classification compared with valence and arousal ratings", EURASIP J. Audio, Speech, and Music Processing, 1, 2012, http://link.springer.com/journal/13636/2012/1/page/1, access 6.10.2015.

[16] K. Hevner, "The affective value of pitch and tempo in music", American Journal of Psychology, vol. 49, pp. 621-630, 1937.

[17] B. Kostek, "Content-Based Approach to Automatic Recommendation of Music", 131st Audio Eng. Soc. Convention, Paper No: 8506, New York, October 21-23, 2011.

[18] B. Kostek, A. Kupryjanow, P. Żwan, W. Jiang, Z. Ras, M. Wojnarski, J. Swietlicka, "Report of the ISMIS 2011 Contest: Music Information Retrieval", Foundations of Intelligent Systems, ISMIS 2011.

[19] P. Hoffmann, B. Kostek, "Music Data Processing and Mining in Large Databases for Active Media", Active Media Technology, LNCS, vol. 8610, pp. 85 – 95. Springer, 2014.

[20] Hoffmann P., Kostek B., Bass Enhancement Settings in Portable Devices Based on Music Genre Recognition; J. Audio Eng. Soc., vol. 63, no. 12, pp. 980 - 989, 12.2015, http://dx.doi.org/10.17743/jaes.2015.0087.

[21] B. Kostek, A. Kaczmarek, "Music Recommendation Based on Multidimensional Description and Similarity Measures, Fundamenta Informaticae, pp. 1001–1017, DOI 10.3233/FI-2012-0000, 2013.

[22] B. Kostek, P. Hoffmann, A. Kaczmarek, P. Spaleniak, "Creating a Reliable Music Discovery and Recommendation System", Intelligent Tools for Building a Scientific Information Platform: From Research to Implementation, Springer Verlag, 2013.

[23] B. Kostek, "Music Information Retrieval in Music Repositories", Chapter 17, in: Rough Sets and Intelligent Systems (Skowron A., Suraj Z., Eds.), vol. 1, ISRL, 42, pp. 463–489, Springer Verlag, Berlin Heidelberg, 2013.

[24] A. Rosner, F. Weninger, B. Schuller, M. Michalak, B. Kostek, "Influence of Low-Level Features Extracted from Rhythmic and Harmonic Sections on Music Genre Classification", International Conference on Man-Machine Interactions, pp. 467-473, 2013.

[25] A. Rosner, B. Schuller, and B. Kostek, "Classification of Music Genres Based on Music Separation into Harmonic and Drum Components," Archives of Acoustics, vol. 39, no. 4, pp. 629–638 (2014), http://dx.doi.org/10.2478/aoa-2014-0068.

[26] O. Lartillot, "MIRtoolbox 1.4: User's Manual", Finnish Centre of Excellence in Interdisciplinary Music Research Swiss Center for Affective Sciences, 2012.

[27] A. Rauber, M. Frühwirth, "Automatically Analyzing and Organizing Music Archives", 5th European Conference on Research and Advanced Technology for Digital Libraries, Springer, London 2001.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 79–87
DOI: 10.15439/2016F295
ACSIS, Vol. 8. ISSN 2300-5963

Clustering based on the Krill Herd Algorithm with Selected Validity Measures

Piotr A. Kowalski[1,2], Szymon Łukasik[1,2], Małgorzata Charytanowicz [2,3] and Piotr Kulczycki[1,2]

[1] Faculty of Physics and Applied Computer Science
AGH University of Science and Technology
al. Mickiewicza 30, 30-059 Cracow, Poland
Email: {pkowal,slukasik,kulczycki}@agh.edu.pl

[2] Systems Research Institute
Polish Academy of Sciences
ul. Newelska 6, 01-447 Warsaw, Poland
Email: {pakowal,slukasik,mchmat,kulczycki}@ibspan.waw.pl

[3] Institute of Mathematics and Computer Science
The John Paul II Catholic University of Lublin
Konstantynów 1 H, 20-708 Lublin, Poland
Email: mchmat@kul.lublin.pl

***Abstract*—This paper describes a new approach to metaheuristic-based data clustering by means of Krill Herd Algorithm (KHA). In this work, KHA is used to find centres of the cluster groups. Moreover, the number of clusters is set up at the beginning of the procedure, and during the subsequent iterations of the optimization algorithm, particular solutions are evaluated by selected validity criteria. The proposed clustering algorithm has been numerically verified using twelve data sets taken from the UCI Machine Learning Repository. Additionally, all cases of clustering were compared with the most popular method of k-means, through the Rand Index being applied as a validity measure.**

I. Introduction

EXPLORATORY Data Analysis is essentially centred upon tasks of clustering, classification, data reduction and outliers detection. The procedure of clustering consists of dividing a large data set into smaller subsets called 'clusters'. This partition is achieved through developing a function which assigns individual elements of the data collection into each subset. This technique has been applied to a wide range of problems, including various technical tasks [1], robotics [2] and control approaches [3], to aspects of economics [4], as well as to many agricultural issues [5].

This procedure is considered to be an unsupervised method, therefore, the division of the data is based on information directly discovered (derived) from the data itself. Hence, the separation into clusters is made in such a way that the elements within the clusters are very similar to each other, but show a difference to that held in other clusters. [6].

In data clustering procedures, a few main groupings of algorithms can be distinguished. The first of these are hierarchical [7]. In this case, the process consists of phases in which available set of clusters are merged or divided. An example of an algorithm implementing the aforementioned task, is the "bottom up" approach of Agglomerative Clustering [8]. It starts from a division in which every object is a separate cluster. In each subsequent iteration, the various groups are combined on the basis of the adopted criteria. Finally, all tested elements are placed within one cluster. A further example, albeit an opposite, is the "top down" approach of Divisive Clustering Algorithm [9]. Here, all data items start in one cluster, and this splits recursively, as one element represents one cluster.

A second algorithm group is that called 'centroid-based clustering'. This is based on minimizing variance within the clusters. Here, the best known and most commonly employed method is 'k-means procedure' [10].

The application of fuzzy-logic-based techniques [11] are a still further way of completing a clustering task. In so doing, the individual elements of a considerate data set are assigned to more than one cluster. This feature imparts a significant difference to this category of algorithms, when compared to the other procedures. The most popular algorithm of this group is 'C-fuzzy-means' [12].

Density based methods are included within another group of clustering procedures. One of the more recently introduced algorithms is that referred to as the Complete Gradient Algorithm [13]. It based on the nonparametric methodology of statistical kernel estimators as used for the recognition of data set density. This information provides the number, as well as the shape of the proposed clusters. An interesting feature of this algorithm is that it possibilities of adjustment to the authentic structure of data, and, consequently, the achieved

978-83-60810-90-3/$25.00

results are more justifiable with regard to natural point of view.

A further, but similar group of procedures of clustering tasks are those of algorithms based on grid technique. These methods are based on the assumption that data space can be partitioned into a finite number of cells - the grid structure. Subsequently, each cell density is calculated, and, after sorting the cells according to their densities, the clusters centres are determined. What is interesting herein is that this group of algorithms allows for the traversal of neighbour cells. The first algorithm in this group was introduced by Warnekar and Krishna [14]. Nowadays, the most well-known algorithms in this group are CLIQUE, MAFIA, ENCLUS, OptiGrid, O-cluster and CBF. It should be noted that such algorithms can be used for high-dimensional tasks [15].

Yet one more group of clustering algorithms is that based on an optimization algorithm inspired by Nature [16], [17]. In this approach, some metaheuristics are applied for the optimization of adopted division criteria. This action enables the coming about of great similarity of items inside the clusters, and, simultaneously, vast diversity between clusters. The mentioned criteria can be expressed as a specific mathematical formula, using a variety of statistical measures. These criteria are called 'clustering indexes', and their properties are used to assess the quality of the assignment of individual elements of the test set to the appropriate clusters.

Because the task of clustering is a NP-hard problem of combinatorial optimization [18], here – in natural manner – we apply KHA [19] as an optimization technique. This is so as to find the best location for placement of the centre point of cluster. Based on these position of centres, the individual elements of the data set are then assigned to defined groups. Completion of the thus defined clustering method is achieved using the selected three indices separately, and the obtained results are compared with the outcomes of k-means method application, taking into account the Rand Index [20] as a common evaluation criterion.

In next section, the reader will familiarize with some general information concerning optimization tasks and KHA. In Section III, the details of the application of the clustering approach, as well as selected clustering validity measures are being covered. The experimental results of our work are discussed in Section IV. Finally, in the last section of this paper, the reader will find some conclusions regarding the application of the proposed clustering algorithm, as well as intended further research and studies.

II. Optimisation based on Krill Herd Algorithm

KHA is an iterative heuristic procedure inspired by the natural phenomena of krill heard behaviour. This technique is mainly used for solving optimization problems in continuous space. Here, the solution of this problem comes about by finding such an argument x° of space under consideration $S \subseteq R^N$, which satisfies the following formula

$$f(x^\circ) = \min_{x \in S} f(x) \tag{1}$$

where $f(x)$ describes value of cost function.

The KHA originally proposed by Amir Hossein Gandomi and Amir Hossein Alavi in the paper [19], imitates the behaviour of the individual krill moving together as a herd. Such herds, move accordingly to environmental factors such as proximity to neighbours (herd density), dispersion of swarm, food position and any other biological and environmental phenomena.

In order to solve the optimization problem, we apply KHA metaheuristic. Herein, particular elements $x_i = x_i^1, \ldots, x_i^N$ of N dimensional solutions space in the form of P herd's individuals are represented. In the kth iteration, the best solution of the optimization problem as represented by the pth individual is given alternatively by these two equations:

$$x^\circ(k) = arg \max_{p=1,\ldots,P} f(x_p(k)) \quad \text{/for maximalization task/} \tag{2}$$

or

$$x^\circ(k) = arg \min_{p=1,\ldots,P} f(x_p(k)). \quad \text{/for minimalization task/} \tag{3}$$

The above best solution are corresponding with extremal value of cost function $f^\circ = f(x^\circ)$ given as (2) or (3).

The full KHA procedure in flow chart form is shown as Figure 1. This algorithm starts from an initialization of all its parameters, and positions of all P individuals are generated randomly ❶. In next step ❷, the cost function values are calculated for all initial P individuals using (2) or (3). The subsequent stage ❸ is of great importance and is characterized by KHA technique. It consists of formulas describing the movement of particular individuals. Such motion viv-a-vis each individual krill is determined by three main components. They are:

- movement induced by other krill individuals,
- foraging activity,
- random diffusion.

In subsequent time units, vector of movement of ith krill in KHA technique is based on the by Lagrangian formula:

$$\frac{dx_i}{dt} = N_i + F_i + D_i, \tag{4}$$

where N_i is the motion induced by other krill individuals, F_i denotes the foraging motion and D_i is the physical diffusion of the krill individuals, respectively.

The first factor ❹ is a reflection of the social inspiration of the swarm's individual members. In the herd, individuals are maintained at a high density. Hence, the velocity of each individual is influenced by the movement of others. Thus, the direction of movement by the α_i parameter is induced by the presence of other herd members. This parameter is determined on the basis of the following components: local effect and target effect. The fraction of motion is formulated as:

$$N_i^{new} = N^{max}\alpha_i + \omega_n N_i^{old}. \tag{5}$$

Here N^{max} represents the maximum possible speed that can be induced, ω_n in the range $[0, 1]$ is the inertia weight of a particular krill and N_i^{old} is the motion induced in the previous turn. The α_i parameter is defined as:

$$\alpha_i = \alpha_i^{local} + \alpha_i^{target}, \tag{6}$$

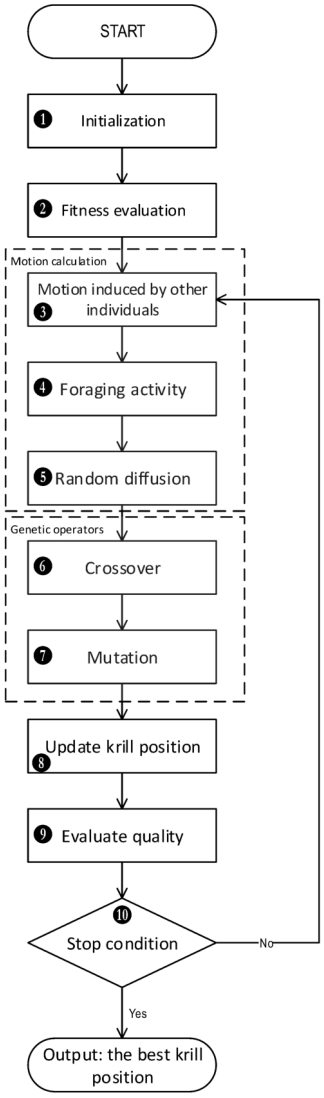


Fig. 1: Flowchart of KHA

where α_i^{local} is the local influence of the neighbours of any particular krill, whereas α_i^{target} is the target direction. The latter is determined by the position and movement of the best individual in a herd.

The α_i^{local} parameters are calculated according to the following formula:

$$\alpha_i^{local} = \sum_{j=1}^{NN} \hat{f}_{ij}\hat{X}_{ij}, \tag{7}$$

where

$$\hat{X}_{ij} = \frac{x_j - x_i}{\|x_j - x_i\| + \epsilon}, \tag{8}$$

and

$$\hat{f}_{ij} = \frac{f_i - f_j}{f^{worst} - f^{best}}. \tag{9}$$

In equation (9), f in describes the fitness value (1) of any investigated krill. Therefore f^{worst} and f^{best} represent, respectively, the worst and the best fitness of individuals in swarm. Additionally, NN provides the identification of the number of reachable krill neighbours, and ϵ is a positive number introduced to avoid singularities in the formula (8).

For determination of distance between particular krills and their neighbours, a parameter designated as being the sensing distance d_s, is introduced. This parameter may be formulated as:

$$d_{s,i} = \frac{1}{5P}\sum_{j=1}^{P}\|x_i - x_j\|. \tag{10}$$

Each individual incorporates its own target vector. This is determined as follows:

$$\alpha_i^{target} = C^{best}\hat{f}_{i,best}\hat{x}_{i,best}, \tag{11}$$

where

$$C^{best} = 2\Big(rand + \frac{k}{K^{max}}\Big). \tag{12}$$

Herein, k, K^{max} designate, respectively, the current iteration number and the maximum number of iterations. Moreover, $rand$ is a random value between 0 and 1, whereas $\hat{f}_{i,best}$ describes the best value of fitness function, while $\hat{x}_{i,best}$ provides the position of the best ith krill individual form the previous iterations.

The next main factor F_i of equation (4), is connected with the food foraging task. This F_i is defined as:

$$F_i = V_f\beta_i + \omega_f F_i^{old}, \tag{13}$$

where V_f is the food foraging speed and ω_f, denotes the inertia of the movement. In this previous equation (13), the food fitness of the ith individual is determined as follows:

$$\beta_i = \beta_i^{food} + \beta_i^{best}. \tag{14}$$

The aforementioned food aspect is determined by way of its location. Therefore, the virtual centre of food concentration is defined via KHA. This conception by the "centre of mass" approach is inspired. Hence, the food concentration in each iteration is calculated according to following formula:

$$X^{food} = \frac{\sum_{i=1}^{P}\frac{1}{f_i}x_i}{\sum_{i=1}^{P}\frac{1}{f_i}}. \tag{15}$$

Moreover, the food attraction for the ith krill individual is described via:

$$\beta_i^{food} = C^{food}\hat{f}_{i,food}\hat{X}_{i,food}. \tag{16}$$

The food coefficient in (16), expresses the global attraction of the food centre (15), and may be calculated as:

$$C^{food} = 2\Big(1 - \frac{k}{K^{max}}\Big). \tag{17}$$

The second part of equation (14) is as follows:

$$\beta_i^{best} = \hat{f}_{i,best}\hat{x}_{i,best}. \tag{18}$$

In this equation, $f_{i,best}$ is the best fit achieved by a given ith krill individual so far. This is characterised by its position $\hat{x}_{i,best}$.

The last element of the Lagrangian equation (4) is related to random physical diffusion ❺, notated as D_i. In essence, this component is of fully random character. This sub-part of movement is focused upon the diversity of population. In addition, it allows the individual krill to escape krill swarm in a situation of local optimum. Moreover, this part of equation (4) represents a trade-off between exploration and exploitation. The following formula describes this aspects of a random diffusion:

$$D_i = D^{max}\Big(1 - \frac{k}{K^{max}}\Big)\delta, \tag{19}$$

where, D^{max} is the maximum diffusion factor and δ describes the random directional vector.

Finally, the motion process can be formally summarized. This employs all the above effective parameters. The position of ith krill during the interval t to $t + \Delta t$ is, thus, determined by the following formula:

$$x_i(t + \Delta t) = x_i(t) + \Delta t \frac{dx_i}{dt}. \tag{20}$$

Here, it must be emphasized that parameter Δt is very sensitive to the speed and accuracy of optimisation task. In this respect, the Δt may be interpreted as being a scale factor of krill movement. This parameter can be obtained by way of the following equation:

$$\Delta t = C_t \sum_{j=1}^{N} (UB_j - LB_j). \tag{21}$$

In this equation, C_t is an empirically found constant number from the interval $[0, 2]$. What is more, UB_j and LB_j are, respectively, the upper and lower bounds of the jth feature $(j = 1, \ldots, N)$ of data set $X = x_1, \ldots, x_P$.

In the next stage of the KHA, the implementation of two basic evolutionary operators is applied. Firstly, in step ❻ the crossover function is considered. This operator is controlled by the crossover probability of the Cr parameter. In this approach, this operator is defined randomly. The crossover results in a change of the mth coordinate of ith krill as shown below by the formula:

$$x_{i,m} = \begin{cases} x_{r,m} & \text{for} \quad \gamma \le Cr \\ x_{i,m} & \text{for} \quad \gamma > Cr \end{cases}, \tag{22}$$

where $Cr = 0.2\hat{K}_{i,best}$; $r \in \{1, 2, ..., i-1, i+1, ..., P\}$ denotes a random index, and γ is a random number drawn from the interval $[0, 1)$ generated according to the uniform distribution. In this approach the crossover operator is calculated upon a single individual.

The last part of the main loop of the KHA employs the mutation operator ❼. This modifies the mth coordinate of the ith krill, doing so via the following formula:

$$x_{i,m} = \begin{cases} x_{gbest,m} + \mu(x_{p,m} - x_{q,m}) & \text{for} \quad \gamma \le Mu \\ x_{i,m} & \text{for} \quad \gamma > Mu \end{cases}, \tag{23}$$

wherein $Mu = 0.05/\hat{K}_{i,best}$; $p, q \in \{1, 2, ..., i-1, i+1, ..., P\}$ and $\mu \in [0, 1)$.

This operation completes the evolutionary procedures. Subsequently, we can now obtain individuals that are readily utilizable within the next iteration. In so-doing, in the last stage ❽ of the main loop, we should calculate the cost function for all the swarm members. Herein, the algorithm's stop condition ❿ decides whether the next iteration or the optimization algorithm is to be completed. The form of stop condition could be that of a time limit, or the reaching of a desired fitness level or a combination of these two.

More information about KHA can be found in [19]. Regarding KHA parameters, the tuning of the KHA is described in publications: [21], [22] and [23]. Notably articles [24] and [25] include other proposed modifications of the algorithm. The KHA procedure has been verified positively in discrete optimization tasks [26]. Furthermore, a parallel version of this algorithm can be found in [27]. It should also be underlined that this heuristic procedure has been applied in data base domains [28], medical tasks [29], in mechanism and machine theory [30], and also in neural learning process [31], e.t.c.

III. Clustering and Selected Clustering Indices

In this section a fusion of KHA with a variety of clustering task assessment methods is to be presented. The stated validation methods are based on characterisation indexes.

Consider a Y as being a data set matrix with dimensions D and M, respectively

$$Y = [y_1, \ldots, y_M]. \tag{24}$$

Herein, each data set element is represented by one column of this matrix. Moreover, the D feature describes each data item. The goal of the clustering task is to devise the particular division of the data set (24) into the individual C subsets, including the assignment of individual elements $y_1, \ldots, y_M$ to clusters $CL_1, \ldots, CL_C$,. In such process, as a rule, the number of clusters C is considerably smaller than the cardinality of set Y, i.e. $C \ll M$.

Individual clusters, along with their associated elements of the set Y, are characterized by points deemed the centroid of clusters $O = O_1, \ldots, O_C$. Each of these is calculated as:

$$O_c = \frac{1}{\#CL_c} \sum_{y_i \in CL_c} y_i, \tag{25}$$

where $\#CL_c$ denotes the number of elements assigned to the cth cluster. In a similar way, the center of gravity for all the investigated elements (24) is defined:

$$O_Y = \frac{1}{M} \sum_{i=1}^{M} y_i. \tag{26}$$

In this paper, the assignment of individual elements of the data set Y (24), to the clusters, is made through employing the KHA procedure. In undertaking this, krills are encoded as vectors that contain the centroid of clusters O_c. In this case, the number of clusters is established in advance, and the grouping of the individual elements of the data set is made on the basis of the rules of the nearest centroid. Thus, for

each point y_i (for $i = 1, \ldots, M$) the distance to each cluster centre O_c is calculated. In so-doing, the ith element belongs to cluster CL_c if the distance $dist(y_i, O_c)$ is the smallest of the tested distances.

Furthermore, the division of elements of the set Y, is evaluated in such a way as to minimize the cost function (1). This aspect is individually determined for each of the clustering index. The formula pertaining to individual functions will be described later in this work.

A. Rand Index

The Rand Index is the first in the sequence that will be presented. This is considered to be a so-called supervised method for validating clustering procedures. To use this index, it is assumed that the reference distribution of membership of individual elements of the data set Y with regard to the pertinent cluster, is known to be similar as that in the case of handling the data for the classification task. This index is expressed as:

$$I_R = \frac{a+d}{a+b+c+d}, \tag{27}$$

where a is the number of elements placed in the same reference group that in the cluster grouping, b denotes the number of elements that are placed in the reference group and in the different cluster sets, c defines the number of elements placed in other reference groups and in the same cluster, and, finally, d indicates the number of elements placed in the different reference groups and in the different cluster's groups.

Building upon above definition, it can be observed that I_R can yields value between 0 and 1. Furthermore, its maximum value points out the degree of full compliance of the clustering division result, with a reference set. In the reported studies, this index is employed for comparing the division by way of applying the clustering procedure that is based on KHA, with the division arising from the structure of the reference data (i.e. the label of classes). With this index, it is possible to compare the obtained results with the reference data, as well as with other clustering indices applied in the optimization cost function.

More information about the Rand Index can be found at [20], [32].

B. Calinski-Harabasz Index

The following indexes are designated as being unsupervised methods for validating clustering procedures. In such, the assessment of the quality of the division stems from the properties of the dataset and the individual clusters. Consequently, such induces, in terms of measuring ability, can be utilized within the evaluation function (1) at the KHA stage.

The Celinski-Harabasz criterion has its foundation within the concept of data set variance. This index is defined as:

$$I_{CH} = \frac{V_B}{V_W}\frac{M-C}{C-1}, \tag{28}$$

where V_B and V_W denote overall between-cluster and within-cluster variance respectively. These are calculated according to the following formulas:

$$V_B = \sum_{c=1}^{C} \#CL_c \|O_c - O_Y\|^2, \tag{29}$$

and

$$V_W = \sum_{c=1}^{C} \sum_{y_i \in CL_c} \|y_i - O_c\|^2, \tag{30}$$

here, $\|\cdot\|$ is the L^2 norm (Euclidean distance) between the two vectors.

It must be underlined that high values of Celinski-Harabasz Index designate well-defined partitions. More information about this index can be found at [33].

C. Davies-Bouldin Index

The Davies-Bouldin Index is one of the more commonly utilized unsupervised evaluations of clustering results criteria. This function consists of a ratio of within-clustering and between-clustering distances. This index is described via:

$$I_{DB} = \frac{1}{C}\sum_{c=1}^{C} \max_{c \neq p}\{D_{c,p}\}, \tag{31}$$

where $D_{c,p}$ denotes within-to-between cluster distance for the cth and pth cluster

$$D_{c,p} = \frac{\overline{d_c} + \overline{d_p}}{d_{c,p}}. \tag{32}$$

In (32) notation $\overline{d_p}$ designates the average distance between each element of the pth cluster and centre point of this group. Moreover $d_{c,p}$ is the distance between the centres of the cth and pth clusters. In this case,the smallest value of the Davies-Bouldin Index delineates a well-defined clustering solution. More information about this measure is obtainable in [34].

D. Silhoutte Value Index

The Silhouette Value Index (SH) is the last clustering index to be dealt within this part of this paper. Herein, for each ith point of data set Y, the distance between all points in the same cluster and the separation distance presented by the nearest neighbours, are calculated. This criteria is defined as follow:

$$I_{SV} = \frac{1}{M}\sum_{c=1}^{C} \sum_{y_i \in CL_c} \frac{b(i,c) - a(i,c)}{max(a(i,c), b(i,c))}. \tag{33}$$

Here, $a(i,c)$ describes the mean distance of the ith point to other points in the same cluster CL_c, while $b(i,c)$ represents a minimum of average distance from the ith point in cluster pth to points in other clusters. These values are obtained through the following formulas:

$$a(i,c) = \frac{1}{\#CL_c} \sum_{y_j \in CL_c \& j \neq i} dist(y_i, y_j), \tag{34}$$

and

$$b(i,c) = \min_{CL_l \in C \setminus CL_c} \frac{1}{\#CL_l} \sum_{y_j \in CL_l} dist(y_i, y_j). \quad (35)$$

For a single ith data point, a high value of the component of this criteria denotes that this element y_i is well-matched to its group, and, simultaneously is weakly-match to other clusters. What is interesting, is that a low value of I_{SV} index reveals that the number of clusters is overestimated.

By way of formulas (33)-(35), one can observe that this criteria yields a value between -1 and $+1$. Of note: a well-defined clustering solution is represented by a value close to 1.

More information concerning the SH Clustering Index can be found in [35], [36]. With regard to clustering quality measures as a whole, more information is obtainable in [37], [38].

IV. Numerical results

This section is intended to inform the reader of several numerical verification procedures that are useful in assessing the quality of the proposed clustering methods. In order to verify the quality of the clustering algorithm, 12 sets of data obtained from the UCI Machine Learning Repository were taken into consideration [39]. With regard to these, Table I provides a characterization of all the data sets that were applied in generating a numerical verification within this paper. Evident in this table is that it includes names, abbreviations, numbers of items, dimensionality, number of classes and references to the description of the presented data sets. Herein, synthetic data collection is placed within the first four rows. These data sets are two-dimensional, and, therefore, they serve as being very good explanatory examples upon which Figure 2 is outlined.

In the presented approach, the vector of cluster centre represents the solution in state space for KHA. Thus, the product value $D{\cdot}C$ expresses the dimensionality of a particular optimization task.

In this work, a quite difficult task that the researcher must undertake is to determine a suitable set of KHA parameters. Thus, for several data sets, pilot-tests are calculated. In each test, one parameter of the KHA optimisation procedure is made variable. In addition, in these studies, the CH Index is applied as a validation parameter. As a result of this research, it is found that for almost all data set cases, the same suboptimal sets with best parameters values are calculated. Indeed, it has been discovered that it is only in the case of the Sonar and Ionosphere data sets that the achieved parameters differ. The reason for this is thought to be the higher dimension of these datasets. The following parameters of KHA were established after pilot-tests:

- $P = 20$,
- $K^{max} = 200$,
- $N^{max} = 0.01$,
- $\omega_n = 0.5$,
- $V_f = 0.02$,
- $D^{max} = 0.01$,
- $C_t = 0.5$.

Each clustering test is made of only 200 iterations of the KHA optimization procedure. For this task, three clustering indexes CH, DB and SV are employed, and these validity measures are applied in assessing the value of the cost function (1) for KHA. Because of their different properties (described in Section III), for the indices used here, the following forms of cost functions are formulated

$$f_{CH} = \frac{1}{I_{CH}} + \#CL_{\text{empty}}, \quad (36)$$

$$f_{DB} = 2I_{DB} + \#CL_{empty}, \quad (37)$$

and, finally,

$$f_{SV} = \frac{1}{I_{SH} + 1.01} + \#CL_{empty}. \quad (38)$$

In the investigation presented here, it is assumed that, firstly, a clustering procedure based on KHA is performed by way of one selected index at a time. The result of this experiment is the clustering of the explored data set. In the next step of the test, the Rand Index calculated versus class labels is employed, as this is a commonly used evaluator of clustering performance. Thus, the obtained KHA optimization procedure solution is compared with the reference label of the class (cluster) which came from the data set. Additionally, for comparison purposes, outcomes from utilizing the k-means algorithm are also reported (with corresponding Rand Index values). Results generated by means of aforementioned steps can be seen in Table II. Throughout the testing runs, both KHA-based clustering procedures, as well as the k-means clustering algorithm were performed 30 times.

Table II consists of two parts. The first incorporates the 2nd and 3rd columns, and it contains the mean values $\overline{R}$ and the standard deviations σ_R of the Rand Index that was obtained while using the k-means clustering function. The second part of the table lists the Rand Index results (as in the first part). However, these were obtained by the way of following the KHA-clustering procedure. Here, each of three subparts provides the application results for Celinski-Harabasz ($\overline{R_{CH}}$ and $\sigma_{R_{CH}}$), Davies-Bouldin ($\overline{R_{DB}}$ and $\sigma_{R_{DB}}$) and Silhoutte Value ($\overline{R_{SV}}$ and $\sigma_{R_{SV}}$) Indexes, respectively.

While comparing all the obtained results, it can be seen that it is only in the case of the ION data set when Rand Index of clustering that was performed with k-means procedure achieves better quality then the one attained by the application of the KHA-clustering procedure. In all other cases, the results obtained via the KHA clustering method yield much better evaluation notes. These cases in Table II are emphasized with a bold font.

Based on presented results, one can observe that the Celinski-Harabasz Index clustering validation measure proved to be the best evaluation index applicable in metaheuristic procedures used in clustering. However, with regard to the other indexes, the results generated by way of the Davies-Bouldin Index are better than that obtained via the k-means

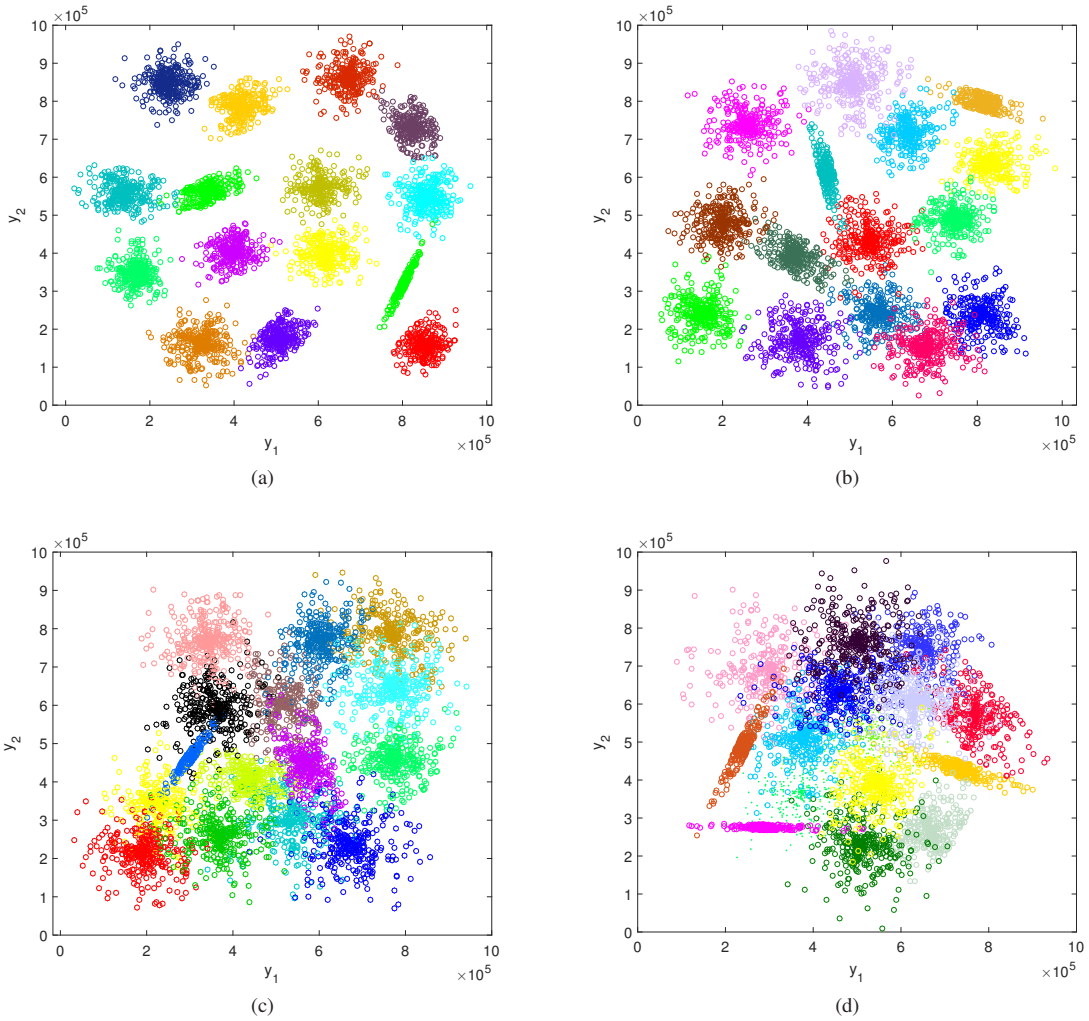


Fig. 2: Plots of 2 dimensional *s1* (a), *s2* (b), *s3* (c) and *s4* (d) datasets

TABLE I: Data sets used for experimental verification

Name of Data set	Abreviation	Number of			Bibliographical
	in paper	elements (M)	features (D)	classes (C)	reference
Synthetic 1	S1	5000	2	15	[40]
Synthetic 2	S2	5000	2	6	[40]
Synthetic 3	S3	5000	2	3	[40]
Synthetic 4	S4	5000	2	6	[40]
Ionosphere	ION	351	34	2	[41]
Iris	Iris	150	4	3	[42]
Seeds	Seeds	210	7	3	[43]
Sonar	SON	208	60	2	[44]
Thyroid	TH	7200	21	3	[45], [46]
Vehicle	VH	846	18	4	[47]
Wisconsin Breast Cancer	WBC	683	10	2	[48]
Wine	Wine	178	13	3	[49]

TABLE II: Results summary

Data set	k–means clustering $\overline{R}$	σ_R	KHA clustering $\overline{R_{CH}}$	$\sigma_{R_{CH}}$	$\overline{R_{DB}}$	$\sigma_{R_{DB}}$	$\overline{R_{SV}}$	$\sigma_{R_{SV}}$
S1	0.9748	0.0093	**0.9782**	0.0078	0.9200	0.0138	**0.9775**	0.0090
S2	0.9760	0.0072	**0.9839**	0.0053	0.9610	0.0128	0.9664	0.0096
S3	0.9522	0.0072	**0.9548**	0.0053	0.9138	0.0177	0.9436	0.0093
S4	0.9454	0.0056	**0.9484**	0.0048	0.8942	0.0229	0.9388	0.0080
Iris	0.8458	0.0614	**0.8872**	0.0145	0.7846	0.0099	0.8321	0.0563
Ionosphere	**0.5945**	0.0004	0.5573	0.0124	0.5393	0.0239	0.5682	0.0090
Seeds	0.8573	0.0572	**0.8709**	0.0156	0.6234	0.0952	0.8341	0.0586
Sonar	0.5116	0.0016	**0.5145**	0.0078	**0.5196**	0.0015	**0.5151**	0.0022
Vehicle	0.5843	0.0359	**0.6076**	0.0194	0.4854	0.0342	0.5192	0.0157
WBC	0.5448	0.0040	**0.5456**	0.0000	**0.5465**	0.0002	**0.5456**	0.0000
Wine	0.7167	0.0135	**0.7257**	0.0073	0.3979	0.0378	0.6708	0.0084
Thyroid	0.5844	0.0982	0.4535	0.0339	**0.8148**	0.1007	**0.8423**	0.0757

algorithm in only three of the applications, and that of the Silhouette Value Index, four.

Looking closely at all the results obtained by way of an application of the KHA procedure, it can be stated that for data collections S1, S2, S3, S4, Iris, Seeds, VH and Wine, the employment of the Celinski-Harabasz Index as a part of the cost function in KHA-clustering procedure gives the best results. Similarly, for the data set SON, applying the Davies-Bouldin Index, and, for the TH data set, using the Silhouette Value Index, yield the best result. Furthermore, in the situation of tests with use the WBC data collection, clusterings incorporating all three indexes provide the same result.

V. Summary

This paper is a presentation of research describing various clustering methods based on metaheuristic procedures and several validation measures. Here, in optimizing the cluster centroid locations, the biologically-inspired KHA procedure was employed. For the evaluation of particular KHA generated solutions, the paper assessed the quality of using Celinski-Harabasz, Davies-Bouldin and Silhouette Value Indexes as three clustering variants. Moreover, the Rand Index was calculated so as to evaluate the quality of the derived solutions of the analyzed clustering procedures. The proposed algorithm, in its three versions, was also confronted via the application of the well-known and commonly enrolled k-means method.

As a result of the study, it was established that the results obtained via the KHA-clustering method are much better than for that which were generated via k-means clustering procedure. What is more, the Celinski-Harabasz Index, as well as the KHA-clustering method, qualify for being considered superior for clustering tasks.

Future research will be targeted on deeper analysis of new clustering quality validation methods, as well as on applying the new procedures of swarm intelligence to the task of clustering.

References

[1] P. Kulczycki, M. Charytanowicz, P. A. Kowalski, and S. Lukasik, "The complete gradient clustering algorithm: properties in practical applications," 2012.

[2] P. A. Kowalski, S. Łukasik, M. Charytanowicz, and P. Kulczycki, "Data-driven fuzzy modeling and control with kernel density based clustering technique," *Polish Journal of Environmental Studies*, vol. 17, pp. 83–87, 2008.

[3] S. Łukasik, P. Kowalski, M. Charytanowicz, and P. Kulczycki, "Fuzzy models synthesis with kernel-density-based clustering algorithm," in *Fuzzy Systems and Knowledge Discovery, 2008. FSKD '08. Fifth International Conference on*, vol. 3, Oct 2008. doi: 10.1109/FSKD.2008.139 pp. 449–453.

[4] S. Breschi and F. Malerba, "The geography of innovation and economic clustering: some introductory notes," *Industrial and corporate change*, vol. 10, no. 4, pp. 817–833, 2001.

[5] M. Charytanowicz, J. Niewczas, P. Kulczycki, P. A. Kowalski, S. Łukasik, and S. Żak, "Complete gradient clustering algorithm for features analysis of x-ray images," in *Information Technologies in Biomedicine*, ser. Advances in Intelligent and Soft Computing, E. Piętka and J. Kawa, Eds. Springer Berlin Heidelberg, 2010, vol. 69, pp. 15–24. ISBN 978-3-642-13104-2. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-13105-9_2

[6] L. Rokach and O. Maimon, "Clustering methods," in *Data Mining and Knowledge Discovery Handbook*, O. Maimon and L. Rokach, Eds. Springer US, 2005, pp. 321–352. ISBN 978-0-387-24435-8. [Online]. Available: http://dx.doi.org/10.1007/0-387-25465-X_15

[7] P. Langfelder, B. Zhang, and S. Horvath, "Defining clusters from a hierarchical cluster tree: the dynamic tree cut package for r," *Bioinformatics*, vol. 24, no. 5, pp. 719–720, 2008.

[8] I. Davidson and S. Ravi, "Agglomerative hierarchical clustering with constraints: Theoretical and empirical results," in *Knowledge Discovery in Databases: PKDD 2005*. Springer, 2005, pp. 59–70.

[9] S. M. Savaresi, D. L. Boley, S. Bittanti, and G. Gazzaniga, "Cluster selection in divisive clustering algorithms." in *SDM*. SIAM, 2002, pp. 299–314.

[10] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proc. 5th Berkeley Symp. Math. Stat. Probab., Univ. Calif. 1965/66*, 1967, pp. 281–297.

[11] M.-S. Yang, "A survey of fuzzy clustering," *Mathematical and Computer modelling*, vol. 18, no. 11, pp. 1–16, 1993.

[12] J. C. Bezdek, R. Ehrlich, and W. Full, "Fcm: The fuzzy c-means clustering algorithm," *Computers & Geosciences*, vol. 10, no. 2, pp. 191–203, 1984.

[13] P. Kulczycki and M. Charytanowicz, *A Complete Gradient Clustering Algorithm*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 497–504. ISBN 978-3-642-23896-3. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-23896-3_61

[14] C. Warnekar and G. Krishna, "A heuristic clustering algorithm using union of overlapping pattern-cells," *Pattern Recognition*, vol. 11, no. 2, pp. 85 – 93, 1979. doi: http://dx.doi.org/10.1016/0031-3203(79)90054-2. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0031320379900542

[15] C. C. Aggarwal and C. K. Reddy, *Data clustering: algorithms and applications*. CRC Press, 2013.

[16] C.-W. Tsai, W.-C. Huang, and M.-C. Chiang, "Recent development of metaheuristics for clustering," in *Mobile, Ubiquitous, and Intelligent Computing*, ser. Lecture Notes in Electrical Engineering, J. J. J. H. Park, H. Adeli, N. Park, and I. Woungang, Eds. Springer Berlin Heidelberg, 2014, vol. 274, pp. 629–636. ISBN 978-3-642-40674-4. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-40675-1_93

[17] T. Niknam and B. Amiri, "An efficient hybrid approach based on pso, {ACO} and k-means for cluster analysis," *Applied Soft Computing*, vol. 10, no. 1, pp. 183 – 197, 2010. doi: http://dx.doi.org/10.1016/j.asoc.2009.07.001. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1568494609000854

[18] W. J. Welch, "Algorithmic complexity: three np- hard problems in computational statistics," *Journal of Statistical Computation and Simulation*, vol. 15, no. 1, pp. 17–25, 1982. doi: 10.1080/00949658208810560. [Online]. Available: http://dx.doi.org/10.1080/00949658208810560

[19] A. H. Gandomi and A. H. Alavi, "Krill herd: A new bio-inspired optimization algorithm," *Communications in Nonlinear Science and Numerical Simulation*, vol. 17, no. 12, pp. 4831–4845, 2012. doi: 10.1016/j.cnsns.2012.05.010. [Online]. Available: http://dx.doi.org/10.1016/j.cnsns.2012.05.010

[20] H. Parvin, H. Alizadeh, and B. Minati, "Objective criteria for the evaluation of clustering methods," *Journal of the American Statistical Association*, vol. 66, pp. 846–850, 1971.

[21] P. A. Kowalski and S. Łukasik, "Experimental study of selected parameters of the krill herd algorithm," in *Intelligent Systems'2014*. Springer Science Business Media, 2015, pp. 473–485. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-11313-5_42

[22] G. P. Singh and A. Singh, "Comparative study of krill herd, firefly and cuckoo search algorithms for unimodal and multimodal optimization," *IJISA*, vol. 6, no. 3, pp. 35–49, 2014. doi: 10.5815/ijisa.2014.03.04. [Online]. Available: http://dx.doi.org/10.5815/ijisa.2014.03.04

[23] P. K. Adhvaryyu, P. K. Chattopadhyay, and A. Bhattacharjya, "Application of bio-inspired krill herd algorithm to combined heat and power economic dispatch," in *2014 IEEE Innovative Smart Grid Technologies - Asia*. IEEE, 2014. doi: 10.1109/isgt-asia.2014.6873814. [Online]. Available: http://dx.doi.org/10.1109/isgt-asia.2014.6873814

[24] L. Guo, G.-G. Wang, A. H. Gandomi, A. H. Alavi, and H. Duan, "A new improved krill herd algorithm for global numerical optimization," *Neurocomputing*, vol. 138, pp. 392–402, 2014. doi: 10.1016/j.neucom.2014.01.023. [Online]. Available: http://dx.doi.org/10.1016/j.neucom.2014.01.023

[25] G.-G. Wang, A. H. Gandomi, and A. H. Alavi, "Stud krill herd algorithm," *Neurocomputing*, vol. 128, pp. 363–370, 2014. doi: 10.1016/j.neucom.2013.08.031. [Online]. Available: http://dx.doi.org/10.1016/j.neucom.2013.08.031

[26] G.-G. Wang, S. Deb, and S. M. Thampi, *Intelligent Systems Technologies and Applications: Volume 1*. Cham: Springer International Publishing, 2016, ch. A Discrete Krill Herd Method with Multilayer Coding Strategy for Flexible Job-Shop Scheduling Problem, pp. 201–215. ISBN 978-3-319-23036-8. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-23036-8_18

[27] A. Nowosielski, P. A. Kowalski, and P. Kulczycki, "Increasing the Speed of the Krill Herd Algorithm through Parallelization," in *Information Technology, Computational and Experimental Physics*. AGH University of Science and Technology Press, 2016, pp. 117–120. ISBN 978-83-7464-838-7

[28] ——, "The column-oriented database partitioning optimization based on the natural computing algorithms," in *2015 Federated Conference on Computer Science and Information Systems, FedCSIS 2015, Łódź, Poland, September 13-16, 2015*, 2015. doi: 10.15439/2015F262 pp. 1035–1041. [Online]. Available: http://dx.doi.org/10.15439/2015F262

[29] A. Mohammadi, M. S. Abadeh, and H. Keshavarz, "Breast cancer detection using a multi-objective binary krill herd algorithm," in *Biomedical Engineering (ICBME), 2014 21th Iranian Conference on*, Nov 2014. doi: 10.1109/ICBME.2014.7043907 pp. 128–133.

[30] R. R. Bulatović, G. Miodragović, and M. S. Bošković, "Modified krill herd (mkh) algorithm and its application in dimensional synthesis of a four-bar linkage," *Mechanism and Machine Theory*, vol. 95, pp. 1 – 21, 2016. doi: http://dx.doi.org/10.1016/j.mechmachtheory.2015.08.004. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0094114X15001895

[31] P. Kowalski and S. Łukasik, "Training neural networks with krill herd algorithm," *Neural Processing Letters*, 2015. doi: 10.1007/s11063-015-9463-0

[32] E. Achtert, S. Goldhofer, H. P. Kriegel, E. Schubert, and A. Zimek, "Evaluation of clusterings – metrics and visual support," in *2012 IEEE 28th International Conference on Data Engineering*, April 2012. doi: 10.1109/ICDE.2012.128. ISSN 1063-6382 pp. 1285–1288.

[33] T. Caliński and J. Harabasz, "A dendrite method for cluster analysis," *Communications in Statistics-theory and Methods*, vol. 3, no. 1, pp. 1–27, 1974.

[34] D. L. Davies and D. W. Bouldin, "A cluster separation measure," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-1, no. 2, pp. 224–227, April 1979. doi: 10.1109/TPAMI.1979.4766909

[35] L. Kaufman and P. J. Rousseeuw, *Finding groups in data: an introduction to cluster analysis*. John Wiley & Sons, 2009, vol. 344.

[36] P. J. Rousseeuw, "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis," *Journal of computational and applied mathematics*, vol. 20, pp. 53–65, 1987.

[37] O. Arbelaitz, I. Gurrutxaga, J. Muguerza, J. M. Pérez, and I. Perona, "An extensive comparative study of cluster validity indices," *Pattern Recognition*, vol. 46, no. 1, pp. 243–256, 2013.

[38] J. Demšar, "Statistical comparisons of classifiers over multiple data sets," *The Journal of Machine Learning Research*, vol. 7, pp. 1–30, 2006.

[39] M. Lichman, "UCI machine learning repository," 2013. [Online]. Available: http://archive.ics.uci.edu/ml

[40] P. Fränti and O. Virmajoki, "Iterative shrinking method for clustering problems," *Pattern Recognition*, vol. 39, no. 5, pp. 761 – 775, 2006. doi: http://dx.doi.org/10.1016/j.patcog.2005.09.012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0031320305003778

[41] V. G. Sigillito, S. P. Wing, L. V. Hutton, and K. B. Baker, "Classification of radar returns from the ionosphere using neural networks," *Johns Hopkins APL Technical Digest*, vol. 10, no. 3, pp. 262–266, 1989.

[42] P. A. Kowalski and P. Kulczycki, "Interval probabilistic neural network," *Neural Computing and Applications*, pp. 1–18, 2015. doi: 10.1007/s00521-015-2109-3. [Online]. Available: http://dx.doi.org/10.1007/s00521-015-2109-3

[43] M. Charytanowicz, J. Niewczas, P. Kulczycki, P. A. Kowalski, S. Lukasik, and S. Zak, "Complete gradient clustering algorithm for features analysis of x-ray images," in *Information Technologies in Biomedicine*, ser. Advances in Intelligent and Soft Computing, E. Pietka and J. Kawa, Eds. Springer Berlin Heidelberg, 2010, vol. 69, pp. 15–24. ISBN 978-3-642-13104-2. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-13105-9_2

[44] R. P. Gorman and T. J. Sejnowski, "Analysis of hidden units in a layered network trained to classify sonar targets," *Neural networks*, vol. 1, no. 1, pp. 75–89, 1988.

[45] J. R. Quinlan, "Induction of decision trees," *Machine learning*, vol. 1, no. 1, pp. 81–106, 1986.

[46] J. R. Quinlan, P. J. Compton, K. Horn, and L. Lazarus, "Inductive knowledge acquisition: a case study," in *Proceedings of the Second Australian Conference on Applications of expert systems*. Addison-Wesley Longman Publishing Co., Inc., 1987, pp. 137–156.

[47] R. Setiono and W. Leow, "Vehicle recognition using rule based methods," *Turing Institute Research Memorandum TIRM-87-018*, vol. 121, 1987.

[48] J. Zhang, "Selecting typical instances in instance-based learning," in *Proceedings of the ninth international conference on machine learning*, 1992, pp. 470–479.

[49] S. Aeberhard, D. Coomans, and O. De Vel, "Comparison of classifiers in high dimensional settings," *Dept. Math. Statist., James Cook Univ., North Queensland, Australia, Tech. Rep*, no. 92-02, 1992.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 89–95
DOI: 10.15439/2016F552
ACSIS, Vol. 8. ISSN 2300-5963

Forming Classifier Ensembles with Deterministic Feature Subspaces

Michał Koziarski*, Bartosz Krawczyk* and Michał Woźniak*
*Department of Systems and Computer Networks
Faculty of Electronics
Wrocław University of Science and Technology
Wrocław, Poland
Email: {michal.koziarski, bartosz.krawczyk, michal.wozniak}@pwr.edu.pl

***Abstract*—Ensemble learning is being considered as one of the most well-established and efficient techniques in the contemporary machine learning. The key to the satisfactory performance of such combined models lies in the supplied base learners and selected combination strategy. In this paper we will focus on the former issue. Having classifiers that are of high individual quality and complementary to each other is a desirable property. Among several ways to ensure diversity feature space division deserves attention. The most popular method employed here is Random Subspace approach. However, due to its random nature one cannot consider this approach as stable one or suitable for real-life applications. Therefore, we propose a new approach called Deterministic Subspace that constructs feature subspaces in a guided and repetitive manner. We present a general framework and three dedicated measures that can be used for selecting diverse and uncorrelated features for each base learner. This way we will always obtain identical sets of features, leading to creation of stable ensembles. Experimental study backed-up with statistical analysis prove the usefulness of our method in comparison to popular randomized solution.**

***Index Terms*—Machine learning, ensemble classification, feature subspaces, diversity, deterministic methods.**

I. Introduction

CONTEMPORARY machine learning deals with the ever-increasing complexity of problems, directly connected to the era of big data and data flood. Standard classifiers cannot properly capture the properties of analyzed data or by trying to do so finally become subject to the overfitting process. Therefore, methods that can take advantage of combining several learners to get at the same time advantages of complex decision boundary and simplified models are of high interest for both researchers and practitioners. Such approaches are known as ensemble learning, multiple classifier systems or classifier committees [1] and are being considered as one of the most efficient tools to handle pattern analysis process. It is assumed that we have at our disposal a number of models and combine their predictions in order to get a more efficient recognition system, as a set of weak models may overcome the limitations of using a single strong one.

For the ensemble to work efficiently one needs to supply a pool of diverse classifiers [2]. The diversity itself can be ensured on several different levels. One of the most popular is to train each learner on the basis of different features, in hope that such an embedding into lower dimensions will at the same time simplify the training procedure and allow classifiers to explore different properties of supplied feature space [3]. Random Subspace (RS) [4] is the most popular implementation of this paradigm. This method assumes that each base learner is trained with a subset of randomly selected features, with the assumption that feature subsets may overlap. This provides simple, but efficient way of managing the diversity in the ensemble. However, there exist a significant drawback of this method, rooted in its randomized nature. Due to lack of any guidance when creating feature subspaces we obtain different sets for each run (e.g., when using cross-validation or repeating experiments). This significantly limits the usability of RS in real-life applications as we are not sure exactly what kind of model we should be using and how stable it is.

To overcome this limitation we introduce a novel approach for forming classifier ensembles based on feature subset named Deterministic Subspace (DS). It is based on the same idea as RS, namely creating a pool of diverse classifiers on the basis of reduced number of features. However, we remove the randomization from it and replace it with a fully guided search approach that guarantee a high stability and repetitiveness of the entire procedure. To obtain a number of equally-sized subspaces we propose a set of metrics dedicated to evaluating the discriminative power of each separate feature and the diversity among created subspaces. A round robin strategy is being employed to evenly distribute features with greedy approach. This leads to the creation of deterministic feature subsets of suitable discriminative power that lead to a creation of efficient ensemble in a guided and deterministic manner. Finally, the created classifiers are combined using a majority voting strategy. The proposed method is as flexible as original RS approach and can work with any kind of base learner.

The main contributions of this works are as follow:

- Novel Deterministic Subspace method for forming classifier ensembles.
- Set of metrics suitable for evaluating the quality and diversity of features being used.
- Greedy round robin approach for creating subspaces of evenly distributed features.
- Thorough experimental evaluation of the proposed approach backed-up by statistical tests.

978-83-60810-90-3/$25.00

The rest of the paper is organized as follows. Next section gives the necessary background in recent advances in ensemble classification. Section III gives the full details regarding the proposed DS method. Section IV, while the final section concludes the paper.

II. Related works on ensemble classification

Ensemble classifiers have several desirable properties for the process of pattern classification system design:

- Ensemble techniques allow to exploit local competencies of base learners, thus leading to a potential gain in accuracy of the combined system.
- Due to their structure they are highly flexible methods that could be easily adjusted by the user according to specific needs.
- They prevent us from selecting the worst model from the pool.
- They are easy to implement in parallel and distributed high-performance commuting environments.

These properties made them highly regarded approaches for a variety of tasks including classification, regression and clustering. In this paper we will concentrate on their usage in supervised classification.

There are three major issues in designing multiple classifier systems:

- How to create a pool of classifiers characterized by a high individual accuracy and diversity [5].
- How to chose the topology of the ensemble.
- How to efficiently combine the outputs of individual classifiers in order to obtain better final predictions [6].

We will focus on the first issue, as this paper deals with the problem on how to form an efficient pool individuals. For such a group of learners to work well we must ensure that they display differing characteristics, as adding similar or identical models to the pool would not contribute to the quality of the ensemble, but would only increase its computational complexity.

There are three main approaches for introducing diversity into a pool of classifiers, depending whether we work with heterogeneous or homogeneous models:

- Different learning algorithms (different models or different parameters for the same model).
- Different inputs (training base classifiers on different data set partitions or choosing different attributes during training) .
- Different outputs (decompose the classification task e.g. into binary tasks).

Heterogeneous ensembles assume that using varying classifier models is enough to properly diversify the pool. By applying different learning paradigms we will get varying decision boundaries that combined together may promote their strong sides, while reducing their weaknesses. However, in some situations different classifiers may return similar or identical boundaries, thus making the selection process crucial. An entire family of dynamic classifier and ensemble selection methods deserves mentioning, as they offer a flexible ensemble line-up for each incoming sample [7].

However, when the considered pool is homogeneous one needs to find an alternative way of introducing diversity. Input manipulation for each base classifier is the most straightforward approach. One can either work in the data or feature space. Former approach assumes that we introduce variance into training instances in order for classifiers to capture properties of different subsets of objects. Here Bagging [8] and Boosting [9] are the most popular solutions, but one can also train ensembles on the basis of clusters to preserve spatial relations among instances [10]. Latter solution introduces diversity by splitting the feature space. This can be done in either randomized manner [11] or in a guided way with the usage of feature selection [12] or global optimization methods [13].

Alternatively one may manipulate the outputs of classifiers in order to get a diverse set of learners. Here the most common solution is a multi-class decomposition, where a divide-and-conquer approach is being used to obtain simplified learners specialized on a reduced number of classes. Then a dedicated combination method like Error-Correcting Output Codes is being used to reconstruct the original multi-class task [14]. Two approaches deserving mentioning are binarization (in one-vs-one or one-vs-all manner) [15] and hierarchical decomposition [16].

Finally when using the same model one may initialize it with different or randomized parameters in hope that training process for each method will return diverse classifiers. This approach is based on the assumption of a complex search space during the classifier training procedure, as multiple starting points could end up in different local extrema, thus offering better coverage of the considered problem. Most popular examples are prematurely stopped neural networks or ensembles of Support Vector Machines with varying kernels [17].

III. Deterministic feature subspace approach

Random Subspace method, while being hugely successful approach for ensemble classification [4], does not avoid some of the pitfalls associated with methods randomly selecting their inputs. Specifically, randomly selected subspaces may lack the discriminant power necessary for proper separation of different classes, which in turn can harm performance of whole ensemble. Additionally, even if individually strong subspaces are produced, we still do not have any guarantees on their diversity. In this section we present a novel approach of creating feature subspaces in a fully deterministic way, with the aim of, at least partially, mitigating mentioned issues.

A. Deterministic subspace algorithm

The idea behind DS approach we propose is to assign features evenly between the subspaces, with some preference of individually strong predictors. To achieve this, we employ separate metrics of both individual feature quality and diversification between subspaces. We then greedily assign features to subspaces using round robin strategy, based on weighted

average of two above metrics. Pseudocode for the proposed method is presented in Algorithm 1.

Algorithm 1 Deterministic Subspace algorithm

```
create k empty subspaces

repeat
  for all subspaces do
    for all features not in current subspace do
      score = α feature quality + (1 − α) diversity
    end for
    add feature with highest score
  end for
until every subspace has n features

return subspaces
```

B. Subspace diversity measure

Many different approaches of measuring diversity in classifier ensembles exist in the literature. Most of them, however, are fairly computationally expensive, as they usually require to train classifiers and compute predictions [18]. We instead propose naive but fast approach of measuring evenness of feature spread between subspace.

Let us denote the set of created so far subspaces by S and current subspace, for which we are considering selecting additional feature f, by S_c. We define metric of diversity d as an average of two components: ratio of subspaces already containing considered feature d_f and nearest distance between subspaces after selecting said feature d_s.

$$d_f = 1 - \frac{|\{S_i : f \in S_i\}|}{|S|} \tag{1}$$

$$d_s = 1 - \max_{i \neq c} \frac{|S_c \cap S_i|}{|S_c|} \tag{2}$$

$$d = \frac{d_f + d_s}{2} \tag{3}$$

Using defined above metrics in the framework of deterministic subspace algorithm ensures even spread of features between subspaces. Furthermore, since its values are bound in range from 0 to 1, it naturally extends to the case in which we consider feature quality as well.

C. Feature quality measures

Ideally, we would like to be able to measure the quality of whole subspace considered. It is especially important taking into account the fact that it can be easily shown that multiple weak features can have increased predictive power when combined. However, due to computational considerations such measures tend to be out of our reach: we have to either employ much less demanding search strategy for subspace creation or rely on rough estimations.

Instead, we propose alternate strategy: giving higher preference to individually strong predictors. While it is not true that using individually stronger features must improve quality of subspace, and even more so whole ensemble, we hypothesize that reducing frequency of appearance of particularly weak features may result in increased performance, especially in cases when classifier used is not resistant to being trained on uninformative features. Furthermore, such strategy is relatively inexpensive, requiring only single computation of ranking between features.

During the experimental study, we evaluate performance of three different measures of features predictive power: accuracy on validation set, mutual information between the feature and target vector, and correlation between the two.

TABLE I
DETAILS OF DATASETS USED THROUGHOUT THE EXPERIMENT.

No.	Name	Features	Objects	Classes
1	winequality	11	6497	11
2	vowel	13	990	11
3	vehicle	18	846	4
4	segment	19	2310	7
5	ring	20	7400	2
6	thyroid	21	7200	3
7	mushroom	22	5644	2
8	chronic kidney disease	24	157	2
9	automobile	25	159	6
10	wdbc	30	569	2
11	ionosphere	33	351	2
12	dermatology	34	358	6
13	texture	40	5500	11
14	biodegradation	41	1055	2
15	spectfheart	44	267	2
16	spambase	57	4597	2
17	sonar	60	208	2
18	splice	60	3190	3
19	optdigits	64	5620	10
20	mice protein expression	80	552	8
21	coil2000	85	9822	2
22	movement libras	90	360	15

IV. EXPERIMENTAL STUDY

In this section we present detailed description of conducted experiments, together with obtained results. Our goal was to compare the performance of proposed deterministic method with random subspace approach. We try to assess whether deterministic method can achieve at least as high accuracy without introducing randomness into the algorithm. We also investigate under what conditions, if any, deterministic approach may actually outperform random subspace method.

A. Datasets

During the experiments 22 datasets with varying number of features and objects were used. All datasets were taken from UCI[1] and KEEL[2] repositories and are publicly available for download. Details of datasets used are presented in Table I.

[1]http://archive.ics.uci.edu/ml/datasets.html
[2]http://sci2s.ugr.es/keel/datasets.php

B. Set-up

All experiments were implemented in Python programming language with usage of scikit-learn machine learning library[3]. In particular, all classification algorithms were taken from dedicated scikit-learn modules to ensure correctness of implementation. Repository with remaining code, sufficient to repeat conducted experiments is publicly available[4].

Throughout the experiments, performance of three different classifiers was tested: linear Support Vector Machine, CART decision tree and k-nearest neighbors. For every classification algorithm default parameters, provided in corresponding scikit-learn modules, were used.

During the experiments number of features in single subspace n was fixed at half the total number of features, rounded down. Different numbers of subspaces $k \in \{5, 10, ..., 50\}$ were evaluated for both random and deterministic method. Additionally, parameter values $\alpha \in \{0.0, 0.1, ..., 0.9\}$ were tested for deterministic subspace approach. Three different feature quality metrics were evaluated, namely: 5-fold cross-validation accuracy on training set, mutual information between features and targets, and absolute correlation between the two. Simple majority voting was used in every case, for both methods. All tests were done employing 5×2 cross-validation with combined F-test [19] performed to assess statistical significance of results.

C. Results and discussion

Summary of the obtained results is presented in Figure 1. It shows averaged classification accuracy across different classifiers and measures of feature quality, with baseline accuracy achieved by random subspace method presented for reference. Figure 2 shows the number of datasets on which deterministic subspace achieved statistically significantly better results than random subspace method minus the number of datasets, on which results were significantly worst. Detailed tables showing the number of statistically significantly better and worse results are presented in Appendix A.

Results of experiments indicate a slightly better performance of proposed method against RS approach for small values of $\alpha \in \{0.0, 0.1, ..., 0.5\}$ and CART, k-NN and SVM classifiers, regardless of feature quality metric chosen. Using larger values of α parameter, however, leads to significantly worse results. This indicates that whereas taking into account individual feature quality may have slight positive influence of quality of subspaces, there exist a threshold after which whole ensemble suffers due to lack of diversity.

V. Conclusions and future works

Method of deterministic feature subspace creation was presented and tested throughout this paper. During the experimental study, its performance was evaluated and compared to random subspace approach. Proposed method, on average, achieved slightly better results compared to its random counterpart.

[3]http://scikit-learn.org/stable/

[4]https://github.com/michalkoziarski/DeterministicSubspace

It is worth noting that both individual feature quality and evenness of feature distribution measures are simplified means of estimating subspace quality and diversity of ensemble, respectively. Proposing alternative, computationally feasible metrics presents possible venue of further investigation. Additionally, different search strategies for feature selection could be considered, allowing using established, more computationally expensive metrics at the cost of depth of the search.

Acknowledgment

This work was supported by the Polish National Science Center under the grant no. DEC-2013/09/B/ST6/02264.

Appendix A. Statistical significance tables

Tables containing number of datasets on which proposed deterministic approach achieved either statistically significantly better (indicated with plus sign) or worst (indicated with minus sign) results than random subspace method.

TABLE II
CART classifier, accuracy as feature quality measure.

	α									
k	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
5	+6/-2	+6/-0	+7/-0	+5/-0	+7/-0	+3/-1	+4/-4	+1/-6	+1/-8	+2/-9
15	+3/-0	+1/-0	+6/-0	+4/-0	+3/-0	+1/-3	+0/-7	+0/-9	+0/-9	+0/-11
10	+3/-0	+4/-0	+4/-1	+4/-2	+5/-0	+2/-3	+0/-11	+1/-12	+1/-12	+0/-15
30	+3/-0	+4/-0	+3/-0	+2/-1	+0/-0	+1/-3	+1/-11	+1/-12	+1/-11	+1/-12
25	+3/-1	+3/-0	+5/-1	+4/-1	+4/-0	+3/-2	+1/-8	+1/-12	+1/-11	+1/-12
20	+2/-1	+4/-1	+2/-0	+5/-0	+6/-0	+3/-2	+2/-10	+2/-12	+2/-12	+2/-14
35	+2/-0	+3/-0	+3/-1	+6/-0	+3/-0	+1/-1	+1/-12	+1/-15	+1/-13	+1/-13
40	+1/-1	+0/-1	+3/-0	+0/-2	+2/-0	+2/-6	+1/-10	+1/-10	+1/-13	+1/-14
45	+2/-0	+1/-0	+1/-0	+1/-0	+0/-2	+1/-3	+1/-13	+1/-12	+1/-13	+1/-13
50	+2/-0	+3/-0	+2/-1	+1/-0	+1/-0	+1/-1	+1/-13	+1/-13	+1/-14	+1/-15

TABLE III
CART classifier, correlation as feature quality measure.

	α									
k	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
5	+4/-0	+7/-2	+6/-1	+4/-0	+6/-0	+5/-1	+2/-7	+2/-9	+2/-9	+1/-10
15	+4/-1	+3/-0	+3/-0	+4/-2	+3/-0	+0/-2	+0/-13	+0/-13	+0/-13	+0/-13
10	+4/-1	+4/-1	+4/-0	+6/-0	+4/-0	+2/-4	+1/-9	+1/-12	+1/-13	+1/-13
30	+2/-2	+2/-0	+1/-1	+3/-0	+1/-2	+2/-3	+1/-14	+1/-14	+1/-14	+1/-15
25	+0/-2	+0/-0	+1/-0	+0/-1	+0/-1	+1/-3	+1/-15	+1/-16	+1/-15	+1/-16
20	+2/-1	+4/-0	+2/-0	+3/-0	+1/-0	+1/-4	+1/-13	+1/-14	+1/-16	+1/-16
35	+3/-0	+1/-0	+4/-0	+3/-0	+1/-0	+3/-4	+1/-14	+1/-14	+1/-15	+1/-15
40	+0/-1	+3/-2	+1/-1	+2/-1	+2/-1	+1/-5	+1/-14	+1/-15	+1/-14	+1/-15
45	+0/-0	+0/-0	+0/-0	+1/-0	+0/-0	+1/-3	+1/-12	+1/-15	+1/-15	+1/-15
50	+3/-1	+3/-0	+2/-1	+1/-0	+2/-1	+2/-7	+1/-17	+1/-17	+1/-17	+1/-17

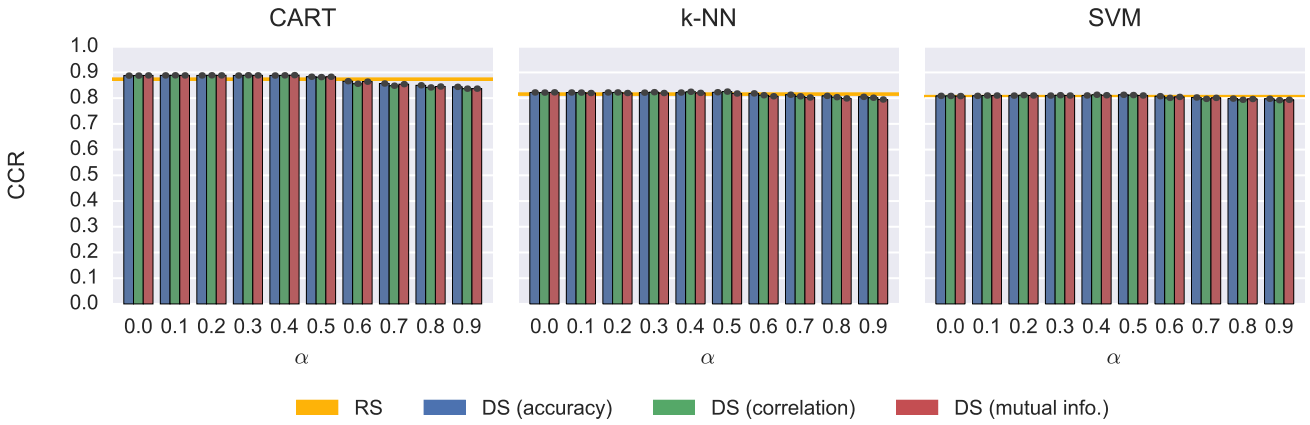


Fig. 1. Correct classification rates averaged over all datasets and examined number of subspaces.

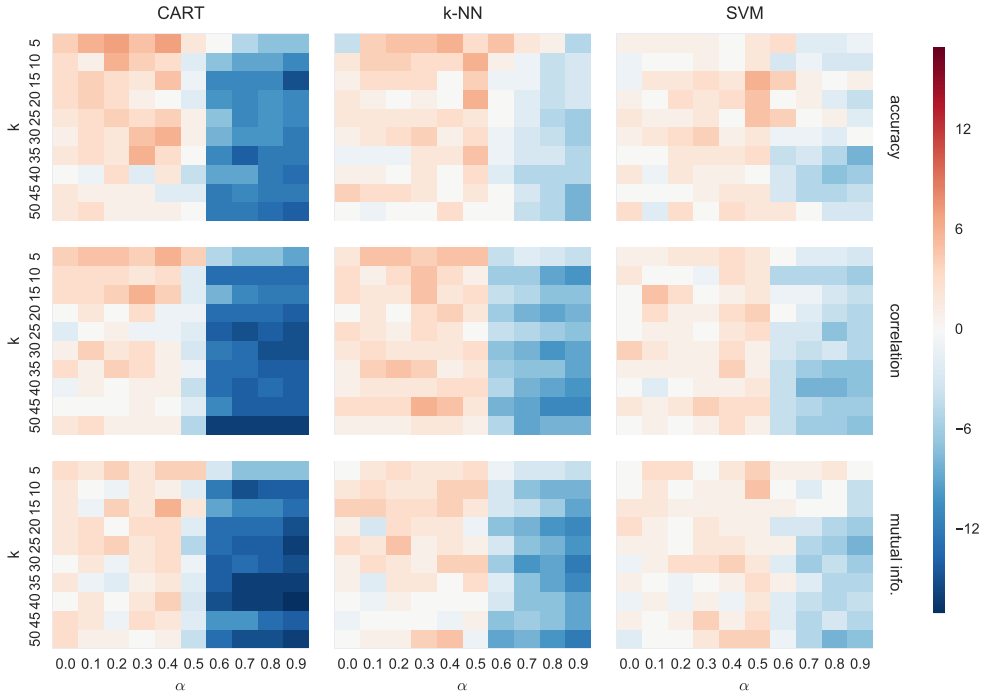


Fig. 2. Differences between the number of datasets on which proposed method achieved statistically significantly better and worst results and their relation to the number of subspaces k used.

TABLE IV
CART CLASSIFIER, MUTUAL INFO. AS FEATURE QUALITY MEASURE.

	α									
k	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
5	+4/-1	+3/-1	+4/-0	+4/-2	+5/-1	+5/-1	+2/-5	+3/-10	+2/-9	+3/-10
15	+2/-0	+1/-1	+1/-2	+2/-0	+2/-1	+1/-3	+0/-12	+0/-15	+0/-14	+0/-14
10	+3/-1	+1/-2	+4/-0	+2/-0	+6/-0	+3/-1	+1/-10	+1/-12	+1/-12	+1/-14
30	+2/-0	+3/-0	+0/-0	+4/-1	+3/-0	+1/-3	+1/-14	+1/-14	+1/-14	+1/-16
25	+2/-0	+3/-0	+6/-2	+2/-0	+3/-0	+1/-6	+2/-15	+1/-15	+1/-15	+1/-17
20	+1/-1	+3/-1	+1/-2	+2/-0	+4/-1	+2/-3	+1/-15	+1/-14	+1/-15	+1/-16
35	+3/-1	+0/-1	+1/-3	+1/-0	+2/-0	+3/-5	+1/-15	+1/-17	+1/-17	+1/-17
40	+1/-1	+2/-0	+1/-1	+2/-0	+4/-0	+3/-6	+1/-16	+1/-17	+1/-17	+1/-18
45	+3/-0	+3/-1	+1/-2	+3/-0	+2/-0	+2/-4	+2/-12	+1/-11	+1/-14	+1/-15
50	+3/-0	+1/-0	+1/-0	+0/-0	+2/-1	+1/-5	+1/-14	+1/-16	+1/-16	+1/-17

TABLE V
K-NN CLASSIFIER, ACCURACY AS FEATURE QUALITY MEASURE.

	α									
k	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
5	+0/-4	+4/-0	+5/-0	+5/-0	+6/-0	+3/-0	+6/-1	+4/-2	+4/-3	+3/-8
15	+2/-0	+4/-0	+4/-0	+3/-0	+3/-0	+7/-1	+5/-4	+3/-4	+3/-7	+4/-7
10	+2/-1	+3/-0	+3/-0	+3/-0	+1/-1	+5/-1	+4/-5	+5/-6	+4/-8	+5/-8
30	+2/-0	+1/-0	+0/-0	+1/-0	+1/-1	+7/-1	+3/-3	+2/-4	+2/-6	+2/-5
25	+3/-1	+2/-0	+2/-0	+2/-0	+3/-0	+4/-2	+3/-3	+2/-4	+1/-6	+1/-7
20	+2/-1	+2/-0	+3/-0	+5/-1	+1/-0	+4/-1	+4/-5	+1/-4	+1/-5	+1/-7
35	+0/-1	+1/-2	+1/-2	+4/-2	+4/-2	+7/-2	+5/-6	+4/-7	+4/-7	+2/-7
40	+1/-0	+1/-0	+1/-1	+2/-0	+1/-1	+6/-3	+4/-6	+2/-7	+2/-7	+3/-8
45	+4/-0	+3/-0	+3/-0	+2/-0	+3/-3	+5/-3	+4/-4	+3/-6	+3/-8	+2/-10
50	+1/-1	+0/-1	+0/-0	+0/-0	+3/-0	+3/-3	+4/-4	+3/-7	+4/-9	+3/-11

TABLE VI
K-NN CLASSIFIER, CORRELATION AS FEATURE QUALITY MEASURE.

	α									
k	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
5	+3/-1	+6/-1	+6/-1	+5/-1	+6/-1	+6/-1	+2/-6	+3/-5	+2/-5	+2/-6
15	+3/-0	+1/-0	+3/-0	+5/-0	+3/-0	+2/-1	+2/-8	+2/-8	+1/-10	+1/-11
10	+3/-0	+3/-1	+2/-0	+5/-0	+3/-1	+5/-2	+2/-6	+3/-8	+1/-8	+1/-8
30	+1/-1	+4/-2	+1/-1	+3/-0	+4/-0	+5/-3	+3/-9	+3/-11	+3/-12	+4/-12
25	+4/-1	+3/-1	+2/-1	+4/-1	+4/-2	+3/-3	+2/-6	+2/-7	+2/-8	+2/-9
20	+3/-1	+5/-2	+3/-1	+2/-0	+2/-0	+7/-4	+3/-10	+2/-10	+2/-12	+2/-11
35	+3/-2	+4/-0	+5/-0	+4/-0	+3/-2	+5/-3	+3/-9	+2/-9	+2/-9	+2/-11
40	+1/-0	+2/-0	+2/-0	+2/-0	+3/-0	+5/-2	+2/-7	+0/-8	+0/-9	+0/-10
45	+4/-1	+3/-0	+3/-0	+6/-0	+5/-0	+5/-3	+2/-10	+2/-11	+1/-12	+1/-12
50	+1/-0	+1/-0	+1/-0	+2/-0	+2/-0	+5/-4	+1/-6	+1/-10	+2/-10	+2/-10

TABLE VII
K-NN CLASSIFIER, MUTUAL INFO. AS FEATURE QUALITY MEASURE.

	α									
k	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
5	+0/-0	+2/-0	+3/-0	+2/-0	+2/-0	+3/-0	+3/-4	+3/-6	+3/-6	+3/-7
15	+2/-0	+3/-0	+2/-0	+2/-0	+4/-0	+5/-1	+3/-6	+3/-10	+2/-10	+2/-10
10	+4/-0	+4/-0	+4/-1	+4/-1	+4/-0	+5/-3	+4/-7	+3/-7	+3/-9	+2/-11
30	+3/-2	+0/-3	+5/-2	+3/-1	+3/-1	+4/-5	+2/-9	+2/-10	+2/-12	+1/-12
25	+4/-1	+2/-0	+5/-0	+2/-1	+2/-0	+4/-3	+3/-6	+1/-8	+1/-9	+1/-11
20	+2/-1	+3/-0	+2/-0	+1/-0	+4/-0	+6/-2	+2/-8	+1/-11	+1/-10	+0/-12
35	+1/-0	+0/-2	+2/-0	+2/-0	+1/-1	+5/-2	+2/-4	+1/-8	+1/-10	+0/-11
40	+1/-1	+1/-2	+2/-1	+1/-1	+1/-1	+3/-3	+3/-6	+1/-8	+1/-8	+1/-10
45	+0/-1	+0/-0	+0/-0	+0/-0	+0/-1	+4/-3	+1/-8	+2/-8	+2/-9	+0/-9
50	+3/-3	+2/-1	+2/-1	+5/-1	+6/-1	+5/-6	+1/-10	+2/-10	+1/-11	+1/-13

TABLE VIII
SVM CLASSIFIER, ACCURACY AS FEATURE QUALITY MEASURE.

	α									
k	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
5	+1/-0	+1/-0	+1/-0	+1/-0	+0/-0	+2/-0	+3/-0	+2/-4	+1/-3	+1/-2
15	+0/-1	+0/-0	+1/-1	+1/-0	+1/-1	+2/-0	+2/-5	+2/-3	+2/-5	+2/-5
10	+1/-2	+2/-0	+3/-1	+3/-0	+2/-0	+6/-0	+5/-1	+5/-4	+4/-4	+5/-4
30	+2/-1	+2/-2	+3/-0	+3/-1	+5/-2	+6/-1	+4/-4	+5/-4	+4/-6	+4/-8
25	+3/-1	+1/-0	+1/-0	+3/-1	+0/-0	+5/-0	+5/-1	+4/-4	+4/-3	+4/-6
20	+2/-1	+2/-0	+3/-0	+4/-0	+3/-2	+4/-2	+2/-3	+3/-4	+3/-5	+4/-4
35	+1/-1	+1/-1	+2/-0	+2/-0	+2/-0	+4/-1	+2/-6	+2/-5	+2/-7	+1/-9
40	+2/-0	+0/-1	+2/-0	+1/-0	+2/-1	+2/-2	+1/-4	+1/-6	+1/-8	+1/-7
45	+1/-1	+1/-1	+2/-2	+1/-0	+2/-0	+3/-1	+3/-6	+2/-7	+2/-7	+2/-6
50	+3/-0	+1/-3	+3/-0	+2/-2	+4/-1	+5/-3	+5/-3	+5/-5	+5/-8	+5/-8

TABLE IX
SVM CLASSIFIER, CORRELATION AS FEATURE QUALITY MEASURE.

	α									
k	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
5	+1/-0	+2/-1	+3/-1	+2/-0	+3/-0	+2/-0	+5/-5	+4/-6	+3/-5	+3/-6
15	+2/-0	+0/-0	+1/-1	+0/-1	+3/-0	+4/-2	+2/-7	+2/-7	+2/-7	+2/-8
10	+2/-2	+5/-0	+3/-0	+1/-1	+3/-2	+2/-0	+4/-5	+4/-5	+4/-7	+4/-8
30	+1/-1	+3/-0	+2/-1	+2/-1	+3/-0	+5/-1	+5/-5	+4/-6	+4/-8	+4/-9
25	+1/-1	+1/-0	+1/-0	+0/-0	+2/-0	+3/-0	+3/-6	+3/-6	+2/-9	+2/-7
20	+4/-0	+2/-0	+3/-2	+1/-0	+3/-0	+3/-2	+3/-6	+4/-8	+3/-6	+3/-8
35	+1/-0	+2/-1	+1/-0	+3/-2	+4/-0	+2/-1	+3/-7	+3/-9	+2/-9	+2/-9
40	+1/-1	+0/-2	+2/-2	+2/-1	+2/-1	+3/-1	+2/-6	+2/-10	+2/-10	+2/-9
45	+3/-1	+2/-1	+3/-1	+4/-0	+3/-0	+5/-2	+3/-7	+3/-8	+3/-9	+3/-9
50	+0/-0	+1/-0	+1/-1	+2/-1	+3/-0	+2/-1	+3/-7	+2/-8	+2/-8	+3/-10

TABLE X
SVM CLASSIFIER, MUTUAL INFO. AS FEATURE QUALITY MEASURE.

	α									
k	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
5	+1/-1	+4/-1	+4/-1	+2/-2	+4/-2	+5/-1	+5/-4	+4/-3	+5/-3	+5/-5
15	+1/-1	+2/-0	+0/-0	+1/-0	+1/-0	+6/-1	+3/-3	+3/-5	+3/-3	+3/-7
10	+2/-0	+2/-0	+2/-1	+2/-1	+1/-0	+2/-1	+4/-3	+4/-4	+4/-4	+3/-7
30	+3/-0	+1/-0	+0/-0	+2/-0	+1/-0	+3/-2	+3/-6	+3/-6	+3/-8	+3/-9
25	+2/-1	+1/-0	+1/-1	+2/-0	+3/-1	+4/-3	+4/-3	+2/-7	+3/-9	+2/-10
20	+0/-1	+1/-0	+4/-1	+3/-0	+5/-1	+4/-2	+5/-7	+4/-10	+4/-8	+4/-10
35	+1/-0	+1/-1	+0/-1	+0/-0	+1/-0	+6/-3	+4/-3	+3/-5	+2/-7	+3/-8
40	+1/-2	+1/-0	+0/-1	+2/-1	+0/-1	+4/-2	+3/-5	+3/-7	+3/-8	+3/-7
45	+1/-0	+0/-0	+2/-1	+4/-0	+1/-0	+4/-1	+3/-5	+3/-8	+3/-6	+3/-8
50	+0/-2	+0/-0	+2/-2	+2/-0	+4/-0	+5/-1	+4/-5	+3/-8	+3/-11	+3/-10

REFERENCES

[1] M. Woźniak, M. Graña, and E. Corchado, "A survey of multiple classifier systems as hybrid systems," *Information Fusion*, vol. 16, pp. 3–17, 2014.

[2] S. Wang and X. Yao, "Relationships between diversity of classification ensembles and single-class performance measures," *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 1, pp. 206–219, 2013.

[3] W. M. Czarnecki, R. Józefowicz, and J. Tabor, "Maximum entropy linear manifold for learning discriminative low-dimensional representation," in *Machine Learning and Knowledge Discovery in Databases - European Conference, ECML PKDD 2015, Porto, Portugal, September 7-11, 2015, Proceedings, Part I*, 2015, pp. 52–67.

[4] T. K. Ho, "The random subspace method for constructing decision forests," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 8, pp. 832–844, 1998.

[5] T. Windeatt, "Accuracy/diversity and ensemble MLP classifier design," *IEEE Trans. Neural Networks*, vol. 17, no. 5, pp. 1194–1211, 2006.

[6] B. Krawczyk and M. Woźniak, "Untrained weighted classifier combination with embedded ensemble pruning," *Neurocomputing*, vol. 196, pp. 14 – 22, 2016.

[7] P. Trajdos and M. Kurzynski, "A dynamic model of classifier competence based on the local fuzzy confusion matrix and the random reference classifier," *Applied Mathematics and Computer Science*, vol. 26, no. 1, p. 175, 2016.

[8] L. Rokach, "Decision forest: Twenty years of research," *Information Fusion*, vol. 27, pp. 111–125, 2016.

[9] P. M. Álvarez, J. Luengo, and F. Herrera, "A first study on the use of boosting for class noise reparation," in *Hybrid Artificial Intelligent Systems - 11th International Conference, HAIS 2016, Seville, Spain, April 18-20, 2016, Proceedings*, 2016, pp. 549–559.

[10] B. Cyganek, "One-class support vector ensembles for image segmentation and classification," *Journal of Mathematical Imaging and Vision*, vol. 42, no. 2-3, pp. 103–117, 2012.

[11] J. Maudes, J. J. R. Diez, C. I. García-Osorio, and N. García-Pedrajas, "Random feature weights for decision tree ensemble construction," *Information Fusion*, vol. 13, no. 1, pp. 20–30, 2012.

[12] A. M. P. Canuto, K. M. O. Vale, A. F. Neto, and A. Signoretti, "Reinsel: A class-based mechanism for feature selection in ensemble of classifiers," *Appl. Soft Comput.*, vol. 12, no. 8, pp. 2517–2529, 2012.

[13] K. Nag and N. R. Pal, "A multiobjective genetic programming-based ensemble for simultaneous feature selection and classification," *IEEE Trans. Cybernetics*, vol. 46, no. 2, pp. 499–510, 2016.

[14] S. Özögür-Akyüz, T. Windeatt, and R. S. Smith, "Pruning of error correcting output codes by optimization of accuracy-diversity trade off," *Machine Learning*, vol. 101, no. 1-3, pp. 253–269, 2015.

[15] M. Galar, A. Fernández, E. Barrenechea, and F. Herrera, "DRCW-OVO: distance-based relative competence weighting combination for one-vs-one strategy in multi-class problems," *Pattern Recognition*, vol. 48, no. 1, pp. 28–42, 2015.

[16] I. T. Podolak and A. Roman, "Theoretical foundations and experimental results for a hierarchical classifier with overlapping clusters," *Computational Intelligence*, vol. 29, no. 2, pp. 357–388, 2013.

[17] T. Sun, L. Jiao, F. Liu, S. Wang, and J. Feng, "Selective multiple kernel learning for classification with ensemble strategy," *Pattern Recognition*, vol. 46, no. 11, pp. 3081–3090, 2013.

[18] R. E. Banfield, L. O. Hall, K. W. Bowyer, and W. P. Kegelmeyer, "Ensemble diversity measures and their application to thinning," *Information Fusion*, vol. 6, no. 1, pp. 49–62, 2005.

[19] E. Alpaydin, "Combined 5 x 2 cv F test for comparing supervised classification learning algorithms," *Neural Computation*, vol. 11, no. 8, pp. 1885–1892, 1999.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 97–103
DOI: 10.15439/2016F280
ACSIS, Vol. 8. ISSN 2300-5963

Modification of the Probabilistic Neural Network with the Use of Sensitivity Analysis Procedure

Maciej Kusy*
*Faculty of Electrical and Computer Engineering,
Rzeszow University of Technology,
al. Powstancow Warszawy 12, 35-959 Rzeszow, Poland,
Email: mkusy@prz.edu.pl

Piotr A. Kowalski†‡
†Faculty of Physics and Applied Computer Science,
AGH University of Science and Technology,
al. A. Mickiewicza 30, 30-059 Cracow, Poland,
Email: pkowal@agh.edu.pl
‡Systems Research Institute,
Polish Academy of Sciences,
ul. Newelska 6, 01-447 Warsaw, Poland,
E-mail: pakowal@ibspan.waw.pl

***Abstract*—In this article, the modified probabilistic neural network (MPNN) is proposed. The network is an extension of conventional PNN with the weight coefficients introduced between pattern and summation layer of the model. These weights are calculated by using the sensitivity analysis (SA) procedure. MPNN is employed to the classification tasks and its performance is assessed on the basis of prediction accuracy. The effectiveness of MPNN is also verified by analyzing its results with these obtained for both the original PNN and commonly known classification algorithms: support vector machine, multilayer perceptron, radial basis function network and k-Means clustering procedure. It is shown that the proposed modification improves the prediction ability of the PNN classifier.**

I. Introduction

PROBABILISTIC neural network (PNN) is a data classifier proposed by Specht in [1] and [2]. It attracts researchers from the field of machine learning. In literature, one can find PNN's applications mainly in medical diagnosis and prediction [3], [4], [5], [6] and image classification and recognition [7], [8], [9]. However, a very good classification performance allows PNN to be utilized in other tasks, e.g.: earthquake magnitude prediction [10], multiple partial discharge sources classification [11], interval information processing [12], [13], phoneme recognition [14], email security enhancement [15] or intrusion detection systems [16].

In its conventional form, PNN is a multilayered feedforward network composed of four layers: input layer (represented by data features), pattern layer (consisting of as many neurons as training patterns), summation layer (with one neuron for each class) and output layer where a single neuron produces a classification result. In majority of cases, the connections between layers in PNN are not equipped with weight coefficients (or all the weights are equally set to 1). Therefore, the output response is not influenced by an impact of a particular class.

In literature, a subtle attention has been paid to determining the weights of PNN. The work of [17] is the first contribution where the weights are introduced to PNN. However, the coefficients are not computed directly; the network operates by using anisotropic Gaussians, i.e. the covariance matrix is utilized, instead of a single smoothing parameter, to compute the output for a particular class. In [18] and [19], the weighted PNN is proposed in the way it adds weighting factors between pattern and summation layer. These factors are calculated from soft labeling probability matrix to carry out a classification. The authors of [20] and [21] create weighted PNN based on their class separability. A single weight is defined as the ratio of 'between-class variance' and 'within-class variance' for a particular training pattern. As in [18], the weighting coefficients are used to connect the pattern and summation layer.

In this study, we propose the use the SA procedure in the computation of the weights for the PNN model. Similar to [18], [19], [20] and [21], the coefficients are inserted between pattern and summation layer. Their values are equal to the aggregated sensitivities normalized to $[0, 1]$ interval. The formulas for the weights are analytically derived. The idea is applied to PNN activated with a product Cauchy kernel. The proposed MPNN is tested in the classification problems of University of California, Irvine machine learning repository (UCI-MLR) data sets [22] by computing a 10-fold cross validation accuracy. The obtained outcomes are thoroughly compared to the ones obtained for original PNN. Furthermore, we verify MPNN accuracy with the accuracy of four reference classifiers: support vector machine, multilayer perceptron, radial basis function neural network and k-Means clustering algorithm.

This work is structured as follows. In section II, the SA procedure is highlighted. Section III, presents the fundamentals of the PNN model. In section IV, the procedure of computing the weights for PNN is proposed. Section V describes the input data sets and the reference classifiers used in current study. Here, the discussion regarding the results obtained in the simulations is also presented. Finally, section VI concludes the work.

This work was supported in part by the Rzeszow University of Technology under Grant No. U-596/DS

II. Sensitivity analysis

SA, in general, is one of many approaches used in determining the importance of particular inputs of a neural network. Therefore, it can be applied in the task of elimination of the irrelevant features in the input vectors. The main idea of SA is based on computing the influence of input features on a neural network output signal after a training process. This influence is characterized by real coefficients [23]

$$S_{j,i}^{(p)} = \frac{\partial}{\partial x_i} y_j \left(x_1^{(p)}, x_2^{(p)}, \ldots, x_N^{(p)} \right), \tag{1}$$

where x_i denotes an input feature and y_j stands for an output signal. In (1), $i = 1, \ldots, N$, $j = 1, \ldots, J$, where N and J indicate the number of features and outputs, respectively.

More specifically, equation (1) represents the sensitivity of the jth neural network output on the ith feature of the input vector $\mathbf{x}$ determined based on the pth training pattern $\mathbf{x}^{(p)}$ for $p = 1, \ldots, P$ where P is a data set cardinality. Taking into account N dimensional data set, (1) can be presented as the following matrix

$$\mathbf{S}^{(p)} = \begin{bmatrix} S_{1,1}^{(p)} & S_{1,2}^{(p)} & \ldots & S_{1,N}^{(p)} \\ S_{2,1}^{(p)} & S_{2,2}^{(p)} & \ldots & S_{2,N}^{(p)} \\ \vdots & \vdots & \ddots & \vdots \\ S_{J,1}^{(p)} & S_{J,2}^{(p)} & \ldots & S_{J,N}^{(p)} \end{bmatrix}. \tag{2}$$

Once $\mathbf{S}^{(p)}$ is computed for all P training patterns, it is possible to find aggregated parameters after application of various types of norms. In the research, one usually utilizes the parameter of the mean square average sensitivity

$$S_{j,i}^{\text{mean}} = \sqrt{\frac{\sum_{p=1}^{P} \left(S_{j,i}^{(p)} \right)^2}{P}}. \tag{3}$$

The absolute value average sensitivity and the maximum sensitivity parameters are also frequently applied in input significance estimation [24]. The appropriate impact of $S_{j,i}^{(p)}$ on the aggregated outcome of $S_{j,i}$ implies the selection of a particular norm.

III. Probabilistic neural network

PNN is composed of four layers. The coordinates of an input vector $\mathbf{x} = [x_1, \ldots, x_N]$ constitute the first input layer. The second layer, called a pattern layer, consists of as many neurons as training examples. Pattern neurons feed their output to the next summation layer. In the summation layer, there are J neurons, however each jth neuron sums the inputs from the neurons of jth class. In literature, two approaches are usually utilized to compute the signals of the summation neurons: the additive Gaussian kernels and the product kernels. In this work, we use the second approach, therefore the summation neuron output is defined as follows

$$f_j(\mathbf{x}) = \frac{1}{P_j \det(\mathbf{h})} \sum_{p=1}^{P_j} \frac{1}{s_p^N} K \left(\frac{\left(\mathbf{x} - \mathbf{x}_j^{(p)} \right)^T \mathbf{h}^{-1}}{s_p} \right), \tag{4}$$

where:

- $K(\cdot)$ is the kernel function calculated in the following way

$$K(\mathbf{x}) = \mathcal{K}(x_1) \cdot \mathcal{K}(x_2) \cdot \ldots \cdot \mathcal{K}(x_N), \tag{5}$$

for which

$$\mathcal{K}(x_i) = \frac{2}{\pi(x_i^2 + 1)^2} \tag{6}$$

denotes the one-dimensional Cauchy multiplicand;
- P_j stands for the number of cases in the jth class ($j = 1, \ldots, J$);
- $\mathbf{x}_j^{(p)} = [x_{j,1}^{(p)}, \ldots, x_{j,N}^{(p)}]$ is the pth training vector of the jth class.

If one regards (5) and (6) as the pattern neuron activation function, the summation layer output for the jth class is determined as follows

$$f_j(\mathbf{x}) = \frac{1}{P_j \det(\mathbf{h})} \sum_{p=1}^{P_j} \frac{1}{s_p^N} \prod_{i=1}^{N} \frac{2}{\pi \left(\left(\frac{x_i - x_{j,i}^{(p)}}{h_i s_p} \right)^2 + 1 \right)^2}. \tag{7}$$

Finally, using the Bayes theorem [2], the output layer of PNN determines the label for a new test vector $\mathbf{x}$

$$C(\mathbf{x}) = \underset{j=1 \ldots J}{\operatorname{argmax}} \, f_j(\mathbf{x}), \tag{8}$$

where $C(\mathbf{x})$ denotes the predicted class. The training algorithm for this network amounts to the appropriate choice of the smoothing parameter h_i by means of the plug-in method [25], and the computation of the modification coefficient s_p [26]. The structure of the PNN model is illustrated in Fig. 1.

IV. Proposed algorithm

In Fig. 2, we present the step-by-step data classification algorithm with the use of modified PNN model. We start with the calculation of the sensitivity coefficients for all $r = 1, 2, \ldots, P_j$ neurons in the pattern layer

$$S = \frac{\partial \hat{f}_j \left(\mathbf{x}_j^{(p)} \right)}{\partial \mathbf{x}_j^{(r)}}, \tag{9}$$

were $\mathbf{x}_j^{(p)} \in \left\{ \mathbf{x}_j^{(1)}, \mathbf{x}_j^{(2)}, \ldots, \mathbf{x}_j^{(P_j)} \right\}$ is the vector argument from jth class. Thus, $\mathbf{S}$ for jth class takes the matrix form

$$\mathbf{S}_j = \left\{ \frac{\partial \hat{f}_j \left(\mathbf{x}_j^{(p)} \right)}{\partial \mathbf{x}_j^{(r)}} \right\}_{P_j \times P_j}. \tag{10}$$

In (10), the elements of rth column represent the sensitivities of KDE in jth class in regard to each rth pattern neuron computed for a specific input pattern p. Since the denominator of each item of $\mathbf{S}_j$ is a vector, the following gradient has to be determined

$$\nabla \hat{f}_j^{(p,r)} = \frac{\partial \hat{f}_j \left(\mathbf{x}_j^{(p)} \right)}{\partial \mathbf{x}_j^{(r)}} = \left[\frac{\partial \hat{f}_j \left(\mathbf{x}_j^{(p)} \right)}{\partial x_{j,1}^{(r)}}, \ldots, \frac{\partial \hat{f}_j \left(\mathbf{x}_j^{(p)} \right)}{\partial x_{j,N}^{(r)}} \right], \tag{11}$$

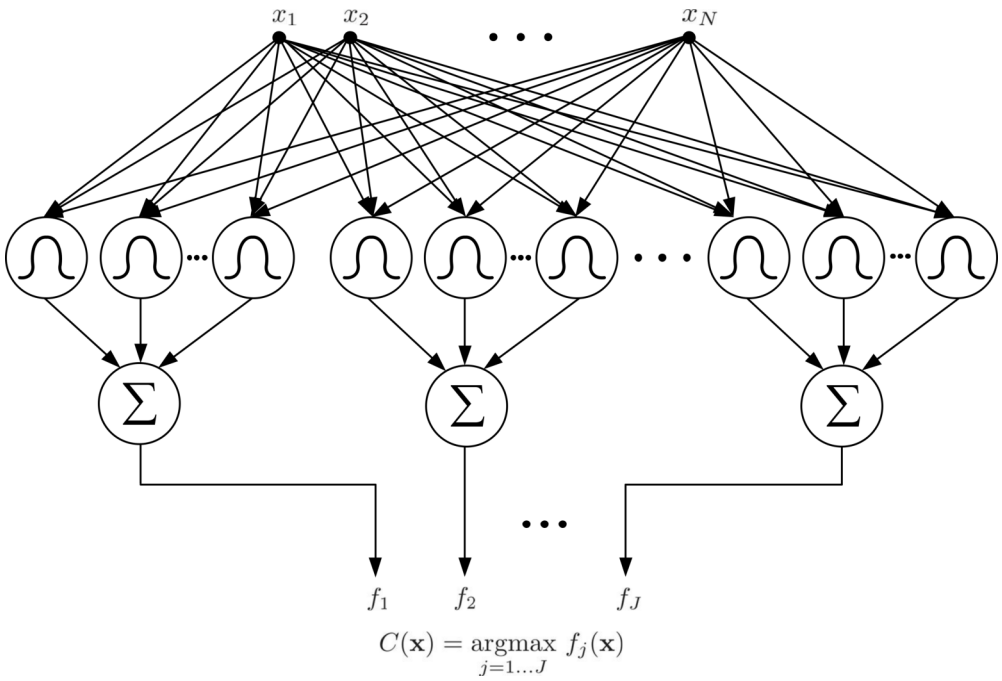


Fig. 1. The architecture of probabilistic neural network.

where

$$\frac{\partial \hat{f}_j\left(\mathbf{x}_j^{(p)}\right)}{\partial x_{j,i}^{(r)}} = \frac{1}{P_j \det(\mathbf{h})} \frac{1}{s_r^N} \frac{\partial}{\partial x_{j,i}^{(r)}} K\left(\frac{\left(\mathbf{x}_j^{(p)} - \mathbf{x}_j^{(r)}\right)^T \mathbf{h}^{-1}}{s_r}\right). \tag{12}$$

The product form of KDE in (5) expands (12) into the following formula

$$\frac{\partial \hat{f}_j\left(\mathbf{x}_j^{(p)}\right)}{\partial x_{j,i}^{(r)}} = \frac{1}{P_j \det(\mathbf{h})} \frac{1}{s_r^N} \mathcal{K}\left(\frac{x_{j,1}^{(p)} - x_{j,1}^{(r)}}{h_1 s_r}\right) \cdots \frac{\partial}{\partial x_{j,i}^{(r)}} \mathcal{K}^\star\left(\frac{x_{j,i}^{(p)} - x_{j,i}^{(r)}}{h_i s_r}\right) \ldots \mathcal{K}\left(\frac{x_{j,N}^{(p)} - x_{j,N}^{(r)}}{h_N s_r}\right). \tag{13}$$

The Cauchy kernel in (6) allows us to determine the ith coefficient of (13)

$$\frac{\partial}{\partial x_{j,i}^{(r)}} \mathcal{K}^\star\left(x_{j,i}^{(p)}\right) = \frac{8\left(x_{j,i}^{(p)} - x_{j,i}^{(r)}\right)}{\pi h_i^2 s_r^2 \left(\left(\frac{x_{j,i}^{(p)} - x_{j,i}^{(r)}}{h_i s_r}\right)^2 + 1\right)^3}. \tag{14}$$

We can see that $\nabla \hat{f}_j^{(p,r)}$ is a vector field, therefore, in order to extract an information on sensitivity of KDE of jth class for a given rth pattern neuron, it is necessary to determine the norm of (11). Thus, $\mathbf{S}_j$ in (10) must be generalized to the following matrix

$$\mathbf{S}_j = \left\{ \|\nabla \hat{f}_j^{(p,r)}\| \right\}_{P_j \times P_j}, \tag{15}$$

where

$$\|\nabla \hat{f}_j^{(p,r)}\| = \sqrt{\sum_{i=1}^{N} \left(\frac{\partial \hat{f}_j\left(\mathbf{x}_j^{(p)}\right)}{\partial x_{j,i}^{(r)}}\right)^2} \tag{16}$$

for the Euclidean norm. The matrix $\mathbf{S}_j$ is computed for $j = 1, \ldots, J$.

Now it is necessary to aggregate all $p = 1, \ldots, P_j$ entries in each rth column of $\mathbf{S}_j$ to obtain the aggregated sensitivity vector. For the mean square average sensitivity measure, this vector takes the following form

$$\mathbf{a}_j = \left[\sqrt{\frac{\sum_{p=1}^{P_j} \left(S_j^{(p,1)}\right)^2}{P_j}}, \ldots, \sqrt{\frac{\sum_{p=1}^{P_j} \left(S_j^{(p,P_j)}\right)^2}{P_j}} \right]. \tag{17}$$

Finally, the normalization of the elements of $\mathbf{a}_j$ to some interval introduces the weight vector for each jth class. In this work, we propose simple "max" normalization

$$\mathbf{w}_j = \left[\frac{a_j^{(1)}}{\max(\mathbf{a}_j)}, \frac{a_j^{(2)}}{\max(\mathbf{a}_j)}, \ldots, \frac{a_j^{(P_j)}}{\max(\mathbf{a}_j)} \right], \tag{18}$$

where $\max(\cdot)$ operator returns the maximum value of the argument.

Including the weights' coefficients, the summation layer output (7) is redefined as

$$f_j(\mathbf{x}) = \frac{1}{P_j \det(\mathbf{h})} \sum_{\{p,r\}=1}^{P_j} w_j^{(r)} \frac{1}{s_p^N} K\left(\frac{\left(\mathbf{x} - \mathbf{x}_j^{(p)}\right)^T \mathbf{h}^{-1}}{s_p}\right), \tag{19}$$

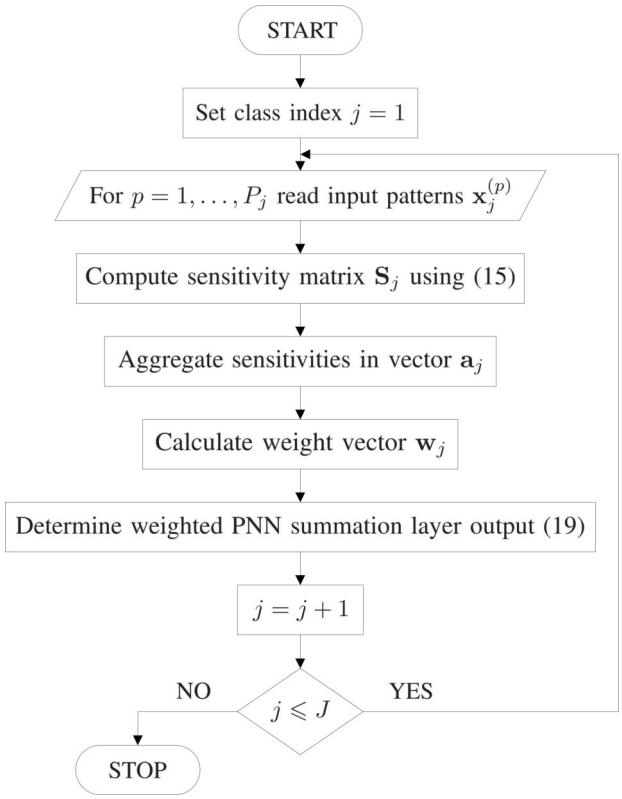


Fig. 2. The algorithm for the computation of the weights in PNN between pattern and summation layer.

where

$$w_j^{(r)} = \frac{\sqrt{\frac{1}{P_j}\sum_{p=1}^{P_j}\left(S_j^{(p,r)}\right)^2}}{\max(\mathbf{a}_j)}, \tag{20}$$

where r refers to rth element of $\mathbf{a}_j$.

V. Experiments

In this section, we present the classification results obtained by PNN with introduced weights and the original network. These results are compared with the outcomes achieved by other classifiers: support vector machine (SVM) algorithm, multilayer perceptron (MLP), radial basis function neural network (RBFN) and k-Means method. Both, input data sets and the reference classifiers used in the simulations are also described.

A. Input data sets

The performance of the proposed MPNN, original PNN and the reference methods (SVM, MLP, RBFN and k-Means) is evaluated on UCI-MLR data sets. Ten well known and commonly tested databases are included: Wisconsin breast cancer (WBC), Statlog heart (SH), Pima Indians diabetes (PID), Ecoli (E), Parkinsons (P), Iris (I), breast tissue (BT), monk (M), seeds (S) and cardiotocography (CTG). The cardinality, dimensionality, number of classes and the class distributions for the utilized databases are presented in Table I.

B. Reference methods

In this subsection, the reference classification models are highlighted. Additionally, we provide the parameter settings used for these classifiers in the simulations.

1) SVM: SVM is the classification algorithm proposed by Vapnik [27]. To perform the classification, SVM requires the solution of the quadratic programming optimization problem. For this purpose, various kernel functions need to be explored. In the current study, we verify the following ones:

- radial basis kernel

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}\right), \tag{21}$$

- polynomial kernel

$$K(\mathbf{x}_i, \mathbf{x}_j) = \left(\alpha\left(\mathbf{x}_i \cdot \mathbf{x}_j\right) + \beta\right)^k. \tag{22}$$

In the classification tasks, the parameters of the kernels (21) and (22) and capacity control parameter C must be appropriately selected.

2) MLP: MLP is a feedforward neural network [28]. This network is composed of an input layer, hidden layers, and an output layer. The number of hidden layers, the optimal number of neurons in hidden layers and the appropriate activation functions must be determined for this model. In this work, the following functions are tested:

- linear identity

$$f(x) = x, \tag{23}$$

TABLE I
UCI-MLR DATA SETS USED TO TEST ALL COMPARED MODELS

Data set	Records	Attributes	Classes	Class distribution
WBC	683	9	2	444–239
SH	270	13	2	150–120
PID	786	8	2	500–268
E	327	5	5	143–77–35–20–52
P	195	22	2	147–48
I	150	4	3	50–50–50
BT	106	9	6	21–15–18–16–14–22
M	432	6	2	216–216
S	210	7	3	70–70–70
CTG	2126	22	3	1655–295–176

- logistic sigmoid

$$f(x) = \frac{1}{1 + e^{-\alpha x}}, \tag{24}$$

- hyperbolic tangent

$$f(x) = \frac{2}{1 + e^{-\beta x}} - 1, \tag{25}$$

where α and β are the coefficients used to control the slope of (24) and (25).

3) RBFNN: RBFNN, similar to PNN and MLP, is a feed-forward neural network [29]. However, this model consists of three layers: an input layer, a radial basis hidden layer and a linear output layer. The number of the neurons in the hidden layer and the parameters of the RBFNN training method must be appropriately selected.

4) k-Means: The k-Means clustering is an unsupervised learning algorithm. It partitions input data into k clusters and provides a center of each cluster [30]. As a result, the records within each cluster are similar to each other and distinct from records in other clusters. The predictions for the unknown cases are made by assigning them the category of the nearest cluster center. The parameter k is increased up to K, which depends on the number of input vectors of a given class (P_j). In the current study, K does not exceed 50% of P_j, $j = 1, \ldots, J$. The step for k is determined empirically.

The training methods and the parameters of the models are presented in Table II.

TABLE II
SIMULATION PARAMETERS OF THE EXAMINED REFERENCE CLASSIFIERS

SVM	
	kernel functions: – radial basis kernel (21) – polynomial kernel (22) The grid search is performed for σ, α, β and k Capacity control coefficient: $C = \{10^{-1}, 10^{0}, 10^{1}, 10^{2}, 10^{3}, 10^{4}, 10^{5}\}$
MLP	
	training method: scaled conjugate gradients number of hidden layers: $\{1, 2\}$ number of hidden neurons: $\{2, 3, \ldots, 30\}$ activation functions: – linear (23) – logistic (24) – tangent (25)
RBFN	
	training method: weighted boosting search [31] neuron tuning parameters: size of population: $\{400, 600, 800, 1000\}$ maximum generation: $\{100, 200, 500\}$ boosting iterations: 50 number of hidden neurons: $\{2, 3, \ldots, 30\}$
k-Means	
	number of clusters: $\{2, \ldots, K\}$ distance measure: Euclidean distance

C. Results and discussion

In Table III, we present the accuracy (*Acc*) determined for MPNN, PNN and the reference methods in the classification problems of WBC, SH, PID, E, P, I, BT, M, S and CTG data sets. The accuracy is computed with the use of a 10-fold cross validation procedure. Table IV shows the optimal parameters of the reference classifiers for which the highest accuracy is achieved in particular classification problems. In the case of the SVM model, the capacity control parameter C and the spread constant σ are shown for the radial basis kernel (21) since for this function, a higher accuracy values are obtained in contrast to the results achieved with the use of kernel presented in (22). For MLP, the network structure is presented in the form *w–x–y–z*, where *w* and *z* denote the number of input and output neurons, respectively while *x* and *y* stand for the number of neurons in the hidden layers. The abbreviations "lin" and "log" refer to linear (23) and logistic (24) activation functions, respectively. For example, in the case of E data classification case, the string "5–13–4–5 log–log–log" describes MLP with 5 input neurons, 13 neurons in the first hidden layer, 4 neurons in the second hidden layer and 5 output neurons where the logistic activation function is applied in hidden and output layers. In the case of the RBFN and k-Means classifiers, n and k denote the number of neurons in the network's hidden layer and the number of cluster centers, respectively. The optimal parameters are obtained after vast number of simulations performed in DTREG software [32].

Comparing two first columns of Table III, one can definitely emphasize the fact that in almost all classification problems, the introduction of additional weight parameters to PNN

TABLE III
CROSS VALIDATION ACCURACY FOR: WEIGHTED PNN, ORIGINAL PNN AND THE REFERENCE CLASSIFIERS: SVM, MLP, RBFN AND K-MEANS IN UCI-MLR DATA CLASSIFICATION TASKS

Data set	MPNN	PNN	SVM	MLP	RBFN	k-Means
WBC	**0.9779**	0.9706	0.9546	0.9722	0.9663	0.9517
SH	**0.8148**	0.7963	0.8074	0.7926	0.7926	0.6852
PID	0.6948	0.6842	0.7474	**0.7669**	0.7435	0.6914
E	**0.8485**	0.8333	0.7982	0.8379	0.8471	0.8410
P	**0.9250**	0.8750	0.9231	0.9128	0.9026	0.8308
I	**1.000**	0.9667	0.9733	0.9733	0.9533	0.9667
BT	**0.7273**	0.7098	0.6792	0.6038	0.7075	0.6038
M	0.8605	0.8408	**0.9884**	0.9236	0.7500	0.9120
S	**0.9524**	**0.9524**	0.9429	0.9286	0.9476	0.9143
CTG	0.8568	0.8545	**0.9751**	0.9417	0.9694	0.8664

TABLE IV
THE PARAMETER VALUES FOR WHICH THE HIGHEST *Acc* IS OBTAINED FOR THE REFERENCE CLASSIFIERS

	SVM		MLP		RBFN	k-Menas
Data set	C	σ	structure	act. funct.	n	k
WBC	10^3	0.5	9–3–2	log–lin	50	57
SH	10^2	2.5	13–19–2	log–log	38	4
PID	10^3	0.3	8–7–2	log–lin	74	4
E	10^4	1.5	5–13–4–5	log–log–log	66	2
P	10^5	0.1	22–7–2	log–log	54	33
I	10^0	1.5	4–5–3	log–log	16	4
BT	10^1	27.1	9–3–6	log–log	34	11
M	10^2	0.7	6–14–8–2	log–log–lin	3	23
S	10^3	0.2	7–8–10–3	log–log–log	41	25
CTG	10^2	2.6	22–13–6–3	log–log–log	100	171

results in significant increase of its accuracy. Note, that in P data classification case, this increase is over 5% which is the highest investigated result. The above observation does not take place in S data set classification task. Here, the proposed modification of neural structure does not change obtained accuracy. Thus, the use of weights for original PNN is not beneficial in terms of prediction ability.

In general, one can note that only in three data set cases, i.e.: PID, M and CTG, the reference methods provide higher *Acc* value. However, among these cases, there is no unequivocal alternative classifier since SVM and MLP yield higher accuracy twice and once, respectively.

If we do not take the MPNN accuracy results into account, it is clear that the original PNN is an average quality classifier because in WBC, SH, E, P and I classification problems, higher *Acc* is obtained by the following methods: MLP, SVM, RBFNN, SVM, and SVM *ex aequo* MLP, respectively.

VI. CONCLUSION

In this paper, the modified probabilistic neural network was proposed. The modification relied on the introduction of the weights' coefficients between pattern and summation layers. The weights for PNN were computed by means of the SA procedure. Their values were equal to the aggregated sensitivities normalized to $[0, 1]$ interval. A PNN with product kernel estimator with Cauchy realization was used. The plug-in method was applied to determine the smoothing parameter. A 10-fold cross validation accuracies of MPNN were compared with the ones obtained for the original PNN and the reference classifiers in the UCI-MLR data sets classification tasks. The results showed, that among all tested models, MPNN achieved the highest *Acc* in seven out of ten classification cases. In contrast to the conventional PNN, the proposed network performed better in nine classification problems, while only in single task (S data set), the accuracies of MPNN and PNN were identical.

REFERENCES

[1] D. F. Specht, "Probabilistic neural networks," *Neural Networks*, vol. 3, no. 1, pp. 109–118, 1990.

[2] ——, "Probabilistic neural networks and the polynomial adaline as complementary techniques for classification," *Neural Networks, IEEE Transactions on*, vol. 1, no. 1, pp. 111–121, Mar 1990. doi: 10.1109/72.80210

[3] R. Folland, E. Hines, R. Dutta, P. Boilot, and D. Morgan, "Comparison of neural network predictors in the classification of tracheal–bronchial breath sounds by respiratory auscultation," *Artificial intelligence in medicine*, vol. 31, no. 3, pp. 211–220, 2004.

[4] D. Mantzaris, G. Anastassopoulos, and A. Adamopoulos, "Genetic algorithm pruning of probabilistic neural networks in medical disease estimation," *Neural Networks*, vol. 24, no. 8, pp. 831–835, 2011.

[5] M. Kusy and R. Zajdel, "Application of reinforcement learning algorithms for the adaptive computation of the smoothing parameter for probabilistic neural network," *Neural Networks and Learning Systems, IEEE Transactions on*, vol. 26, no. 9, pp. 2163–2175, 2015.

[6] ——, "Probabilistic neural network training procedure based on q(0)–learning algorithm in medical data classification," *Applied Intelligence*, vol. 41, no. 3, pp. 837–854, 2014.

[7] Y. Chtioui, S. Panigrahi, and R. Marsh, "Conjugate gradient and approximate newton methods for an optimal probabilistic neural network for food color classification," *Optical Engineering*, vol. 37, no. 11, pp. 3015–3023, 1998.

[8] S. Ramakrishnan and S. Selvan, "Image texture classification using wavelet based curve fitting and probabilistic neural network," *International Journal of Imaging Systems and Technology*, vol. 17, no. 4, pp. 266–275, 2007.

[9] X.-B. Wen, H. Zhang, X.-Q. Xu, and J.-J. Quan, "A new watermarking approach based on probabilistic neural network in wavelet domain," *Soft Computing*, vol. 13, no. 4, pp. 355–360, 2009.

[10] H. Adeli and A. Panakkat, "A probabilistic neural network for earthquake magnitude prediction," *Neural networks*, vol. 22, no. 7, pp. 1018–1024, 2009.

[11] S. Venkatesh and S. Gopal, "Orthogonal least square center selection technique–a robust scheme for multiple source partial discharge pattern recognition using radial basis probabilistic neural network," *Expert Systems with Applications*, vol. 38, no. 7, pp. 8978–8989, 2011.

[12] P. A. Kowalski and P. Kulczycki, "Data sample reduction for classification of interval information using neural network sensitivity analysis," in *Artificial Intelligence: Methodology, Systems, and Applications*, ser. Lecture Notes in Computer Science, D. Dicheva and D. Dochev, Eds. Springer Berlin Heidelberg, 2010, vol. 6304, pp. 271–272.

[13] ——, "Interval probabilistic neural network," *Neural Computing and Applications*, 2016. doi: 10.1007/s00521-015-2109-3 Online available.

[14] K. Elenius and H. G. Tråvén, "Multi-layer perceptrons and probabilistic neural networks for phoneme recognition." in *EUROSPEECH*, 1993.

[15] T. P. Tran, T. T. S. Nguyen, P. Tsai, and X. Kong, "Bspnn: boosted subspace probabilistic neural network for email security," *Artificial Intelligence Review*, vol. 35, no. 4, pp. 369–382, 2011.

[16] T. P. Tran, L. Cao, D. Tran, and C. D. Nguyen, "Novel intrusion detection using probabilistic neural network and adaptive boosting," *International Journal of Computer Science and Information Security*, vol. 6, no. 1, pp. 83–91, 2009.

[17] D. Montana, "A weighted probabilistic neural network," in *NIPS*, 1991, pp. 1110–1117.

[18] T. Song, M. Jamshidi, R. R. Lee, and M. Huang, "A novel weighted probabilistic neural network for mr image segmentation," in *Systems, Man and Cybernetics, 2005 IEEE International Conference on*, vol. 3. IEEE, 2005, pp. 2501–2506.

[19] T. Song, C. Gasparovic, N. Andreasen, J. Bockholt, M. Jamshidi, R. R. Lee, and M. Huang, "A hybrid tissue segmentation approach for brain mr images," *Medical and Biological Engineering and Computing*, vol. 44, no. 3, pp. 242–249, 2006. doi: 10.1007/s11517-005-0021-1

[20] S. Ramakrishnan and M. Emary, Ibrahiem, "Comparative study between traditional and modified probabilistic neural networks," *Telecommun. Syst.*, vol. 40, no. 1-2, pp. 67–74, 2009. doi: 10.1007/s11235-008-9138-5

[21] D. Nanjundappan *et al.*, "Hybrid weighted probabilistic neural network and biogeography based optimization for dynamic economic dispatch of integrated multiple-fuel and wind power plants," *International Journal of Electrical Power & Energy Systems*, vol. 77, pp. 385–394, 2016.

[22] M. Lichman, "UCI machine learning repository," 2013. [Online]. Available: http://archive.ics.uci.edu/ml

[23] J. M. Zurada, A. Malinowski, and S. Usui, "Perturbation method for deleting redundant inputs of perceptron networks," *Neurocomputing*, vol. 14, no. 2, pp. 177–193, 1997.

[24] J. Zurada, A. Malinowski, and I. Cloete, "Sensitivity analysis for minimization of input data dimension for feedforward neural network," in *Circuits and Systems, 1994. ISCAS '94., 1994 IEEE International Symposium on*, vol. 6, May 1994, pp. 447–450.

[25] P. A. Kowalski and P. Kulczycki, "A complete algorithm for the reduction of pattern data in the classification of interval information," *International Journal of Computational Methods*, vol. 13, no. 03, p. 1650018, 2016. doi: 10.1142/S0219876216500183

[26] B. W. Silverman, *Density estimation for statistics and data analysis*. CRC press, 1986, vol. 26.

[27] V. Vapnik, *The nature of statistical learning theory*. Springer Science & Business Media, 2013.

[28] D. E. Rumelhart, J. L. McClelland, and C. PDP Research Group, Eds., *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 1: Foundations*. Cambridge, MA, USA: MIT Press, 1986.

[29] D. S. Broomhead and D. Lowe, "Multivariable functional interpolation and adaptive networks," *Complex Systems*, vol. 2, pp. 321–355, 1988.

[30] J. A. Hartigan and M. A. Wong, "Algorithm as 136: A k-means clustering algorithm," *Applied statistics*, pp. 100–108, 1979.

[31] S. Chen, X. Wang, and C. J. Harris, "Experiments with repeating weighted boosting search for optimization signal processing applications," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 35, no. 4, pp. 682–693, 2005.

[32] P. H. Sherrod, "Dtreg predictive modelling software." [Online]. Available: http://www.dtreg.com

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 105–111
DOI: 10.15439/2016F513
ACSIS, Vol. 8. ISSN 2300-5963

Position tracking using inertial and magnetic sensing aided by permanent magnet

Michał Meina
Department of Informatics,
Faculty of Physics, Astronomy and Inf.
Nicolaus Copernicus University,
Grudziadzka 5, 87-100 Toruń, Poland

Krzysztof Rykaczewski
Faculty of Mathematics and Comp. Science,
Nicolaus Copernicus University,
Chopina 12/18, 87-100 Toruń, Poland

Andrzej Rutkowski
Faculty of Mathematics and Comp. Science,
Nicolaus Copernicus University,
Chopina 12/18, 87-100 Toruń, Poland

***Abstract*—This paper describes a method for spatial tracking of a strapdown device that can be used for design of human-computer interfaces. Inertial Measurement Unit (IMU) is used to obtain 6-dof position exploiting the so-called ZUPT technique by the means of the Kalman Filter. Additional corrections of position are done using magnetometer readings in the presence of static magnetic field induced by permanent magnet that overshadow geomagnetic field. This correction allows us to overcome drifting errors of integration of IMU readings. We have also presented comparisons of different models for magnetic field reconstruction that is crucial for this system.**

I. Introduction

HUMAN-Computer interfaces that uses hand gestures, object manipulation or any other pervasive technology always require position tracking subsystem. Recent developments in micro-electro mechanical systems (MEMS) enables researchers to build new wearable, position aware devices by utilization of miniaturized and low-power sensing devices. For example, inertial sensors are commonly used for construction of 6-dof (orientation and position) spatial tracking systems [1], [2]. Nevertheless, precision of such sensing devices is still very low, therefore it is needed to adopt more complex computations and signal filtering techniques. In this paper, we present an inertial position tracking system that uses additional correction by positioning in magnetic field induced by permanent magnet. Previous works exploited an array of magnetic sensors to unambiguously estimate position, while we propose a data fusion algorithm that uses just one magnetic sensor along with inertial sensor.

Data fusion algorithms are used commonly with inertial devices—in order to estimate orientation of the sensor, state-of-the art filter [3] integrate angular rate (obtained from gyroscope) in order to calculate rough quaternion rates and then uses gravitational force (obtained by accelerometer) to correct roll and pitch angles. Finally, magnetometer readings are exploited to correct yaw angle. The best known algorithm that can be used in this setup for position estimation exploits the so-called "zero-velocity" update (ZUPT). Second integral of acceleration outputs position with accumulative error. Additional statistics are used to test if the sensor is still—velocity, therefore, should be zero (most of the time it is not and one can use this information to correct position). This technique was proposed in [4] for inertial navigation using foot-mounted inertial measurement unit. We have extended the technique introducing Magnetic Update (MUPT) that introduces small correction of position if the sensor is exposed to magnetic field. In presented setup properties of magnetic field must be known beforehand, basing on extensive calibration procedure.

Contribution of this paper is as follow: (1) design and evaluation of position tracking system by fusion of inertial and magnetic sensing in the presence of magnetic field induced by permanent magnet, (2) evaluation of magnetic field reconstruction techniques, (3) positioning algorithm in that use

II. Related Work

Measurements of magnetic field have long been used in the problem of spatial tracking. Starting with the problem of locating buried magnet, one solution [5] used measurements of magnetic field generated by beacon device. In 1979 Raab et al. [6] proposed a complete system for relative position and orientation tracking using active 3-axis magnetic dipole source. Technique was further described and analysed on simulated data by Raab [7]. Similar methods are now used in commercial tracking solutions like *Polhemus Fastrack* or gaming controller *Razer Hydra*. Although these methods are useful in some situations, power requirements and high weight of magnetic coils disqualify it in context of wearable devices, with exception of specific usages as described in [8]. Since the advent of MEMS inertial and magnetic sensors, many solutions have been proposed for problem of position and orientation tracking using fusion of inertial and passive magnetic measurements [9], [10], [11]. Unfortunately, inertial systems suffer from systematic error accumulation. It is common [12], [13] to address this problem with ZUPT, as proposed in [4]. Although, as pointed out in [14], this solution suffers from many drawbacks and its usefulness is limited.

Geomagnetic field measurements are also susceptible to errors such as interference from ferromagnetic mass and electric devices. Zachmann tried correcting magnetic measurements by gathering calibration samples of magnetic field in a volume of tracking space [15]. Some authors tried using these distortions for detection and low accuracy tracking of metallic objects [16], [17]. Magnetic field produced by permanent magnets locally overshadow earth's, thus extensive research have been done on reverse problem of tracking permanently

978-83-60810-90-3/$25.00

magnetised markers with relation to magnetometer and (most often) arrays of magnetometers [18], [19], [20], [21].

Finally, in 2015 Kortier et al. [22] explored a system for tracking neodymium magnet paired with accelerometer and gyroscope with respect to array of four magnetometers in parallel with inertial sensor.

We propose a system where magnetometer is coupled with inertial sensors in a package and tracking is carried out with respect to neodymium magnet.

III. Modelling Vector Field Of A Permanent Magnet

In order to obtain the position of magnetic sensor in the presence of static magnetic field the straightforward idea is to use a model of the magnetic field and match the empirical measurement (obtained by magnetometers) to the theoretical one. The idea is to construct error function that will be minimized in order to find position of the sensor (or array of the sensors). This idea was investigated in many papers, e.g. [19], [21]. The technique requires finding a method for determining theoretical value of magnetic vector field.

More formally, we are searching for a function P such that

$$P_{\Theta_s^m}(x, y, z) = \vec{B}, \tag{1}$$

where $\vec{B}$ is a vector of magnetic induction in given position $(x, y, z) \in \mathbb{R}^3$ and Θ_s^m is an orientation of the sensing device s in the magnet frame of reference m (one can understand it as a transformation from sensor frame of reference to magnet frame of reference).

Magnetic induction vector in our case consists of three components: geomagnetic field $\vec{B}_{\text{geo}}$, field of permanent magnet $\vec{B}_{\text{mag}}$ and ferromagnetic part considered as environment noise $\vec{B}_{\text{env}}$. Therefore,

$$\vec{B} = \vec{B}_{\text{mag}} + \vec{B}_{\text{geo}} + \vec{B}_{\text{env}}. \tag{2}$$

For strong magnets, however, we can omit environmental and geomagnetic component, hence $\vec{B} \approx \vec{B}_{\text{mag}}$.

The model of the magnetic field is difficult to calculate in real time, so we are looking for some approximation of it. For this reason, we have investigated and compared different techniques: (1) dipole model, (2) integration from Maxwell equations using finite element method (FEM) simulators, and (3) field interpolation on empirical (and sparse) sampling. These methods are described below.

A. The Magnetic Dipole

Magnetic field of strong magnet can be approximated using *magnetic dipole* model [23]:

$$\vec{B}_{\vec{\mu}}(\vec{r}) = \frac{1}{\|\vec{r}\|^5}\left(3(\vec{\mu}\vec{r})\vec{r} - \vec{\mu}\|\vec{r}\|^2\right), \tag{3}$$

where μ is a *magnetic dipole moment* (that characterize magnetic field source) and $\vec{r}$ is a vector from the magnetic dipole source to the observation point. It is important to note that, since we assume that magnetic field source is infinitely small point, this model is accurate only when $\|\vec{r}\| \gg r_m$ (where r_m is a size of magnet).

We can rewrite Eq. (3) into components to obtain formulas that characterize components of $\vec{B} =: [B_x, B_y, B_z]$:

$$\begin{aligned} B_x &= 3|\mu|\tfrac{zy}{(x^2+y^2+z^2)^{\frac{5}{2}}}, \\ B_y &= 3|\mu|\tfrac{zx}{(x^2+y^2+z^2)^{\frac{5}{2}}}, \\ B_z &= |\mu|\tfrac{2(x^2-y^2)-z^2}{(x^2+y^2+z^2)^{\frac{5}{2}}}. \end{aligned} \tag{4}$$

From Equation (4) we see that in order to use the above model, we must carry out a calibration of the magnet (find the magnetic dipole moment μ for a particular magnet). This characteristic is included in the documentation that came with the magnet, nevertheless, we have found this parameter in the experimental section by defining an error function and making its optimization. In order to find this parameter, only one measurement in some known position is needed.

Usage of magnetic dipole moment for defining position estimation error function enables us to carry out the gradient function in analytical form. That makes minimization very fast.

B. Magnetostatic simulation using FEM

FEM (Finite Element Method) is a method of integrating differential equation used to find approximate solutions of many physical models. In our context it is used to solve the Maxwell equations. Magnetic dipole moment, which is a parameter for this FEM model, was calculated by experiment. The boundary conditions were chosen at far distance from field source in the manner that changing them will not affect the magnetic field in region of interest.

We used `QuickField`[1] library, that outputs the exact solution at some points and interpolates the rest. The position of the points are optimized.

C. Interpolation of a vector field

In this section, we propose a fast method to approximate the magnetic field using empirical readings—our magnetic field calibration procedure.

The method described below is based on measurements of magnetic field $\vec{B}$ at predetermined positions in the vicinity of the magnet, and then determining the interpolating function for any point in the area of interest D. For this purpose, we used the *Delaunay triangulation* (DT). The idea of this method is to find simplex, inside which there is a point of interest, and then counting the weighted sum of the vertices of the simplex (on which empirical measurement is known).

More formally, suppose we have made n measurements of magnetic field $\vec{B}(p_i)$, $i = 1, \ldots, n$, in points $p_i := (x_i, y_i, z_i)$, $i = 1, \ldots, n$.

Let us consider the division of space $\mathbb{R}^d$ on $(d+1)$-simplices $\{\sigma_j := \sigma(\{p_{i_k}\}_{k=0}^{K_j})\}_{j\in J}$, where $\sigma(\{p_l\}_{l=1}^{L})$ is the simplex determined by vertices p_l, $l = 1, \ldots, L$, such that:

- for $k \neq l$, k, $l \in J$, intersection of simplices $\sigma_k \cap \sigma_l$ is either a common wall of σ_k and σ_l, or is empty,

[1] http://www.quickfield.com/

- the interior of the circumscribed sphere on any simplex σ_j does not contain any point $\{p_i\}_{i=1}^n$.

Our area of interest D is the sum of all the above simplices. In the present context we only consider the division of $\mathbb{R}^2$ and $\mathbb{R}^3$, therefore, we can informally say that Delaunay triangulation divides the space into triangles and tetrahedrons, respectively, such that none of these objects is deformed (i.e. do not contain "very sharp" angles). A popular method for determining triangulation is based on the dual space (we show an example for $\mathbb{R}^2$): we consider the projection $(p_x, p_y) \mapsto (p_x, p_y, p_x^2 + p_y^2)$ and then calculate the convex hull of this set of points. Then the corresponding points will create DT.

Let $\vec{B}_{DT}(p)$ denotes the magnetic induction vector calculated by Delaunay interpolation at point p contained in the convex hull of p_i. For determining $\vec{B}_{DT}$ DT uses barycentric coordinates $\{\lambda_k\}_{k=1}^{d+1}$ of p with respect to $\{p_{i_k}\}_{k=1}^{d+1}$. In conclusion, we can represent it as a linear combination

$$\hat{\vec{B}}_{DT}(p) = \sum_{k=0}^{d+1} \lambda_k \hat{\vec{B}}(p_{i_k}) \tag{5}$$

such that $\sum_{k=1}^{d+1} \lambda_k = 1$ and $\lambda_1, \ldots, \lambda_{d+1} \geq 0$, where points p_{i_k} are vertices of simplex σ such that $p \in \sigma$. It is worth noting that calculation of coefficients λ_k for point $p = (p_x, p_y)$ in $\mathbb{R}^2$ (and, analogically, in $\mathbb{R}^3$) using reference points p^1, p^2, p^3, is limited to determining the solution of the following system of equations

$$\begin{bmatrix} p_x^1 & p_x^2 & p_x^3 \\ p_y^1 & p_y^2 & p_y^3 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{bmatrix} = \begin{bmatrix} p_x \\ p_y \\ 1 \end{bmatrix}. \tag{6}$$

From the computational complexity point of view, the most expensive operation is finding a simplex, inside which the point is located. To do this, one can apply a hierarchical decomposition of simplices, which represents collection of simplices in the form of binary search tree. Asymptotic cost of finding simplex is, therefore, $\mathcal{O}(\log n)$, however, because of need for preprocessing and higher storage requirements, such representation may be not suitable for implementation on small microprocessors. Best alternative seems to be a classic *walking* algorithm [24]. As noted in [25], Mücke shows [26] that careful use of *walking* algorithm can bring down expected time close to $\mathcal{O}(n^{1/4})$ (for $d = 3$).

D. Magnetic Field Reconstruction

In order to check its reliability our methodology was as follows: we recorded vectors of magnetic induction in many places on common plane and compared it with the theoretical model. In this section, we will present experimental result of reconstruction of magnetic field.

The magnetic field was generated by arrangement of three cylindrical neodymium magnets ($\oslash = 22\,[\text{mm}]$, $h = 10\,[\text{mm}]$, direction of magnetization was along shorter axis). According to specification, at the distance of $0.7\,[\text{mm}]$ in the direction of magnetization the magnetic field strength was $\approx 0.380\,[\text{T}]$. Measurement was performed with MEMS magnetometer LSM303D, that is able to measure magnetic field of maximum strength of $\pm 1.2\,[\text{mT}]$ with 16 bits resolution (values with 16 meaningful bits). See Fig. 1 for our experimental setup.

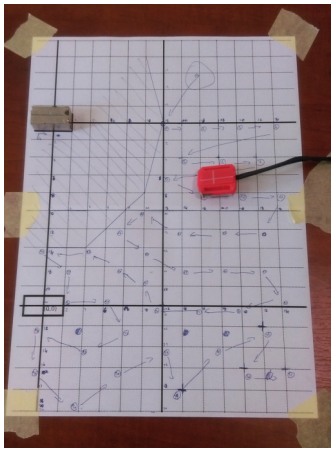


Fig. 1: Experimental setup.

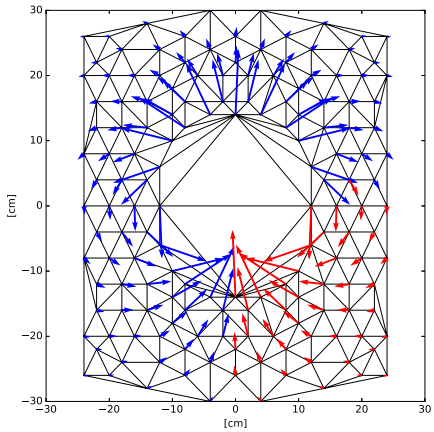


Fig. 2: Magnetic field by Delaunay triangulation. Red arrows depicts empirical measurements whereas blue arrows represent symmetrical reflection of those.

At the beginning, we checked properties of the field with magnets arranged in the above manner.

We've done several experiments, in order to show that magnetic field around the cylinders is symmetrical. Moreover, several experiments have been performed to check the accuracy of different models. Geomagnetic field was measured beforehand and discarded from both calibration dataset and test dataset. Experimental station is illustrated in Fig. 1. Table III-D shows the results of different tests.

Test 1 was divided into several subtest: Iq, IIq, IIIq, IVq, r<20, r>=20. It consists of measurements when magnet was not rotated.

Test 2, 3 and 4 consists of measurements with magnet rotated along magnetization axis by 90, 180 and 270 degrees, respectively, without changing its location. Each test consisted of 54 measurements in places which were different than

TABLE I: Comparison of magnetic field reconstruction using different models.

	$\measuredangle \vec{B_e}\vec{B_m}$ [deg]			$10^{-2}\|B\|$ [gauss]		
	dipole	fem	interpolant	dipole	fem	interpolant
test 1 / 0°	19.37 ± 9.28	7.89 ± 4.96	3.62 ± 3.57	72.17 ± 72.09	113.71 ± 50.49	14.15 ± 16.11
Iq	17.96 ± 9.12	7.67 ± 4.71	1.62 ± 1.32	65.16 ± 55.32	127.09 ± 62.41	5.62 ± 7.94
IIq	22.81 ± 8.03	10.40 ± 4.41	5.82 ± 3.81	59.31 ± 56.11	108.68 ± 23.94	15.22 ± 19.41
IIIq	20.57 ± 9.29	7.68 ± 5.24	3.99 ± 4.78	70.38 ± 63.26	111.28 ± 50.20	20.91 ± 19.18
IVq	18.70 ± 8.15	6.41 ± 4.57	3.15 ± 2.10	58.69 ± 54.25	124.10 ± 39.39	11.07 ± 6.31
r<20	23.09 ± 10.74	11.69 ± 5.02	2.67 ± 1.99	128.76 ± 91.99	122.44 ± 75.47	25.12 ± 22.40
r>=20	17.52 ± 7.83	5.99 ± 3.66	4.09 ± 4.06	43.88 ± 34.11	109.34 ± 30.30	8.66 ± 6.95
test 2 / 90°	19.46 ± 9.50	8.24 ± 5.31	4.07 ± 3.60	72.07 ± 72.16	113.55 ± 51.35	14.06 ± 15.79
test 3 / 180°	19.31 ± 9.24	7.98 ± 4.75	3.70 ± 3.64	72.08 ± 71.99	114.08 ± 51.12	13.56 ± 14.43
test 4 / 270°	19.08 ± 9.21	7.63 ± 5.05	3.77 ± 3.56	71.92 ± 71.63	113.67 ± 50.58	15.53 ± 17.49

calibration points. Since magnetometer has maximum range of measurements, readings were taken at distances of less that 12 [cm] from the magnet.

Left part of the table shows angle of deviation between model magnetic field $\vec{B}_m$ and experimentally measured magnetic field $\vec{B}_e$. Right part of the table shows means and standard deviations of norms between $\vec{B}_m$ and $\vec{B}_e$ (i.e. $|\,\|\vec{B}_m\| - \|\vec{B}_e\|\,|$).

In Table III-D one can see that error is similar in each model, indicating that the measured magnetic field is indeed symmetrical along the axis of magnetization. Differences between quarters of the coordinate system are larger, but this is probably due to inaccuracies of measurement. Dipole model (as expected) makes smaller error (in particular, when it comes to the norm of $\vec{B}$) when we measure at distant places ($r \geq 20$ [cm]) from the source field. Field calculated using FEM is, roughly speaking, two times better than the dipole model. Our method was marked as "interpolant" and is about five times better that dipole and about 2 times better that FEM.

Bearing in mind the above, we assume that around cylindrical magnetic the field is symmetrical, therefore calibration measurements were taken only on the one selected quadrant of a plane.

Delaunay triangulation is depicted in Fig. 2. We see there 42 measurements of $\vec{B}$ (depicted as red arrows) and 115 vectors (blue arrows) that were derived using symmetrical reflection of those 42.

IV. Positioning in empirically interpolated magnetic field

Positioning within the magnetic field using a magnetometer is usually performed using the layout of sensors and the dipole model. Dipole model considered for its own sake is biased by error. However, it can be compensated by using more readings and their relative positions. Additionally, a more accurate position can be determined with a greater number of sensors.

Therefore, in order to calculate the precise position it is usually necessary to use at least two sensors. However, in the paper we want to get rid of this constraint and estimate the position with a single magnetometer in the magnetic field. Hence, we cannot mitigate errors of the model by simultaneous readings and, therefore, we can find only one parameter: either the location (having a fixed orientation) or the orientation (having a fixed position). For that we will minimize the *cost function* defined as follows

$$f_{\vec{B}_e}(p) = |\vec{B}_m(p) - \vec{B}_e|^2 + (|\vec{B}_m(p)| - |\vec{B}_e|)^2, \qquad (7)$$

where $\vec{B}_e$ is experimentally measured magnetic field and $\vec{B}_m(p)$ is the value of model magnetic field at point p. Calculation of $\vec{B}_m(p)$ is the most computationally expensive operation.

Since this function is convex in a neighbourhood of current position, we know that (locally) the minimum value of function (7) is uniquely determined. As it was stated above, in the presence of at least two sensors we can use dipole model. Therefore, error function is smooth and minimization can be done by gradient descent algorithm which it terminates after just a few iterations. In the case of our model, the error function is created from DT and since the data is collected empirically we do not have smoothness any more. That is why we use nongradient methods, e.g. Nelder-Mead. This algorithm also terminates after few iterations, but it needs many more error function evaluations.

In Fig. 3 we show the magnitudes of error function (Eq. 7) with respect to number of evaluations during the execution of minimization, starting from two initial positions. For each minimization these were selected by first randomly choosing known destination point and then selecting random position from its neighbourhood within the range of 1 [cm] and 4 [cm], respectively. The thick line on the chart indicates the average and is surrounded by a plot of the standard deviation at each iteration.

V. Fusion of inertial positioning and magnetic

The fusion algorithm is implemented by the means of Kalman Filter which estimates the error state vector $\delta x_k = [\delta v_k\ \delta p_k\ \delta C_k]$, which represents velocity, position and orientation (roll, pitch, yaw) errors. Details of the algorithm are described in Alg. 1—as an input it takes readings from IMU (a, ω, m—which are acceleration, angular rates and magnetic induction vector respectively).

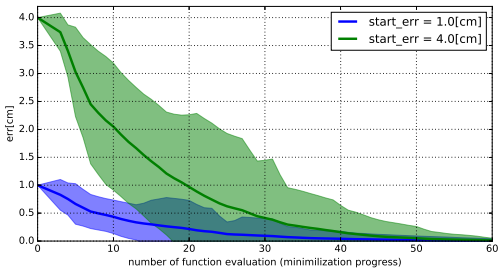


Fig. 3: Converge of minimization.

State transition matrix F_k is defined as follow:

$$F_k := \begin{bmatrix} I_{3\times3} & \Delta t I_{3\times3} & 0_{3\times3} \\ 0_{3\times3} & I_{3\times3} & 0_{3\times3} \\ \Delta t S_k & 0_{3\times3} & I_{3\times3} \end{bmatrix}, \tag{8}$$

where S_k is a skew-symmetric cross-product operator matrix given by:

$$S_k := \begin{bmatrix} 0 & -a^N_{z,k} & a^N_{y,k} \\ a^N_{z,k} & 0 & -a^N_{x,k} \\ -a^N_{y,k} & a^N_{x,k} & 0 \end{bmatrix}. \tag{9}$$

The S_k matrix is used for accumulating orientation errors with respect to velocity error estimates. This method is inspired by technique called "alignment transfer" [27] and was introduced to ZUPT-based inertial navigation systems by Foxlin [4]. In our algorithm we updated the orientation estimation from original Foxlin work by usage of complementary filter (line 6) which was developed by Madgwick [3]. In our previous work [28] we showed his setup works better for orientation estimation for the object in motion with zupt-based filters.

At line 11 two statistical test are performed: one for zero velocity hypothesis testing (`IsZUPT`) and one for position estimation error test (`IsMUPT`). For the zero-velocity hypothesis testing we apply the following formula at k-th reading:

$$\mathrm{Var}\left(\left\{\sqrt{\omega_x^2(t)+\omega_y^2(t)+\omega_z^2(t)} \mid t = k-w,\ldots,k\right\}\right) < \epsilon_1 \tag{10}$$

Simply speaking Eq. 10 is a windowing w-length function that checks if variance of gyroscope readings do not exceed certain threshold (ϵ_1). For zero-velocity hypothesis testing there exist lots of solution in the literature (see [29], [30], [31])

In order to check if the correction based on magnetic positioning observation need to be performed another test is executed: `IsMUPT`. This test checks if covariance error didn't exceed certain threshold:

$$\sum_{i=3}^{6}(P_{i\times i}) < \epsilon_2 \tag{11}$$

Eventually at lines 19-23 the Kalman Update is performed basing on two version of observation matrix H; one for velocity pseudo-observation update (line 13) and second one or true position observation (line 16).

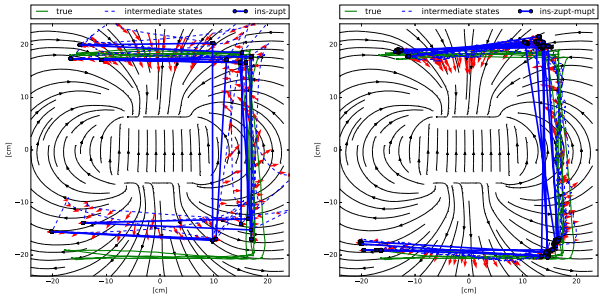
Algorithm 1 Psuedocode for inertial positioning system with zero-veolcity and magnetic positioning updates.

1: **procedure** INS-ZUPT-MUPT(a, ω, m, g_0)
2: $k \leftarrow 0$,
3: $q_0 \leftarrow$ INITIALIZEORIENTATION(g_0)
4: **loop**
5: $k \leftarrow k+1$
6: $q_k \leftarrow$ UPDATEORIENTATION(a_k, ω_k)
7: $a_{nav} \leftarrow q_k^{-1} \otimes a_k \otimes q_k - g_0$
8: $v_k^{(+)} \leftarrow v_{k-1} + a_{nav}\Delta t$
9: $p_k^{(+)} \leftarrow p_{k-1} + v_{k-1}\Delta t$
10: $P_k^{(+)} = FP_{k-1}F^T + Q$
11: **if** ISZUPT(k, a, ω) or ISMUPT($P_k^{(+)}$) **then**
12: **if** ISZUPT(k, a, ω) **then**
13: $H \leftarrow \begin{bmatrix} 0_{3\times3} & I_{3\times3} & 0_{3\times3} \end{bmatrix}$
14: $z_k \leftarrow 0_{3\times1}$
15: **else**
16: $H \leftarrow \begin{bmatrix} I_{3\times3} & 0_{3\times3} & 0_{3\times3} \end{bmatrix}$
17: $z_k \leftarrow \mathrm{argmin}\, f_m(p_k^{(+)})$
18: **end if**
19: $K \leftarrow P^{(+)}H^T(HP^{(+)}H^T + R)^{-1}$
20: $P_k \leftarrow I - KHP_k^{(+)}$
21: $[\delta v_k\ \delta p_k\ \delta C_k] \leftarrow K(z_k - Hx_k)$
22: $q_k \leftarrow$CORRECTORIENTATION(C_k)
23: $[p_k, v_k] \leftarrow [p_k, v_k] - [\delta p_k, \delta v_k]$
24: **end if**
25: **end loop**
26: **end procedure**

VI. EXPERIMENTAL RESULTS

Experiment was conducted using sensor and a magnet, which properties were described in section III-D. The sensor was moved freely over the table with moderate speed in various trajectories (with and without still phases). The true path was obtained using optical system, composed from two cameras that were observing diode mounted on top of the sensor. Example of such movement is depicted on the Fig. 4. Fig. 4b depicts trajectories computed by algorithm described in previous section (blue lines), whereas Fig. 4a shows the same algorithm in which no magnetic update has been made. Note that path is depicted only between zero-velocity phases—the *a posterori* states. The *a priori* states are depicted by the dotted line as "intermediate states". The read arrows indicates magnetic field measurements done at particular place and the green line is the true path observed by optical system. As a visual aid magnetic field was enclosed using interpolation method—black field lines.

Error drift in inertial navigation system describes the situation where there is a random inaccuracy introduced at each step of computation. It source lays in (1) sensor error and (2) floating-point computation. It is very challenging (or impossible) to model this error, therefore there is no method for position tracking basing only on IMU readings.



(a) Inertial System with zero-velocity updates.

(b) Inertial system with zero-velocity updates and magnetic positioning.

Fig. 4: Comparison of position tracking algorithms.

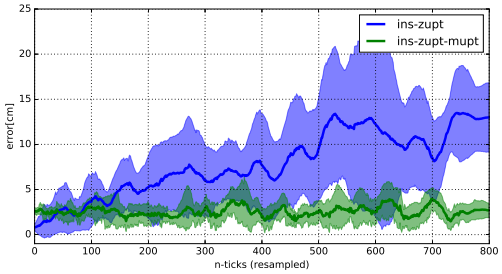


Fig. 5: Position estimation errors.

After prototyping a system with another source of correction the most important thing to check is if it's error does not accumulate over time.

The absolute error (measured as Euclidean distance between the true and estimated path of two methods over time) is depicted on Fig. 5. The error is computed as an average over all 10 samples of different trials along with standard deviation. The most important observation is that that ins-zupt-mupt does not drift over time ($p < 0.01$ using *Augmented Dickey-Fuller Test* for unit root testing of processes—stationary test), while ins-zupt accumulates errors over time. The trials without zero-velocity phases were not enclosed into the results since its error was too significant.

VII. Conclusion and Future Works

In this paper, position tracking system based on data fusion from inertial measurement unit and positioning in magnetic field was presented. Inertial navigation was performed using the so-called zero-velocity updates and modelled by the means of the Kalman Filter. The common problem with error accumulation of inertial system has been solved by positioning sensor in the presence of static magnetic field.

The presented method enables us to build more robust human-computer interfaces which requires positioning subsystem, because it requires just one sensor for magnetic field sensing. Static magnetic field can be easily generated with no energetic cost by various (in terms of size and strength) neodymium magnets and the magnetometer most of the time is already mounted into IMU unit.

The previous works consisted of array of sensor that was rigidly mounted, which can introduce some problem in real-life systems. Usage of just one sensor, however, introduces some drawbacks: (1) the field of magnet needs to be known precisely in advance, therefore a calibration procedure needs to be introduced in advance (2) from theoretical point of view the additional corrections that are made are not unambiguous—the cost function Eq. 7 minimizes position error with given orientation. The second problem could lead to error in yaw-angle estimation.

In this paper authors did not investigated orientation estimation errors. Nevertheless, the presented algorithm corrects the orientation estimates in zero-velocity phases. The estimates are sufficiently good for most applications because orientation estimation is less prone to errors since it can be corrected by data fusion filter and needs only one integration. Nevertheless, in future work promising idea is to estimate orientation errors taking the whole path into consideration by forming some lag-introducing corrections. The naive idea is to correct orientation minimizing the magnetic field vector derivation from theoretical model in more than one points, assuming some tuned springiness between the measuring points.

References

[1] F. Ayazi, "Multi-dof inertial mems: From gaming to dead reckoning," in *2011 16th International Solid-State Sensors, Actuators and Microsystems Conference*, 2011.

[2] D. E. Serrano, "Design and analysis of mems accelerometers," in *IEEE Sensors*, 2013.

[3] S. O. Madgwick, "An efficient orientation filter for inertial and inertial/magnetic sensor arrays," *Report x-io and University of Bristol (UK)*, 2010.

[4] E. Foxlin, "Pedestrian tracking with shoe-mounted inertial sensors," *Computer Graphics and Applications, IEEE*, vol. 25, no. 6, pp. 38–46, 2005.

[5] R. Olsen and A. Farstad, "Electromagnetic direction finding experiments for location of trapped miners," *Geoscience Electronics, IEEE Transactions on*, vol. 11, no. 4, pp. 178–185, 1973.

[6] F. H. Raab, E. B. Blood, T. O. Steiner, and H. R. Jones, "Magnetic position and orientation tracking system," *Aerospace and Electronic Systems, IEEE Transactions on*, no. 5, pp. 709–718, 1979.

[7] F. H. Raab, "Quasi-static magnetic-field technique for determining position and orientation," *Geoscience and Remote Sensing, IEEE Transactions on*, no. 4, pp. 235–243, 1981.

[8] D. Roetenberg, P. J. Slycke, and P. H. Veltink, "Ambulatory position and orientation tracking fusing magnetic and inertial sensing," *Biomedical Engineering, IEEE Transactions on*, vol. 54, no. 5, pp. 883–890, 2007.

[9] E. R. Bachmann, I. Duman, U. Usta, R. B. Mcghee, X. Yun, and M. Zyda, "Orientation tracking for humans and robots using inertial sensors," in *Computational Intelligence in Robotics and Automation, 1999. CIRA'99. Proceedings. 1999 IEEE International Symposium on*. IEEE, 1999, pp. 187–194.

[10] R. Zhu and Z. Zhou, "A real-time articulated human motion tracking using tri-axis inertial/magnetic sensors package," *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, vol. 12, no. 2, pp. 295–302, 2004.

[11] X. Yun, J. Calusdian, E. R. Bachmann, and R. B. McGhee, "Estimation of human foot motion during normal walking using inertial and magnetic sensor measurements," *Instrumentation and Measurement, IEEE Transactions on*, vol. 61, no. 7, pp. 2059–2072, 2012.

[12] P. Robertson, M. Angermann, B. Krach, and M. Khider, "Inertial systems based joint mapping and positioning for pedestrian navigation," in *Proc. ION GNSS*, 2009.

[13] Ö. Bebek, M. A. Suster, S. Rajgopal, M. J. Fu, X. Huang, M. C. Çavuşoğlu, D. J. Young, M. Mehregany, A. J. Van den Bogert, and C. H. Mastrangelo, "Personal navigation via high-resolution gait-corrected inertial measurement units," *Instrumentation and Measurement, IEEE Transactions on*, vol. 59, no. 11, pp. 3018–3027, 2010.

[14] J.-O. Nilsson, I. Skog, and P. Händel, "A note on the limitations of zupts and the implications on sensor error modeling," in *2012 International Conference on Indoor Positioning and Indoor Navigation (IPIN), 13–15th November 2012*, 2012.

[15] G. Zachmann, "Distortion correction of magnetic fields for position tracking," in *Computer Graphics International, 1997. Proceedings*. IEEE, 1997, pp. 213–220.

[16] R. Alimi, N. Geron, E. Weiss, and T. Ram-Cohen, "Ferromagnetic mass localization in check point configuration using a levenberg marquardt algorithm," *Sensors*, vol. 9, no. 11, pp. 8852–8862, 2009.

[17] N. Wahlstrom, J. Callmer, and F. Gustafsson, "Magnetometers for tracking metallic targets," in *Information Fusion (FUSION), 2010 13th Conference on*. IEEE, 2010, pp. 1–8.

[18] W. Weitschies, R. Kötitz, D. Cordini, and L. Trahms, "High-resolution monitoring of the gastrointestinal transit of a magnetically marked capsule," *Journal of pharmaceutical sciences*, vol. 86, no. 11, pp. 1218–1222, 1997.

[19] V. Schlageter, P.-A. Besse, R. Popovic, and P. Kucera, "Tracking system with five degrees of freedom using a 2d-array of hall sensors and a permanent magnet," *Sensors and Actuators A: Physical*, vol. 92, no. 1, pp. 37–42, 2001.

[20] X. Wang, M. Q. Meng, and C. Hu, "A localization method using 3-axis magnetoresistive sensors for tracking of capsule endoscope," in *Engineering in Medicine and Biology Society, 2006. EMBS'06. 28th Annual International Conference of the IEEE*. IEEE, 2006, pp. 2522–2525.

[21] J. T. Sherman, J. K. Lubkert, R. S. Popovic, and M. R. DiSilvestro, "Characterization of a novel magnetic tracking system," *Magnetics, IEEE Transactions on*, vol. 43, no. 6, pp. 2725–2727, 2007.

[22] H. G. Kortier, J. Antonsson, H. M. Schepers, F. Gustafsson, and P. H. Veltink, "Hand pose estimation by fusion of inertial and magnetic sensing aided by a permanent magnet," *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, vol. 23, no. 5, pp. 796–806, 2015.

[23] J. B. M. Mark A. Heald, *Classical Electromagnetic Radiation*, 3rd ed. Brooks Cole, 1995.

[24] P. J. Green and R. Sibson, "Computing dirichlet tessellations in the plane," *The Computer Journal*, vol. 21, no. 2, pp. 168–173, 1978.

[25] H. Ledoux, *Modelling three-dimensional fields in geoscience with the Voronoi diagram and its dual*. University of Glamorgan, 2006.

[26] E. P. Mücke, I. Saias, and B. Zhu, "Fast randomized point location without preprocessing in two-and three-dimensional delaunay triangulations," in *Proceedings of the twelfth annual symposium on Computational geometry*. ACM, 1996, pp. 274–283.

[27] O. Hallingstad, "Design of a kalman filter for transfer alignment," DTIC Document, Tech. Rep., 1989.

[28] M. Meina, A. Krasuski, and K. Rykaczewski, "Model fusion for inertial-based personal dead reckoning systems," in *Sensors Applications Symposium (SAS), 2015 IEEE*. IEEE, 2015, pp. 1–6.

[29] A. Peruzzi, U. D. Croce, and A. Cereatti, "Estimation of stride length in level walking using an inertial measurement unit attached to the foot: A validation of the zero velocity assumption during stance," *Journal of Biomechanics*, vol. 44, no. 10, pp. 1991–1994, 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0021929011003666

[30] A. R. Jimenez, F. Seco, C. Prieto, and J. Guevara, "A comparison of pedestrian dead-reckoning algorithms using a low-cost mems imu," in *Intelligent Signal Processing, 2009. WISP 2009. IEEE International Symposium on*, Aug 2009, pp. 37–42.

[31] I. Skog, J. O. Nilsson, and P. Händel, "Evaluation of zero-velocity detectors for foot-mounted inertial navigation systems," in *Indoor Positioning and Indoor Navigation (IPIN), 2010 International Conference on*, Sept 2010, pp. 1–6.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 113–120
DOI: 10.15439/2016F187
ACSIS, Vol. 8. ISSN 2300-5963

Classification Algorithms in Sleep Detection—A Comparative Study

Aleksandra H. Pasieczna, Jerzy J. Korczak
Wrocław University of Economics
ul. Komandorska 118/120, 53-345 Wrocław, Poland
Contact: wwwpasiecznapl@gmail.com (AHP), jerzy.korczak@ue.wroc.pl (JJK)

***Abstract*—This paper presents a comparison of different machine learning algorithms applied to automatic sleep detection which uses electroencephalogram signals as a differentiating basis. The Single-Layer Perceptron, Multi-Layer Perceptron, Support Vector Machine, Boosted Tree and the Multi-Agent (comprising of the earlier models) models are developed and analyzed with training and testing datasets. The results of the models are evaluated using a cross-validation technique. The models are compared with each other using the Cohen's index, the True Positive Rate and True Negative Rate. The models are very successful with sleep stage detection reaching up to 94 %, and Cohen's index reaching up to 0.69, showing considerable promise for deployment and future studies.**

I. Introduction

ROAD accidents are responsible for many deaths and economic losses around the globe. Drowsiness is a cause of about 40,000 non-fatal injuries and 1550 fatalities annually in the United States alone [1]. A press release by the National Sleep Foundation in 2009 indicates that 1.9 million American drivers have had a car crash or 'near-miss' due to fatigue or drowsiness [2]. The National Sleep Foundation has data which shows that even some flight and train accidents have drowsiness and fatigue as probable causes [3].

In general, to detect a driver's sleep stage the frequency spectra of brainwave signals are used as a differentiating factor. A good detection algorithm should accurately detect a drowsy driver, while not causing many false alarms. In principle, a model may detect drowsiness all the time, i.e., a success rate of 100 % for sleep detection, but this will create many false alarms, which would affect the driving behavior. A false alarm is triggered when a wake stage is detected as a sleep stage. For our classification purposes, a sleep stage is defined as the first stage of sleep. Earlier work using multichannel electroencephalogram (EEG) signals obtained 92.8 % internal sleep stage classification accuracy [4]. This work did not classify wake stages, which would be included in our models. The aim of our project is to obtain similar or better results for sleep detections using only one channel, i.e. fewer data input parameters. It is more important for us to detect a sleep stage than have fewer false alarms, which is why the sleep detection success rate has a high benchmark.

To train and test the algorithms before grading them, the dataset used was obtained from the DREAMS project [5]. This dataset consisted of raw polysomnographic (PSG) signals. A data-mining plan was designed for transforming the dataset, for training and testing the models, and for evaluation of the models. This plan expanded on the existing widely-used CRISP-DM model [6].

In the DREAMS project, PSG data was collected from sleeping patients, which included the electroencephalograph signals, the backbone of our project. Automatic classification of the sleep stages and sleep disorders was a major part of the different studies performed by the DREAMS project team. They showed that it is possible to perform automatic classifications using machine learning algorithms.

In 2001, Van Hese P. *et al.* [7] performed an automatic detection of the sleep stages using only EEG data. They used a modified K-means algorithm, and the parameters they used for classification were the parameters of Hjorth – Activity, Mobility and Complexity – expressed in terms of the frequency spectrum. The mobility was a measure of the central frequency, the complexity was a measure of the bandwidth of the signal, and the activity was a measure of the variance. They showed that clusters were created and it was possible to distinguish between the different stages, but that extra information, like electrocardiograph (ECG) and electrooculograph (EOG) data, was necessary for a clear discrimination between the stages.

Zhovna I. and Shallom I. proposed in 2008, an automatic sleep stage detection and classification of multichannel EEG signals [4]. They used a method based on the Multichannel Auto Regressive (MAR) model, which incorporated the cross-correlation information existing between the different EEG signals. Their approach involved training and testing phases, including a continuous unknown 7-hour subject dataset. They obtained a promising rate of detection of 92.8 %. However, they had not trained the system to detect wake stages or movements.

A more recent study in 2013 by Malaekah E. and Cvetkovic D. tried to perform an automatic detection scheme to classify the sleep stage 1 and the wake stage using the EEG sub-epoch approach [8]. They divided 30 second epochs into 6 second sub-epochs from which the Relative Spectral Energy Band (RSEB) was calculated and used to create the feature space. The RSEB for a given frequency band is defined as the ratio of the Power Spectral Density (PSD) of the band divided to the total power (sum of PSD of all bands). Their best results showed an average of 77 % and 55.8 % success of detecting the wake and first sleep stages respectively.

978-83-60810-90-3/$25.00

The work done by different research groups shows that EEG signals are good indicators of sleep and wake stages. For manual classification, there are two references, [9] and [10], both of which rely on the frequency of the EEG signals. Following up on the works of [7] and [8], it seems that using the complete frequency spectrum instead of few parameters might improve the classification accuracy.

The main goal of the paper is to create an automatic detection mechanism using data mining approach. In the project, the standard CRISP DM methodology was applied [6]. The first step is to understand the important physiological processes of sleep and their mathematical representation as signals. The following section provides a brief description of the brainwave signals along with their classification based on the frequency range. Then the data preparation phase is explained followed by an introduction to the classification models used in this project. In the third section, the experimental results for the different models are tabulated in terms of the true positive and true negative rates, also known as sensibility and specificity respectively. The paper ends with the conclusions drawn from our work and perspectives to continue research in this domain.

II. Data Sources and Classification Models

A. EEG signals

An electroencephalograph (EEG) represents electrical signals which reflect the elec-trical activity of the neurons. Based on their frequency range, an EEG signal is divided into bands denoted by Greek letters [11]:

- Beta waves – β-waves are brainwaves with frequencies between 13 and 30 Hz. They indicate full awareness and high brain activity.
- Alpha waves – α-waves have frequencies between 8 and 13 Hz. These are generally linked to relaxed states.
- Theta waves – θ-waves lie in the frequency range of 4 to 8 Hz. These are connected to Non-Rapid Eye Moment (NREM) sleep.
- Delta waves– δ-waves are below 4 Hz in the frequency spectrum. They generally correspond to slow wave sleep stages.

There are two standards to classify EEG data into sleep stages. The older Rechtschaffen and Kales (R&K) standard is composed of the following stages – Wake stage, REM stage, Sleep stage S1, Sleep stage S2, Sleep stage S3, and Sleep stage S4 [9]. In the more recent American Academy of Sleep Medicine (AASM) standard, the stages are classified as – Wake stage, REM stage, Sleep stage N1, Sleep stage N2, and Sleep stage N3 [10]. However, the DREAMS dataset contains additional stages. There is one sleep stage movement (transition) present in the R&K classification. The additional sleep stages are defined as *unknown sleep stages* by the classifying experts. The terminology used to denote the sleep stages is defined as:

- 0 = Wake stage or REM (R&K and AASM)
- 1 = Sleep stage 1 (R&K and AASM) or Sleep stage 2 (R&K and AASM)
- 2 = Sleep stage 3 (R&K and AASM) or Sleep stage 4 (R&K) or unknown sleep stage (R&K)
- 3 = Sleep stage movement (R&K) or an unknown sleep stage (AASM)

The reason why data points which might not seem as a wake stage or sleep stage S1/N1, are included is that when one model predicts a sleep stage 1, the other model might predict a sleep stage 3, an unknown sleep stage, or a sleep stage movement (transition).

B. Data Preparation

The dataset used in the current work was collected as part of the DREAMS project in Belgium and consists of 32-channel polygraph whole-night polysomnographic readings of 20 healthy subjects [5]. At least three of the 32 channels were EEG channels. The data was collected with a sampling frequency of 200 Hz and stored in the standard European Data Format (EDF). The channel chosen by us for the project was the channel CZ-A1, a central lobe channel. Only the first sleep and wake stages were selected from the files along with the corresponding EEG data. Points, which did not have sleep stage 1 or the wake stage in one rating model but in the other, were included, and the total number of data points after this phase was 32924. Each data point consists of one hypnogram rating based on the R&K model, and one hypnogram rating based on the AASM model and 1000 raw EEG signal points corresponding to a manual rating of a 5-second time-window.

These data points were then read into the statistical software R for further pro-cessing. This is a very large volume of data (32924×1000) to be provided for a machine learning algorithm. To reduce the size, feature selection was performed. Each data point was first transformed as per the Fourier transform to obtain the frequency spectrum from 0 Hz to 100 Hz (half the sampling frequency) with an accuracy of 0.2 Hz (inverse of the length of the signal). The sleep manuals indicate that frequencies above 30 Hz do not provide information for sleep stages and thus frequencies above 30.5 Hz were rejected. The size of the data was still large (32924×153). The frequencies were then averaged to the nearest integer frequency to reduce the dimension of the problem. After the data selection, transformation, and feature selection, the size of the data was 32924×31 (31, including the 0 Hz value). This final data was passed to the R platform, a statistical software package extensively used for analyzing large data.

C. Quantitative Indicators Used for the Models

It is important to describe the quantitative indicators used to grade the models in this work. Two models for each machine learning algorithm were created, one corresponding to the AASM scoring method and the other to the R&K scoring method. Thus, for consistency purposes, it is vitally important to *not* compare the AASM models with the R&K models.

There are many indicators, which can be used to grade a classification model. In our paper, the confusion matrix (see Table I) is first created, where TN stands for *True Negatives*, TP for *True Positives*, FN for *False Negatives* and FP for

TABLE I
EXAMPLE OF A CONFUSION MATRIX

		Model's Output 0	Model's Output 1
Manual Rating	0	TN	FP
	1	FN	TP

False Positives. These values can be used to grade the models, however they depend directly on the number of sleep and wake states. Since the database consists of unequal number of sleep and wake states, and to thus properly normalize the results, the following indicators, along with the Cohen's Index, were used:

- *True Positive Rate* (TPR) – This value is known as the sensitivity of the model, and provides an estimate for the successful sleep detection rate by the model. It is defined,
$$\mathrm{TPR} = \frac{\mathrm{TP}}{(\mathrm{TP} + \mathrm{FN})}, \tag{1}$$
and is also linked to the *False Negative Rate* (FNR) through TPR = 1 - FNR [12].
- *True Negative Rate* (TNR) – This value is known as the specificity of the model, and provides an estimate for the successful wake detection rate by the model. It is defined as,
$$\mathrm{TNR} = \frac{\mathrm{TN}}{(\mathrm{TN} + \mathrm{FP})}, \tag{2}$$
and is also linked to the *False Positive Rate* (FPR) through TNR = 1 - FPR [12].
- Cohen's Index (κ) – κ is an indicator that measures the inter-rater agreement for categorical objects, and that takes into account the agreements (TP and TN) occurring by chance [13]. It is defined as:
$$\kappa = \frac{p_o - p_e}{1 - p_e}, \tag{3}$$
where p_o is the observed agreement between the raters, given as:
$$p_o = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}}, \tag{4}$$
and p_e is probability of chance agreement, described as:
$$\begin{aligned} p_e = &\frac{(\mathrm{TP} + \mathrm{FN})(\mathrm{TP} + \mathrm{FP})}{(\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN})^2} \\ &+ \frac{(\mathrm{TN} + \mathrm{FN})(\mathrm{TN} + \mathrm{FP})}{(\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN})^2}. \end{aligned} \tag{5}$$
It can be easily verified that $\kappa = 1$ when the two raters are in complete agreement. In general, a value of $\kappa > 0.6$ indicates a good level of agreement between the raters.

In classification models, there is a trade-off between the TPR and TNR, since increasing one generally reduces the other. Thus it is necessary to study both simultaneously, making the TPR and TNR important indicators of the models. These are simple indicators and to consider random agreement, κ was also used to grade the models.

The following subsections will present the classifying algorithms used in our experiments – Single-Layer Perceptron, Multi-Layer Perceptron, Support Vector Machine, XGBoost and Multi-Agent model – whose classification accuracies were compared, in order to choose the best model. The SLP is a simple linear classifier that calculates the sum of the weighted input variables. It is simple to implement and forms a good reference model. The MLP, also known as neural networks, consists of layers of neurons. This provides the next level of complexity and allows for non-linear behavior to be captured. The SVM models have better predictive performance than the MLP, since they create multiple support vectors, which provide more flexibility in the classifying criterion. The XGBoost model is a variation of the gradient boosting model, an ensemble model that is a collection of decision trees added to minimize a loss function. The Multi-Agent classifiers are two algorithms proposed by us that consists of all the previous models to better classify the EEG data, one by a majority vote and the other by treating the sleep and wake states on a more equal footing.

D. Single-Layer Perceptron

E. Malaekah and D. Cvetkovic [8] had used the ratios of the EEG bands to perform classification of the sleep and wake states. The classifying criterion was manually tuned in their case. Instead of using the ratios of the EEG bands and manually tuning the ratio to match a specific quantity, we used an automatic learning algorithm that can find the classifying criterion without any human interference.

A Single-Layer Perceptron (SLP) is an algorithm for binary classification of an input vector. It is a linear classifier that makes its decisions by a scalar product of the input vector with a set of weights. In machine learning, a perceptron is considered to be the simplest neural network consisting of one neuron. Mathematically, the SLP algorithm can be expressed as:
$$y(\overrightarrow{x}) = \begin{cases} 1 & \text{if } \overrightarrow{w} \cdot \overrightarrow{x} + b > 0 \\ 0 & \text{otherwise} \end{cases}. \tag{6}$$
Here, $\overrightarrow{x}$ is the input vector, $\overrightarrow{w}$ is the weight vector and b is the bias. The output of the perceptron is either 0 or 1 and is determined by the condition provided by the scalar product. The bias alters the position of the decision line. The weight vector is changed by learning algorithms. The learning algorithm used here was the Particle Swarm Optimization (PSO) [14]. PSO optimizes a problem by having a set of candidate solutions (called particles), and moving these particles in the feature space by simple rules for updating the particles' positions and velocities. This optimization technique does not make assumptions about the parameters being optimized and so can be used for noisy and irregular problems. PSO possesses certain advantages of Monte Carlo techniques, where a large space of solutions is searched by the various particles, as well as the advantages of classical optimization techniques, such as gradient descent, where the particles are guided by a 'force' towards the final solution.

Each particle is attracted by its best known position, as well as the best position among all the particles. There is also an inertia term, which keeps the particle moving along a straight line. These three 'forces' act together in different proportions to guide the particle towards better solutions. A PSO algorithm consists of the following main steps, along with part of the code from our implementation in R:

- Initialize `N` particles' positions `particle_pos` and velocities `particle_vel` randomly within the boundaries of the feature space. `part` indicates the particle index, maximum `N`, and `i` the dimension, 31 dimensions in our case.
```
particle_pos[part,i] <- runif(1, min=-10,
max=10) #Random starting positions
particle_vel[part,i] <- runif(1, min=-5,
max=5) #Random starting velocities
```

- Update each particle's best position `best_particle_pos` with the randomly assigned starting position.
```
best_particle_pos[part,i] <-
particle_pos[part,i]
```

- Find the particle with the best position and update the global best position, `coeffs_aasm` or `coeffs_rnk`, with its position. This step depends on the error function and with random starting positions. In our case, the error function was the sum of absolute errors between the expected result and the actual result. This is similar to using the mean of squared errors, since only the absolute value of the error is considered. `#perr` is the particle error.
```
if(perr[part] < global_min_err) {
global_min_err <- perr[part]
coeffs_aasm <- particle_position[part,]
}
```

- Until the stopping criterion is satisfied, i.e. number of iterations run or error criterion is below threshold, do the following
  - Update each particle's position by adding the velocity to it.
```
particle_position[part,] <-
particle_position[part,] +
particle_velocity[part,]
```
  - Update each particle's best position with its current position if and only if the current position is better than the particle's best position. `#min_perr` is the particle's minimum error.
```
if (perr[part] < min_perr[part]) {
min_perr[part] <- perr[part]
best_particle_position[part,] <-
particle_position[part,]
}
```
  - Find the particle with the best current position and update the global best position if and only if this best current position is better than the global best position.
```
if(perr[part] < global_min_err) {
global_min_err <- perr[part]
coeffs_aasm <- particle_position[part,]
}
```
  - Update each particle's velocity as a sum of three contributions multiplied by factors indicating their proportions. The first contribution is the difference between the particle's best position and its current position. The second contribution is the difference between the global best position and the particle's current position. The final contribution is simply the previous velocity of the particle.
```
particle_velocity[part,] <-
particle_velocity[part,] +
runif(31,min=0,max=0.4) *
(best_particle_position[part,]
- particle_position[part,]) +
runif(31,min=0,max=0.7) * (coeffs_aasm
- particle_position[part,])
```
  - Update the number of iterations or calculate the error.
- The global best position is the best solution.

The factors which multiply the three contributions, the inertia term, the force due to the global best position and the force due to the particle's best position, are chosen by the user. These control the 'swarming rate' and hence the behavior and response of the system. In our implementation, the inertia term was multiplied by 1, the force due to the particle's best position was multiplied by a random number between 0 and 0.4 (excluding the extreme values), and the force due to the global best position was multiplied by a random number between 0 and 0.7. The inertia term was kept as a unit value because we wanted the particles to move to not stop swarming. The maximum values of 0.4 and 0.7 were chosen so that the particles were attracted towards their individual best positions and the global best positions, but the contribution of these attractions did not exceed the contribution of the inertia term. The higher average value of the attractive force towards the global best position indicated a preference of each particle to explore positions near the global best position, which is the best known solution so far. The non-zero value of the force towards the individual particle's best position ensured that the particles explored positions near their own best positions. The final values were chosen based on the "goodness" of the results.

E. Multi-Layer Perceptron

In our case, the data does not satisfy the condition of linear separability; therefore non-linear classifiers were considered.

A well-known neural network model, a Multi-Layer Perceptron (MLP) was chosen and its architecture was evaluated.

Neural networks contain at least two layers, with the possibility of additional *hidden* layers. Earlier experiments [15] were performed with two and three hidden layers, the latter of which gave maximum sleep and wake detection rates of about 69 % and 90 % respectively. Compared to the architecture with three hidden layers, this architecture drastically decreased the number of connections and thus the learning time for the neurons [15]. Based on these results, a neural network with two hidden layers 31-46-10-1 was chosen. The learning algorithm used for this project was the default algorithm, *resilient backpropagation* with weight back tracking. The error function minimized in this project was the mean squared errors.

Neural networks have one major disadvantage – risk of *overfitting*, when a model begins to describe the noise. The underlying relationship between the input variables and the output value may fail to have a predictive power completely. In our case 1,953 weights had to be found using 16,462 training points. Thus there is a high possibility of overfitting in our neural network models, which can lead to poor predictions. Unfortunately, reducing the number of neurons in the hidden layers led to a drastic decrease in the performance of the neural network, while increasing it led to overfitting, where the testing phase results were measurably poor as compared to the training phase results. Therefore, the data parameters for our neural network could not be classified correctly without increasing the risk of overfitting. Due to a lower number of sleep stages, the neural network models did better at detecting the wake stages.

F. Support Vector Machine

Support Vector Machines (SVMs) are supervised learning algorithms used for classification and regression analysis [16], [17]. An SVM model maps data points in the feature space as categories which are as far apart as possible, by constructing a hyperplane or a set of hyperplanes in a space, generally of a higher dimension than the feature space. Generally, two or more hyperplanes are selected with no points between them, and then the distances between these hyperplanes are maximized.

Two types of SVM models, a simple SVM model and a multi-class SVM model, were used. The first one interprets the hypnogram sleep stage rating as having a "meaning", i.e. this value can be expressed as the output of a mathematical function of the SVM. The second model merely assigns the hypnogram value as a label for a given SVM input vector. This model is hence more robust to clustering of data that is not linearly separable.

Before starting the training of the SVMs, tuning was performed for two parameters, `gamma` and `cost`, using the functionality provided in the package `e1071` through the function `tune.svm()` on 10 % of the data. The parameter `gamma` is internal to the implementation and is required for all the non-linear kernels, while *cost* is the cost of violating the constraints of the SVM model. The results of this tuning indicated a value of 0.1 for `gamma`, and a value of 1 for `cost`, which were recommended by the SVM implementation based on the 10 % of the data. The same values were obtained for both AASM and R&K classification methods, and were used for the final training of the models. In the multi-class variant of SVM (unlike in the default-SVM), the wake states were given a weight of 0.2 and the sleep states were given a weight of 1.0, to compensate the influence of the higher number of wake states in the dataset.

G. Extreme Gradient Boosted Model

Extreme Gradient Boosted Model (XGBoost) is a supervised machine learning algorithm used for classification and regression analysis, which comprises of an ensemble of 'weak' prediction models, generally in the form of decision trees. It is an efficient implementation of the gradient boosting framework, and has been known to be many times faster than standard implementations.

The boosting process deals with transforming the weak prediction models into a strong model. For example, decision trees with only one or two levels would be considered weak, since their accuracies are slightly better than random classification, and thus are more robust to overfitting. A boosted model creates an ensemble of these models with the intention of creating a strong learner. In many implementations, the weak learners are functions of the loss (error) functions, which the ensemble is trying to minimize. The gradient boosting model consists of calculating the loss function at each iteration and correcting itself by adding the gradient of the loss function in the next iteration [18], [19].

In this work, the maximum depth (nodes) was set to 5 to allow better decision trees at each level. This reduced the number of iterations, i.e. the number of additional trees required to correct the loss function, which was set to 40. The learning process was made more conservative (less overfitting) by scaling the contribution of each tree by 0.15, compared to the default value of 0.3. These values were selected after running multiple tests on the data, and were used for the AASM and R&K models.

H. Multi-Agent Model

To maximize the sleep and wake detection rates, two Multi-Agent (MA) models were developed. Typically, an MA model consists of many 'agents', each of which is an independent classification model. In our work, these MA models were constructed from all the aforementioned algorithms. There are different ways to implement such algorithms [20], and for this project, a democratic MA model and a weighted MA model were chosen. The democratic MA model determined the outcome of a given input by a majority vote, while the weighted MA model determined the outcome of a given input by weighing the outputs of each of the agents.

The democratic MA model did not have any training or testing phase, due to it directly relying on its component models. The outputs from the agents were directly used to

analyze this model on the training and testing data sets. Two models were made, one for the AASM classification and one for the R&K classification.

The training phase for the weighted MA model consisted of optimizing the weights using an error function, and the optimization was achieved with the Particle Swarm Optimization (PSO) as discussed in the SLP model. The error function chosen was the sum of the FNR and FPR. Since the rates are independent of the number of states, this error function is not biased towards one of the sleep or wake states. Minimizing this function would thus maximize the sleep and wake detections simultaneously (TPR = 1 - FNR and TNR = 1 - FPR).

The parameters of the PSO algorithm were identical to those used to determine the weights of the SLP, with the exceptions that the number of dimensions were reduced to 5 and the bias was not required. Two such models were made for each training set, one corresponding to the AASM classification and the other corresponding to the R&K classification.

III. Experimental Results

The data was divided randomly into training and testing sets for cross-validation purposes to check how well the model can perform under more general data. We performed 4-fold cross-validation, where the data was divided into four equal random sets. Earlier experiments [15] indicated similar results to 2-fold cross-validation results, and 4-fold cross-validation has more training and testing phases, which is why it was used to obtain more results. The final results (averaged) are presented for comparison purposes. To compare the models, all learning algorithms were provided the same data for training and testing.

Since there were two scoring models (R&K and AASM) for each data point, two models were made for each learning algorithm, e.g. MLP_Rnk and MLP_AASM. Because of the two *different* standards, it only makes sense to compare all the AASM models together and all the R&K models together. After splitting the data and dividing them as *train-sets* and *test-sets*, the data workspace in R is stored for future use and reference.

In the confusion matrix tables, two outlier points (Outlier 1 and 2) are obtained because of having the same data point classified as sleep stages other than 0 or 1 in one of the two models. These points were neglected to better analyze our results. The confusion matrix was reinterpreted for comparison purposes in terms of TPR, TNR and κ. These quantities are provided for the 4-fold cross-validation studies in table II.

A general observation is that the multi-class SVM (R&K) performs the best, followed closely by the multi-class SVM (AASM). In addition, the TPR for the multi-class SVM is about 94 %, better than previous results of 92.8 % classification accuracy obtained by I. Zhovna and I. Shallom [4], where the cross-correlation information existing between the multichannel EEG signals was used. The TNR, on the other hand, is about 75 %, slightly lower than the 77 % obtained by E. Malaekah and D. Cvetkovic [8]. This is not so far behind, and thus the multi-class SVM is a good choice for further study. As perspectives, the multi-class SVM might be improved by providing more data points corresponding to sleep stages so that the FPR is reduced.

Looking at the Cohen's indices for the different algorithms, we see that the MA, XGBoost, MLP and SVM models are at least 0.60. However, in our study, we place more emphasis on sleep detection, i.e. a TPR higher than 90 %. This is only achieved by the multi-class SVM. It should be pointed that the number of sleep states are lower than those of wake states, which is why the XGBoost and the MA models have a TNR greater than the TPR.

From the tables, we infer that the default SVM does better than the SLP and MLP methods. Both default SVM models (AASM and R&K) have a TPR of over 70 %, similar to the TNR of the multi-class SVM, and the same can be said for the TNR of the default SVM models. Thus, when compared with the results of the multi-class SVM, the TPR and TNR values appear to be swapped. This could be due to the larger number of the wake stages.

A value of $\kappa > 0.60$ indicates a good agreement between the different raters, and thus the SVM, XGBoost and MA models performed well. Comparison of the Cohen's indices shows that the default SVM, XGBoost and MA models perform well with $\kappa > 0.65$ for both scoring methods.

To better understand the trade-off between TPR and TNR, the Receiver-Operator Characteristic (ROC) curves for all the models were plotted as shown in Fig. 1. The FPR and TPR form the $x-$ and $y-$axes of the graph respectively, and a random guessing model would have TPR = FPR as shown with the blue dotted-dashed diagonal in the figure. Points above the diagonal have TPR $>$ FPR, indicating a better predictive performance and a 'perfect' classification model would be at the (0,1) point on the plot, i.e. FPR = 0 and TPR = 1.

Based on the ROC curve, we see that the weighted MA (PSO) model does the best in terms of the trade-off, being closest to the (0,1) point, followed closely by the democratic variant. This is due to the fact that the error function in the PSO algorithm treats sleep and wake stages on an equal footing. In the democratic MA model, there is no training beyond that for each of the individual agents. As a result, the bias towards the wake states exists due to the 'majority' of the models (MLP, XGB and SVM) being biased towards the wake states, a consequence of the previously mentioned larger number of wake states.

Depending on the requirement, detection of one stage might be more critical and the 'best' model has to be accordingly chosen. The purpose of this project was to focus on sleep detection over wake detection and thus, based on the Cohen's index and the requirement of TPR $>$ 90 %, the multi-class SVM is a better model.

It is important to underline that the R&K models have performed better than the AASM models. The R&K standard for sleep scoring was developed in 1968 [9], while the AASM standard was developed recently, in 2007 [10]. These standards have differences which might have affected the manual scoring of the dataset used. To choose between them, inputs from

TABLE II
COMPARISON OF RESULTS (4-FOLD CROSS-VALIDATION)

	AASM Criteria			R&K Criteria		
	TPR (%)	TNR (%)	Cohen's Index (κ)	TPR (%)	TNR (%)	Cohen's Index (κ)
SLP	72.3	76.9	0.47	73.0	77.1	0.46
MLP	69.0	88.2	0.58	70.5	89.3	0.61
Default-SVM	73.8	90.5	0.65	73.9	92.8	0.68
Multi-class SVM	94.0	67.0	0.53	94.0	74.9	0.60
XGBoost	74.5	90.6	0.66	76.2	92.2	0.69
Multi-Agent (Democratic)	77.2	87.8	0.65	79.1	90.1	0.69
Multi-Agent (PSO)	82.0	85.7	0.66	83.3	88.3	0.69

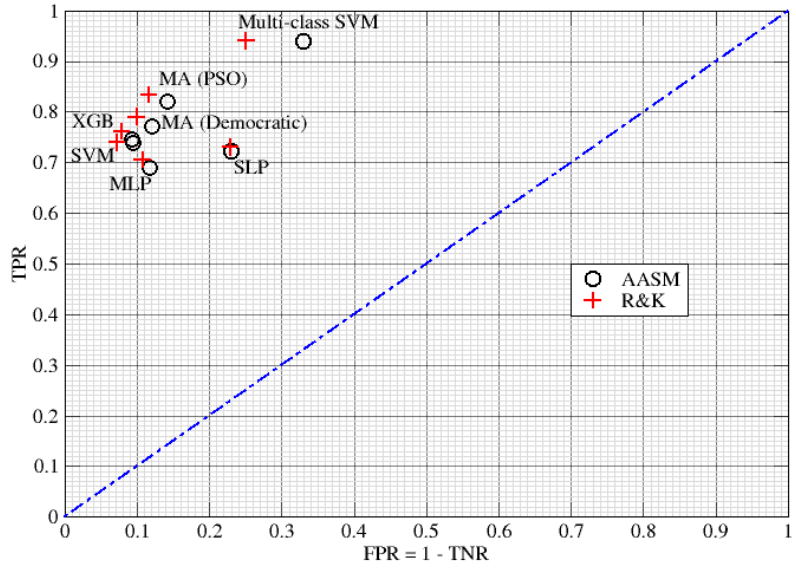


Fig. 1. ROC curve for AASM (black circles) and R&K (red crosses) classification models. The blue dotted-dashed line indicates the ROC curve for a random guess model.

a neuroscientist or more *in vivo* data based on a standard of quantifying drowsiness are required. In addition, using a more *balanced* dataset with similar numbers of wake and sleep stages might help develop the methods further.

IV. CONCLUSIONS AND PERSPECTIVES

The main objective of this work was to develop an automatic sleep detection system to warn a sleepy driver and prevent road accidents. With this purpose, the DREAMS database [5] was used and the Fourier transforms of the EEG signal formed the input vector space for the machine learning algorithms. Multiple experiments were performed on different models, such as SLP, MLP, SVM, multi-class SVM and XGBoost. A MA model was subsequently developed that treated the sleep and wake stages equally. This model was observed to be the closest to a perfect classification as can be confirmed with the ROC curve. In addition, the multi-class SVM model for the R&K scoring system maintains the benchmarks of TPR $\geq$ 90 % and TNR $\geq$ 70 %, and the AASM scoring system is not far behind. Using these benchmarks, the potential reduction in economic costs due to loss of human lives was calculated for France to be about 1,104 million EUR for the year 2014 [15].

For future work, the detection models will need to be tested across different datasets to confirm their reliability and to ensure that the models do not fail when presented with new data, either through future laboratory studies or *in vivo* experiments. Pragmatically, the models can be said to be relatively compatible with different datasets, a desirable property, which is useful in the event of retraining the models for certain medical cases where the EEG signals might not be the same as those in an average human being.

Furthermore, it might be fruitful to validate our results from a strictly medical perspective, with the ability to quantify *drowsiness* rather than the first sleep stages, preferably using EEG signals. Obtaining a more *balanced* dataset with equal wake and sleep stages would ensure that there is no bias of the models towards a particular stage. Including more physiological data, like body pulse or blood pressure, may improve the model, since they also decrease when a person tends to fall asleep.

REFERENCES

[1] National Highway Traffic Safety Administration. Drowsy Driving and Automobile Crashes. http://www.nhtsa.gov/people/injury/drowsy_driving1/Drowsy.html (Accessed 08 May, 2014).

[2] 1.9 Million Drivers Have Fatigue-Related Car Crashes or Near Misses Each Year. National Sleep Foundation (2009). http://www.sleepfoundation.org/media-center/press-release/19-million-drivers-have-fatigue-related-car-crashes-or-near-misses-each (Accessed 08 May, 2014).

[3] Crashes Where Fatigue Was a Contributing Factor. National Sleep Foundation (2012). http://sleepfoundation.org/sites/default/files/Crashes%20Fatigue%20a%20Factor.pdf (Accessed 20 May, 2015).

[4] Zhovna I. and Shallom I. D.: Automatic detection and classification of sleep stages by multichannel EEG signal modelling. Engineering in Medicine and Biology Society, 2008. 30th Annual International Conference of the IEEE, 2665-2668 (2008).

[5] Devuyst S., Dutoit T., Kerkhofs M.: DREAMS Project, The DREAMS Sleep Subjects Database. http://www.tcts.fpms.ac.be/~devuyst/Databases/DatabaseSubjects/ (Accessed 16 April, 2014).

[6] KDnuggets Polls, Data Mining Methodology (2007). http://www.kdnuggets.com/polls/2007/data_mining_methodology.htm (Accessed 20 May, 2015).

[7] Van Hese P., Philips W., De Koninck J., Van de Walle R., Lemahieu I.: Automatic detection of sleep stages using the EEG. Engineering in Medicine and Biology Society. Proceedings of the 23rd Annual International Conference of the IEEE, 2, 1944-1947 (2001).

[8] Malaekah E. and Cvetkovic D.: Automatic detection of the wake and stage 1 sleep stages using the EEG sub-epoch approach. Engineering in Medicine and Biology Society (EMBC) 2013, 35th Annual International Conference of the IEEE, 6401-6404 (2013).

[9] Rechtschaffen A., Kales A.: A Manual of Standardized Terminology, Techniques, and Scoring System for Sleep Stages of Human Subjects. US Department of Health, Education, and Welfare Public Health Service (1968).

[10] Iber C., Ancoli-Israel S., Chesson Jr. A. L., Quan S. F.: The AASM Manual for the Scoring of Sleep and Associated Events. American Academy of Sleep Medicine (2007).

[11] Sucholeiki R.: Normal EEG Waveforms (2014). http://emedicine.medscape.com/article/1139332-overview#aw2aab6b3 (Accessed 14 April, 2014).

[12] Lalkhen A. G. and McCluskey A.: Clinical tests: sensitivity and specificity. Continuing Education in Anaesthesia, Critical Care & Pain, 8, 221-223 (2008).

[13] Cohen J.: A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement 20, 37-46 (1960).

[14] Kennedy J. and Eberhart R.: Particle swarm optimization. Proceedings of IEEE International Conference on Neural Networks, 4, 1942 (1995).

[15] Pasieczna A. H.: An Approach to Driver Sleep Detection. Master Thesis Report, Wrocław University of Economics (2015).

[16] Boser B. E., Guyon I. M., Vapnik V.: A training algorithm for optimal margin classifiers. Proceedings of the fifth annual workshop on Computational learning theory, 144-152 (1992).

[17] Cortes C., Vapnik V.: Support-Vector Networks. Machine Learning 20, 273-297 (1995).

[18] Mason L., Baxter J., Bartlett P. L., Frean M. R.: Boosting Algorithms as Gradient Descent. Advances in Neural Information Processing Systems 12, 512-518, *MIT Press* (2000).

[19] Friedman J. H.: Greedy Function Approximation: A Gradient Boosting Model, IMS 1999 Reitz Lecture (1999).

[20] Dietterich T. G.: Ensemble Methods in Machine Learning. Lecture Notes in Computer Science 1857, 1-15 (2000).

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 121–128
DOI: 10.15439/2016F140
ACSIS, Vol. 8. ISSN 2300-5963

Analysis of the Changes in Processes Using the Kosinski's Fuzzy Numbers

Piotr Prokopowicz
Kazimierz Wielki University
in Bydgoszcz
ul. Chodkiewicza 30, 85-064 Bydgoszcz, Poland
Email: piotrekp@ukw.edu.pl

***Abstract*—This paper presents the analysis of potential of the Kosinski's Fuzzy Number (KFN) idea in the modeling trends of the processes which are described imprecisely. KFNs conception is an alternative for the classical fuzzy numbers ideas as model to represent of the imprecise quantitative data. They introduces new feature into vagueness of the information - a direction. It is base for good arithmetical properties of calculations. Furthermore, a direction also extends a potential in the modeling of information by the additional interpretation, what is a subject of this article. This new potential is presented and explained basing on the example of modeling of quantity of liquid in a reservoir. The environment is changing dynamically what is described as the changes in inflow and outflow. Proposed example explains how to interpret the direction of KFN and how to understand the results of calculations and its influence.**

***Index Terms*—Ordered Fuzzy Numbers, Kosinski's Fuzzy Numbers, direction of imprecision, trend in the data, interpretation of fuzziness**

I. Introduction

THE CLASSIC idea of fuzzy sets and numbers is widely known methodology for modeling the imprecision of real world. It wins with statistical methods, especially in the problems where partial imprecision is needed.

Unfortunately the computational properties of classical fuzzy numbers (convex fuzzy numbers) have drawbacks connected with a rapidly growing imprecision after sequence of operations. To improve their arithmetics several additional solutions were introduced. They are usually connected with defining additional operations or constraints ([1][2][3]).

An alternative solution is the Kosinski's Fuzzy Numbers mathematical model [4], [5]. This model takes into account the order of the characteristic parts of a fuzzy number giving the fuzzy number an additional feature - direction. By the consideration of the order in calculations we get the opportunity to reduce the imprecision of the following operations. The KFN computational model have a number of properties which were presented in the publications [4], [6], [7].

This model was previously called 'Ordered Fuzzy Numbers'. However, to honor the late Professor Witold Kosiński, his work, the contribution and commitment to the development, analysis and popularization of this model since the 2015 [8] the name "Kosinski's Fuzzy Numbers" (KFN for short) is used instead.

New property - a direction - apart good properties in calculations also introduces new interpretative potential. We have additional information in processing imprecise data associated with the direction. The papers [9], [10] presents a practical use of KFNs in the modeling of financial data, and [11] applies them to the modeling of diversity of opinions in the social networks. Furthermore, in [12] is presented proposition of the application of KFNs for the ant colony optimization algorithm.

A direction is property which can contain an additional portion of information in the same object which also represents an imprecise fuzzy value. However, the interpretation of this new feature, should be consistent and intuitive with methods of the calculations/processing and also with the final results of them. Such coherent proposal how to understand a direction of KFN was presented in [13], [14]. This proposition connect this new feature with a representation of trend in a process. Despite the fact that in the literature can be found various proposals of applications of the model KFN, it lacks exact discuss and full consistent example how and where to get the direction and how it translates into results of operations. Thus the purpose of this publication is to provide such complete analysis of the quite simple and easy to understand process where actions are described by KFNs.

It will be presented on the example of process of changes in the level of liquid in the reservoir. Example presents a problem, in which we want to describe / determine the state of the liquid in the container, under conditions of high variability due to rapidly changing an inflow and an outflow. This publication demonstrates the effectiveness of KFNs in the modeling such problems. The example will show that by modeling the process with the KFNs, we can get more than imprecise description of situation, we can get also the informations about the trends of changes.

A. The Organization of the Paper

In next section the Kosinski's Fuzzy Number model is shortly introduced with the genesis of the idea. In subsections are also presented the definitions of arithmetic operations on KFNs and practical interpretation of the specific feature of the model - the direction. The following - third section contains the description of the process, which is modeled in this paper. First subsection presents assumptions of the model, next the

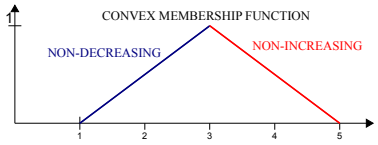


Fig. 1. Parts of the convex membership function

description of subsequent actions with use of KFN model. Third subsection presents the analysis of the appropriateness of using KFNs for modeling of actions of dynamic process. Next - fourth section provides some discussion about different ways for concerning of the direction in the model of process. Final section provides extended summary and conclusions for this paper.

II. Kosinski's Fuzzy Number (KFN)

Main concepts of the idea of Kosinski's Fuzzy Numbers were introduced and developed in the series of papers [15], [4], [5], [6], [13], [16], [14]. As it is an unusual model of imprecise information, before introducing a formal definition, it will be useful to clarify a background of concept.

At the beginning, very important fact must be emphasized. The KFNs are not the fuzzy sets. They are connected conceptually with the idea of fuzzy sets and they can be used in similar way, but, in formal, they are not Zadeh's fuzzy sets. Thus comparing the operations on KFNs with adequate operations based on Zadeh's extension principle [17] can be realized only in the context of the results of calculations, but not by the definitions.

It is also worth noting the fact, although KFNs by definition are other mathematical objects than fuzzy sets, they can represent the majority of situations, which can be described by the convex fuzzy numbers including calculations on them.

A. Genesis of the KFNs

Idea of KFNs have a source in the quasi-concavity of membership functions of fuzzy numbers [18]. It is not a new observation that the each convex membership function of fuzzy number can be split into two parts: first is non-decreasing and second is non-increasing (see fig.1). It is a base of widely known classical fuzzy numbers model called the (L, R) fuzzy numbers [19]. The KFNs idea also generally is based on such point of view. However, the new model treat separately non-decreasing and non-increasing parts of fuzzy number. Additionally, it defines an order between these parts as independent from the domain values. Such conditions leads to the new possibilities in calculations and also in processing of imprecise data.

B. Definition of the KFN

Following the papers [15], [4], [5], [6], [13], [20], [8], [21] fuzzy number will be identified with the pair of functions defined on the interval $[0, 1]$.

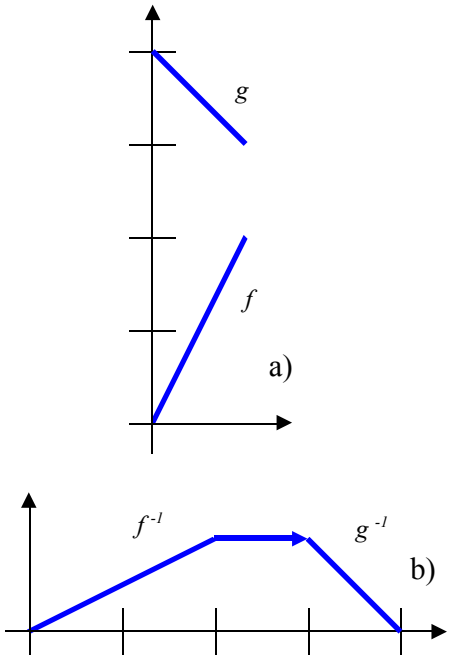


Fig. 2. a)Kosinski's Fuzzy Number, b)The Kosinski's Fuzzy Number as convex fuzzy number with an arrow.

Definition 1: The Kosinski's Fuzzy Number (KFN in short) A is an ordered pair of two continuous functions

$$A = (f_A, g_A) \tag{1}$$

called the up-part and the down-part, respectively. Both are defined on the closed interval $[0, 1]$ with values in **R**.

If the functions f and g are monotonic (Fig.2a), they are also invertible and possess the corresponding inverse functions defined on the real axis with the values in interval $[0, 1]$. If these two inversed functions are not connected, we linking them with constant function with the value 1. In such way we receive an object which directly represents the classical fuzzy number. For the finalization of transformation, we need to mark an order of f and g with an arrow on the graph (see Fig.2b). Notice that pairs (f, g) and (g, f) are the two different Kosinski's Fuzzy Numbers, unless $f = g$. They differ by their orientation or direction. We can distinguish two orientations giving them names. If the down-part g is greater than the up-part f we will call that "positive" orientation and opposite to it - "negative".

For the later use it will be more convenient to adopt the following general indications of the KFN boundaries:

$$\begin{aligned} UP &= (s, 1^-) \\ CONST &= [1^-, 1^+] \\ DOWN &= (1^+, e) \end{aligned} \tag{2}$$

These boundaries allow for simply representation of the KFNs where f and g (up-part and down-part) are linear functions. In such situation we can precisely represent a given KFN by four $(s, 1^-, 1^+, e)$. In fact in this paper in the example presented in

later sections the $CONST$ interval will be minimized to the one point $k = 1^- = 1^+$, thus the KFNs will be represented by triples (s, k, e) (see fig.3).

It should be emphasized that, these intervals can be improper intervals (see the KFNs B and C on the fig.3) in the sense of Kaucher's extended interval arithmetic [22] and called by him 'directed intervals', i.e. such $[a, b]$ where a may be greater than b.

For monotonous f and g we may point the membership function within the meaning of classical fuzzy numbers:

$$\mu(x) = \begin{cases} f^{-1}(x), \text{ if } x \in [f(0), f(1)] = [s, 1^-], \\ g^{-1}(x), \text{ if } x \in [g(1), g(0)] = [1^+, e], \\ 1 \text{ when } x \in [1^-, 1^+]. \end{cases} \quad (3)$$

It is worth to point out that a class of Kosinski's Fuzzy Numbers represents the very wide class of convex fuzzy numbers with continuous membership functions (regarding "classical fuzzy numbers" [18], [23], [3], [24]).

C. Arithmetic Operations

Operations on KFNs we define as calculations with the up-parts and down-parts as follows:

Definition 2: Let $A = (f_A, g_A), B = (f_B, g_B)$ and $C = (f_C, g_C)$ are mathematical objects called Kosinski's Fuzzy Numbers. The sum $C = A + B$, subtraction $C = A - B$, product $C = A \cdot B$, and division $C = A \div B$ are defined by the formula

$$f_C(y) = f_A(y) \star f_B(y) \qquad \wedge \qquad g_C(y) = g_A(y) \star g_B(y) \quad (4)$$

"$\star$" replaces "$+$", "$-$", "$\cdot$", and "$/$". The A/B is determined only if B does not contain zero. The $y \in [0, 1]$ is the domain of functions f and g.

The properties of these operations and their results have already been discussed and analyzed in various publications (for example see [25], [6], [7]). Additionally the paper [7] presents also a different examples of calculations. Nevertheless a brief description of the properties KFN calculations is needed for a better understanding of analysis included in the next sections of this paper. For example it is important property, that the subtraction is equal to the addition of the opposite number, where the opposite number is obtained by multiplying the given value by the -1 (real number - singleton). By using the above-mentioned method in calculation of $A - A$ we obtain exact zero (crisp number). Using the arithmetical operations on KFNs every simple equation type $A + X = B$, where A and B are fuzzy numbers with any membership functions, can be solved. We calculate result exactly the same way as with real numbers. Such possibilities are a consequence of adopting operations for KFNs directly from real numbers. After a closer investigation of the definitions 1 and 2 it can be noted that the operations on parts of the KFNs are executed through operations on functions representing these parts. Finally, the operations on the functions are, in fact, operations on their values. Thus, if space of values of function

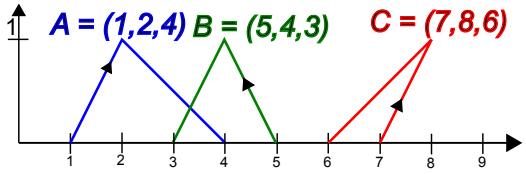


Fig. 3. Examples of KFNs represented by triples.

are the real numbers (as with all KFNs), then practically, calculations are executed as operations on the real numbers universe. Summarizing, the KFN model grants flexibility of calculating on imprecise data in a similar way like with real numbers on crisp data. It retains fuzzy quantitative character, but without necessity of growing an imprecision.

However, using the KFNs we should remember that their basics are different than the Zadeh's fuzzy sets and some objects like C on the fig.3 can appear (see also [14], [7]). Such elements are consistent with the definition of KFN (see def.1), although, their shape can not be defined as a membership function. Such objects are called improper KFNs. This aspect of model was commented in [4], [5], [7]. Despite the unusual shape (as for fuzzy numbers), such KFNs still contain important information needed for the calculations. In the example presented in next sections of this paper such object also will be part of analysis as an important portion of data about the modeled process.

D. Imprecision interpretation and direction of the KFN

The direction is a key element of the KFN model. Basically it is defined as an order of the parts of fuzzy number and is independent from the real numbers universe. Proposition of practical interpretation for this new property was presented in [13], [14]. KFNs are considered in these papers as the values representing an observation, which passes in time, regardless of the order of numerical values. So, the time dependence can be a natural interpretation of the direction. Such context is also used in this publication. Following [13], [14]imprecision here is interpreted as a consequence of dynamic changes. So the up-part represents, relatively short, past behavior of the value represented by given KFN, and the down-part indicate expected change of the value in the next step of process. However, it should be clarified that this is not the only possible interpretation of the direction, but this publication focuses just on such variant.

III. Modeling process using the KFNs

At start it should be noted, that there is a paper [26] that presents an example of calculations which context is connected with the water lever for the cofferdam. But that publication focuses on comparison of arithmetics of LR fuzzy numbers and KFNs based on solving simple equation. The issue of the direction's source is there minimized.

Possibilities of modeling an imprecision in the processes with use of the KFNs will be presented here on the simple intuitive example of reservoir with one outflow and one inflow.

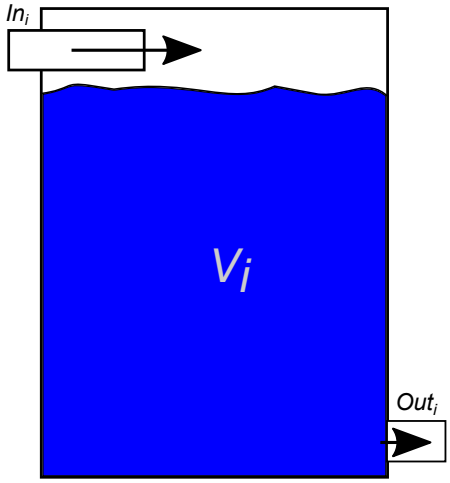


Fig. 4. General idea of the model.

KFNs are used here to describe a behavior of the inflow and the outflow, and also, to indicate the amount of liquid in the reservoir during each step of the process. Following [13], [14], here, the source of imprecision will be also connected strictly with a dynamic changes of parameters. Thus, when no changes are applied the situation will be considered as crisp and precise.

It is worth noting that the problem that we want to solve with this example is to find a simple method to determine the current state of the liquid in the tank in a high variability situation. The input data, in addition to the initial state, are the changes in inflow and outflow. Generally, the problem can be formulated as monitoring the status of the tank only on the basis of the monitoring of sources (an outflow is interpreted as a negative value).

A. Assumptions

General idea of the model is presented on the fig. 4. Before the analysis, some assumptions should be clarified. The time of analyzed example is discrete and divided onto intervals with a similar duration. The units are not specified, as it is not relevant whether they will be seconds, minutes or hours. The actions during the process will be described as changes of values (of inflow and outflow) which lasts by one time interval denoted by Δt. Similarly, the determination of the state of actual liquid level in the reservoir will be denoted by V_i without specifying units. As the natural consequence of these assumptions the performance of inflow In_i and outflow Out_i will be described as cumulative change ΔV_i during Δt with units adequate to the description of capacity of analyzed reservoir and time-steps. The index i will indicate the subsequent time-steps in the analysis.

B. The sequence of actions in the process

The exemplary of process starts with static situation. Inflow and outflow is zero. At the beginning, volume of liquid in the reservoir is described by the KFN singleton $V_0 = (20, 20, 20)$. The whole process will be divided onto time-steps numerated from 1 to 11. Each step is described by the changes in inflow and outflow which are usually consequent continuations of previous states. For each step, the actual volume of liquid will be calculated by the formula:

$$V_i = V_{i-1} + In_i + Out_i \tag{5}$$

where i - indicates the number of subsequent - actually analyzed - time interval. In general, the inflow values are positive numbers as they represent incoming liquid to the reservoir. So, for the outflow the values are negative.

Steps in the process are as follows:

1) The inflow starts slowly from 0 to 1. It is described by the KFN $In_1 = (0, 0, 1)$. The outflow is no changing so the value is $Out_1 = (0, 0, 0)$. The state of liquid in the reservoir is calculated $V_1 = V_{i-1} + In_i + Out_i = (20, 20, 21)$. As we can see the fuzziness is growing due to the changes in inflow. The result represents not only the actual state but also an information about a trend in the process (which is presently growing due to positive orientation of V_1).
2) The inflow is growing with the same speed $In_2 = (0, 1, 2)$. It means is increasing from 1 to 2 unit of volume in Δt. The outflow still is closed: $Out_2 = (0, 0, 0)$. The level of liquid: $V_2 = V_{i-1} + In_i + Out_i = (20, 21, 23)$. We can see that we again have reasonable imprecise information about actual volume of liquid. Summarizing, it started with 20 then in first time interval the inflow was about 0 however not precisely. During next time interval, the inflow is growing and is about 1. So after these two time intervals we have in result the content of reservoir about 21. The orientation is positive so this is increasing trend and as we can see, generally, it is the right conclusion for this moment.
3) The inflow is growing with the same speed $In_3 = (1, 2, 3)$. Now the outflow is activated and it is rapid: $Out_3 = (0, 0, -3)$. At this point it is worth to calculate the cumulative change: $\Delta V_3 = In_3 + Out_3 = (1, 2, 0)$. The level of liquid: $V_3 = (21, 23, 23)$. In this step we see, that the starting rapid outflow eliminates the effect of inflow. The KFN V_3 tells us that the growing trend in filling reservoir is stopping. More interesting however, is the cumulative change ΔV_3, which is in fact the improper KFN as it was earlier introduced. As we can see such object represents important information - sudden change of a trend.
4) The inflow is growing with the same speed: $In_4 = (2, 3, 4)$. The outflow is growing with the same high speed: $Out_4 = (0, -3, -6)$. The cumulative change: $\Delta V_4 = In_4 + Out_4 = (2, 0, -2)$. The level of liquid: $V_4 = (23, 23, 21)$. In this step we have a situation when the actual volume of content of reservoir changes orientation. Inflow and outflow are more or less balancing, however trend starts decreasing.
5) The inflow stops increasing at the level 4: $In_5 = (3, 4, 4)$. The outflow still is growing, however, much

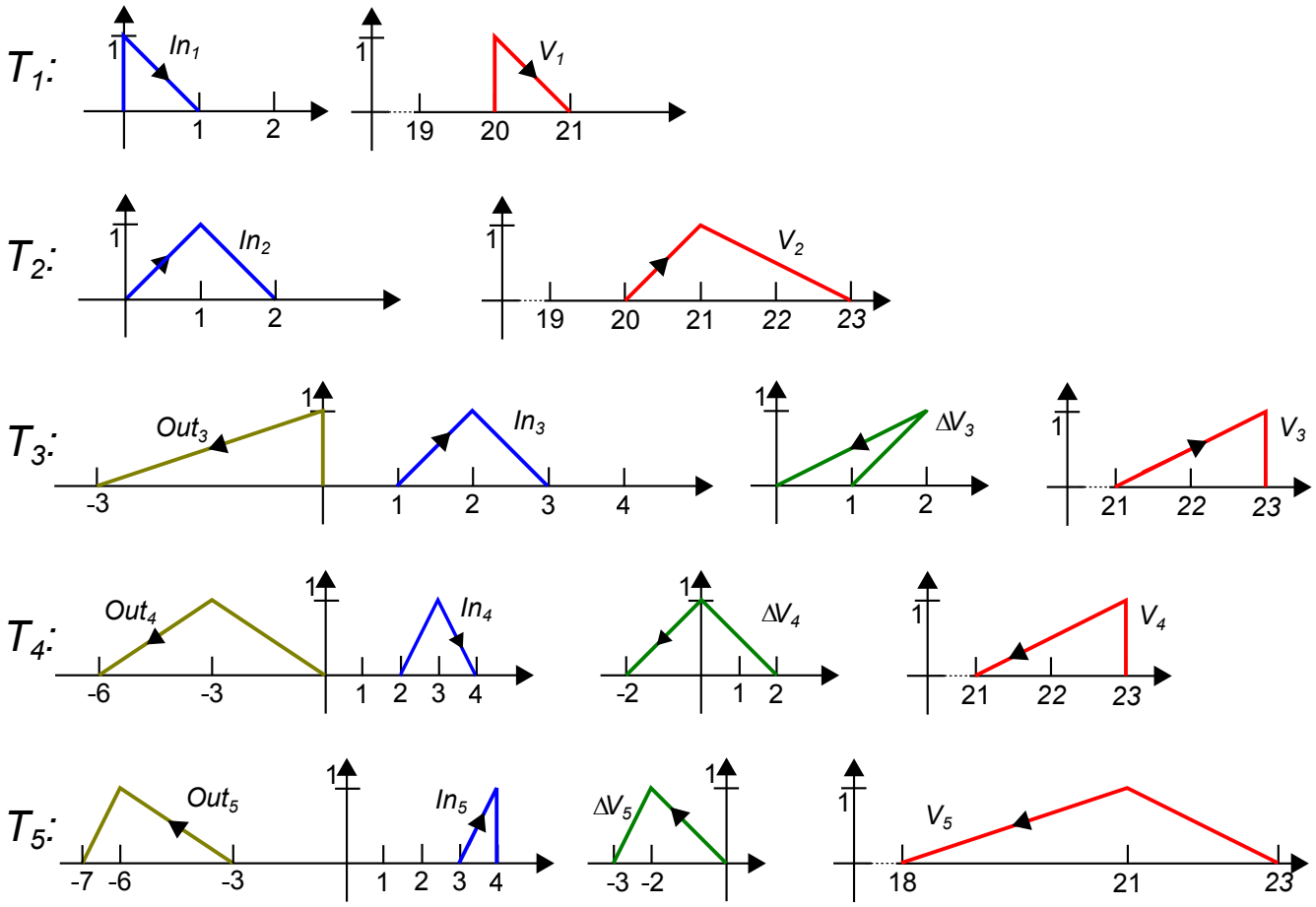


Fig. 5. KFNs in first five time intervals of the process.

more slowly than before: $Out_5 = (-3, -6, -7)$. The cumulative change: $\Delta V_5 = (0, -2, -3)$. The level of liquid: $V_5 = (23, 21, 18)$. Clearly as the V_5 posses negative orientation the liquid volume is about 21 in the decreasing trend.

6) The inflow is constant and equal 4: $In_6 = (4, 4, 4)$. The outflow stops growing at the value -7: $Out_6 = (-6, -7, -7)$. The cumulative change: $\Delta V_6 = (-2, -3, -3)$. The actual level of liquid: $V_6 = (21, 18, 15)$. At this moment we have a situation where the outflow and the inflow stabilizing. So, fuzziness of cumulative change is not high, however, the imprecision of actual volume of liquid is clearly greater. Although the cumulative changes are not too large, the situation in the reservoir is quite dynamic. In comparison with the initial value of 20, currently we have an income of liquid on the level 4 and outcome - about 6. This is generally quite big change, even if cumulatively it is not so significant.
7) The inflow starts decreasing with high dynamic: $In_7 = (4, 4, 2)$. The outflow is stabilized at -7: $Out_7 = (-7, -7, -7)$. The cumulative change: $\Delta V_7 = (-3, -3, -5)$. The level of liquid $V_7 = (18, 15, 10)$. Now we have clear decreasing trend and this decreasing is growing.
8) The inflow decreasing with the same speed: $In_8 = (4, 2, 0)$. The outflow starts rapidly decreasing: $Out_8 = (-7, -7, -4)$. The cumulative change: $\Delta V_8 = (-3, -5, -4)$. The level of liquid: $V_8 = (15, 10, 6)$. Once again we have to deal with the improper KFN as cumulative change.
9) The inflow is closing: $In_9 = (2, 0, 0)$. The outflow is decreasing with the same very high speed: $Out_9 = (-7, -4, -1)$. The cumulative change: $\Delta V_9 = (-5, -4, -1)$. The level of liquid: $V_9 = (10, 6, 5)$.
10) The inflow is cut off: $In_{10} = (0, 0, 0)$. The outflow is closing: $Out_{10} = (-4, -1, 0)$. The level of liquid: $V_{10} = (6, 5, 5)$. This KFN shows that situation in reservoir is stabilizing. The whole process stops, so a fuzziness/imprecision is low.
11) Finally, the inflow is cut off and also the outflow definitely ends: $Out_{11} = (-1, 0, 0)$. The final level of liquid: $V_{11} = (5, 5, 5)$. There is no change in the reservoir. Therefore, according to previous assumptions (a source of imprecision are changes), the result is not fuzzy.

C. The KFNs in the example

For a better understanding and analysis of the example the full graphical representation of KFNs which describes clue values for first five steps of the process are presented on fig.5. Summing up, we can identify the following general activities during the process:

- First, the inflow is slowly turning on and it is increasing steadily giving more and more income of the liquid during next few time intervals. After that it stabilizes for a moment.
- During slow increasing of inflow, the rapid outflow starts, which quickly slow down and stabilizes in short time.
- Then, the inflow is slowly closing and then, almost simultaneously, also the outflow is closing.

The KFN model can be represented, quite intuitively, by Japanese Candlestick Chart commonly used for financial data [10]. We can use it to present a set of KFNs which have a common context. This kind of chart is especially compatible with the KFNs because it can represent not only values, but also intervals and more important - a direction. It is not full precise description but gives general outlook about main character of such set. Figure 6 represents the chart for volume of liquid V_i in the reservoir during each time interval. The actual level of content during $i-th$ time interval is represented by the rectangles. Height of the rectangle represents the range of imprecision. Additionally, white color represents the increasing trend and black - the decreasing. So, we can read from the chart informations about the content of reservoir during subsequent time-steps as follows:

- The trend is positive in the first three time intervals. The fuzziness is not large as the changes are not so dynamic.
- From the fourth time-step the trend changes and becomes negative. Additionally the fuzziness is growing. If we refer back to the detailed description of the example, in third step the outflow starts rapidly and lasts for a while. So, its consequence is the change of trend in fourth step and growing imprecision during next few periods.
- From the ninth time interval, imprecision is decreasing, however trend of the liquid's volume is still negative. If we again refer back to the detailed description of the process, we notice that from ninth period the dynamics of inflow and outflow is decreasing so the imprecision is smaller.

IV. Determining direction of the KFNs

The source of a direction of KFNs used in the example is basing on the fundamentals of conception 'a change', which has generally two possible trends. It is either increasing or decreasing - thus positive or negative orientation for the KFNs. It is simple and intuitive.

In the linguistic description of actions in the example from previous section, we consider the inversion of ideas: the inflow and the outflow. Therefore, if the inflow means adding liquid to the reservoir, then enlarging an outflow means that, the KFN becoming more and more negative. The KFN model grants

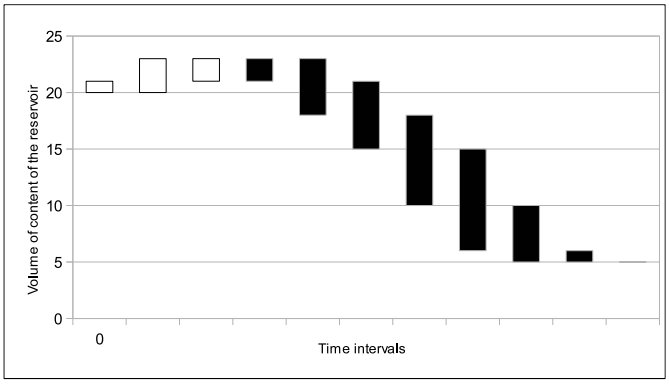


Fig. 6. Candlestick chart for the level of liquid in the subsequent time-steps.

flexibility of calculations, so we could describe outflow with absolute values. However, in such situation the opposition of outflow against inflow should be referred in the calculating of actual volume of content. Instead the formula 5 we should use:

$$V_i = V_{i-1} + In_i - Out_i. \tag{6}$$

Finally, the results would not change, due to the fact that subtracting the KFNs is, in real, adding the opposite number. We get the opposite number like in the real numbers (but unlike with classical fuzzy numbers) by multiplying the original number by -1.

V. Summary and Conclusions

The example in the previous section presents application of KFNs for the modeling of changes in the process of filling and outpouring content in the reservoir. Although this example is simply, it presents not only the changes of actual volume of liquid. In fact, the fuzziness of inflow and outflow refers to the 'change of changes'. As the In_i and Out_i represents $\Delta V_i/\Delta t$, so fuzziness of them is a change $\Delta(\Delta V_i)$ during Δt. It may be easier to understand if we compare this with the process of object's movement. If x is a position, then Δx is a velocity. And furthermore, change of speed is an acceleration. It generally shows a good potential of the KFNs as a good tool for modeling more complex dynamic action than just simple changes.

Presented process is analyzed step by step. It shows also particular usefulness of KFN model for a linguistic describing of changes. Although in the example we have one inflow and outflow, it is not difficult to expand conception to the multi-inflow/outflow situations. The formula 5 becomes as follows:

$$V_i = V_{i-1} + \Sigma In_i^j + \Sigma Out_i^k \tag{7}$$

where: j - subsequent number of the inflow, k - subsequent number of the outflow.

However, because the indication of source of changes, if it is either inflow or inflow, bases on the numerical values, we can simplify the notation. Instead In^j and Out^k we can use just Src:

$$V_i = V_{i-1} + \Sigma Src_i^m \tag{8}$$

where: $m \in (j + k)$ - subsequent number of the source of change of any kind.

If we analyze concrete values of actual level of liquid V_i in each time-step (let's look more closely at the fig.5) we can notice that KFNs accumulates the fuzziness of changes. However, if changes are opposite - they are canceling. This is right and consistent with the intuition. If the liquid is inflowing and simultaneously the same amount of content is outflowing, there is no change at all.

Furthermore we can see also on fig. 5, that KFNs representing actual volume of content can be referred to the classical fuzzy numbers. Such objects can be defuzzified or we can ignore a direction and use them, for example, as input values for the next stage of processing, even with use of classical fuzzy system. But these KFNs contains more information than only imprecision. We can read from them, the actual trend of the process. The papers [16], [27], [8], [21] describes the research of the methods of processing of KFNs where a direction is considered - Direction Sensitive Fuzzy Information Processing. So, instead ignoring orientation we can use full information and process it in a more effective way.

By application of KFNs we can distinguish a situation "about 5 in increasing trend" and "about 5 in decreasing trend". It is not hard to find many real life circumstances where these informations should generate significantly different reactions. An example can be the getting dressed for a trip. If we have "15 degrees of Celsius and temperature is decreasing" we should get more warm cloth than in opposite trend. More to that, there is also significance either a dynamic of changes is large or small.

The example presented in this paper can be easy transferred into other areas than the controlling content of reservoir. Balancing of the incomes and outcomes is very popular pattern. If it is inflow of liquid or income of money, or increasing number of some web-portal users, it can be described by the KFNs. So, if we need a tool for processing an imprecise data and we want to model a trend, the KFNs can be used.

ON closing, it should be stressed that, this publication presents the effectiveness of the use of KFNs in the situations where we focus on the dynamic changes as the source of the inaccuracy. However, there is many cases where a vagueness is more complex. In such situations a good idea is to design the hybrid methods which will consider all significant sources of an imprecision.

REFERENCES

[1] E. Sanchez, "Solution of fuzzy equations with extended operations," *Fuzzy Sets and Systems*, vol. 12, no. 3, pp. 237 – 248, 1984. [Online]. Available: http://dx.doi.org/10.1016/0165-0114(84)90071-X

[2] G. J. Klir, "Fuzzy arithmetic with requisite constraints," *Fuzzy Sets and Systems*, vol. 91, no. 2, pp. 165 – 175, 1997, fuzzy Arithmetic. [Online]. Available: http://dx.doi.org/10.1016/S0165-0114(97)00138-3

[3] M. Wagenknecht, R. Hampel, and V. Schneider, "Computational aspects of fuzzy arithmetics based on archimedean t-norms," *Fuzzy Sets and Systems*, vol. 123, no. 1, pp. 49 – 62, 2001. [Online]. Available: http://dx.doi.org/10.1016/S0165-0114(00)00096-8

[4] W. Kosiński, P. Prokopowicz, and D. Ślęzak, "Ordered fuzzy numbers," *Biulletin of the Polish Academy of Sciences Mathematics*, vol. 51, no. 3, pp. 327 – 338, 2003.

[5] W. Kosiński, P. Prokopowicz, and D. Ślęzak, *Intelligent Information Processing and Web Mining: Proceedings of the International IIS: IIPWM'03 Conference held in Zakopane, Poland, June 2–5, 2003*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2003, ch. On Algebraic Operations on Fuzzy Numbers, pp. 353–362. [Online]. Available: http://dx.doi.org/10.1007/978-3-540-36562-4_37

[6] ——, "Calculus with fuzzy numbers," in *Intelligent Media Technology for Communicative Intelligence*, ser. Lecture Notes in Computer Science, L. Bolc, Z. Michalewicz, and T. Nishida, Eds. Springer Berlin Heidelberg, 2005, vol. 3490, pp. 21–28. [Online]. Available: http://dx.doi.org/10.1007/11558637_3

[7] P. Prokopowicz, "Flexible and simple methods of calculations on fuzzy numbers with the ordered fuzzy numbers model," in *Artificial Intelligence and Soft Computing*, ser. Lecture Notes in Computer Science, L. Rutkowski, M. Korytkowski, R. Scherer, R. Tadeusiewicz, L. Zadeh, and J. Zurada, Eds. Springer Berlin Heidelberg, 2013, vol. 7894, pp. 365–375. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-38658-9_33

[8] P. Prokopowicz and W. Pedrycz, "The directed compatibility between ordered fuzzy numbers - a base tool for a direction sensitive fuzzy information processing," in *Artificial Intelligence and Soft Computing*, ser. Lecture Notes in Computer Science, L. Rutkowski, M. Korytkowski, R. Scherer, R. Tadeusiewicz, L. A. Zadeh, and J. M. Zurada, Eds. Springer International Publishing, 2015, vol. 9119, pp. 249–259. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-19324-3_23

[9] A. Marszałek and T. Burczyński, "Modelling financial high frequency data using ordered fuzzy numbers," in *Artificial Intelligence and Soft Computing*, ser. Lecture Notes in Computer Science, L. Rutkowski, M. Korytkowski, R. Scherer, R. Tadeusiewicz, L. Zadeh, and J. Zurada, Eds. Springer Berlin Heidelberg, 2013, vol. 7894, pp. 345–352. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-38658-9_31

[10] D. Kacprzak, W. Kosiński, and K. W. Kosiński, *Artificial Intelligence and Soft Computing: 12th International Conference, ICAISC 2013, Zakopane, Poland, June 9-13, 2013, Proceedings, Part I*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, ch. Financial Stock Data and Ordered Fuzzy Numbers, pp. 259–270. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-38658-9_24

[11] M. Kacprzak, W. Kosiński, and K. Węgrzyn-Wolska, *Artificial Intelligence and Soft Computing: 12th International Conference, ICAISC 2013, Zakopane, Poland, June 9-13, 2013, Proceedings, Part I*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, ch. Diversity of Opinion Evaluated by Ordered Fuzzy Numbers, pp. 271–281. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-38658-9_25

[12] J. M. Czerniak, Ł. Apiecionek, and H. Zarzycki, *Beyond Databases, Architectures, and Structures: 10th International Conference, BDAS 2014, Ustron, Poland, May 27-30, 2014. Proceedings*. Cham: Springer International Publishing, 2014, ch. Application of Ordered Fuzzy Numbers in a New OFNAnt Algorithm Based on Ant Colony Optimization, pp. 259–270. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-06932-6_25

[13] W. Kosinski and P. Prokopowicz, "Fuzziness - representation of dynamic changes?" in *New Dimensions in Fuzzy Logic and Related Technologies, Vol. I, Proceedings,* M. Stepnicka, V. Novak, and U. Bodenhofer, Eds. European Soc Fuzzy Log & Technol, 2007, pp. 449–456, 5th Conference of the European-Society-for-Fuzzy-Logicand-Technology, Ostrava, Czech Republic, Sep. 11–14, 2007.

[14] W. Kosiński, P. Prokopowicz, and D. Kacprzak, *Views on Fuzzy Sets and Systems from Different Perspectives: Philosophy and Logic, Criticisms and Applications*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, ch. Fuzziness – Representation of Dynamic Changes by Ordered Fuzzy Numbers, pp. 485–508. [Online]. Available: http://dx.doi.org/10.1007/978-3-540-93802-6_24

[15] W. Kosiński, P. Prokopowicz, and D. Ślęzak, *Neural Networks and Soft Computing: Proceedings of the Sixth International Conference on Neural Networks and Soft Computing, Zakopane, Poland, June 11–15, 2002*. Heidelberg: Physica-Verlag HD, 2003, ch. On Algebraic Operations on Fuzzy Reals, pp. 54–61. [Online]. Available: http://dx.doi.org/10.1007/978-3-7908-1902-1_8

[16] P. Prokopowicz, "Adaptation of rules in the fuzzy control system using the arithmetic of ordered fuzzy numbers," in *Artificial Intelligence and Soft Computing - ICAISC 2008*, ser. Lecture Notes in Computer Science, L. Rutkowski, R. Tadeusiewicz, L. Zadeh, and J. Zurada, Eds. Springer Berlin Heidelberg, 2008, vol. 5097, pp. 306–316. [Online]. Available: http://dx.doi.org/10.1007/978-3-540-69731-2_30

[17] L. Zadeh, "The concept of a linguistic variable and its application to approximate reasoning I," *Information Sciences*, vol. 8, no. 3, pp. 199 – 249, 1975. [Online]. Available: http://dx.doi.org/10.1016/0020-0255(75)90036-5

[18] H. T. Nguyen, "A note on the extension principle for fuzzy sets," *Journal of Mathematical Analysis and Applications*, vol. 64, no. 2, pp. 369 – 380, 1978. [Online]. Available: http://dx.doi.org/10.1016/0022-247X(78)90045-8

[19] D. Dubois and H. Prade, "Operations on fuzzy numbers," *International Journal of Systems Science*, vol. 9, no. 6, pp. 613–626, 1978, cited By 1218. [Online]. Available: http://dx.doi.org/10.1080/00207727808941724

[20] W. Kosiński, P. Prokopowicz, and A. Rosa, "Defuzzification functionals of ordered fuzzy numbers," *Fuzzy Systems, IEEE Transactions on*, vol. 21, no. 6, pp. 1163–1169, Dec 2013. [Online]. Available: http://dx.doi.org/10.1109/TFUZZ.2013.2243456

[21] P. Prokopowicz, *Proceedings of the Second International Afro-European Conference for Industrial Advancement AECIA 2015*. Cham: Springer International Publishing, 2016, ch. The Directed Inference for the Kosinski's Fuzzy Number Model, pp. 493–503. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-29504-6_46

[22] E. Kaucher, *Fundamentals of Numerical Computation (Computer-Oriented Numerical Analysis)*. Vienna: Springer Vienna, 1980, ch. Interval Analysis in the Extended Interval Space IR, pp. 33–49. [Online]. Available: http://dx.doi.org/10.1007/978-3-7091-8577-3_3

[23] W. Pedrycz and F. Gomide, *An introduction to fuzzy sets: analysis and design. With a foreword by Lotfi A. Zadeh.* Cambridge, MA: MIT Press, 1998.

[24] A. Piegat, *Fuzzy modeling and control.* Heidelberg: Physica-Verlag, 2001.

[25] R. Kolesnik, P. Prokopowicz, and W. Kosinski, "Fuzzy calculator - useful tool for programming with fuzzy algebra," in *Artificial Intelligence and Soft Computing - ICAISC 2004, 7th International Conference, Zakopane, Poland, June 7-11, 2004, Proceedings*, 2004, pp. 320–325. [Online]. Available: http://dx.doi.org/10.1007/978-3-540-24844-6_45

[26] J. M. Czerniak, W. Dobrosielski, Ł. Apiecionek, and D. Ewald, "Representation of a trend in ofn during fuzzy observance of the water level from the crisis control center," in *Proceedings of the 2015 Federated Conference on Computer Science and Information Systems*, ser. Annals of Computer Science and Information Systems, M. Ganzha, L. Maciaszek, and M. Paprzycki, Eds., vol. 5. IEEE, 2015, pp. 443–447. [Online]. Available: http://dx.doi.org/10.15439/2015F217

[27] P. Prokopowicz and S. Golsefid, "Aggregation operator for ordered fuzzy numbers concerning the direction," in *Artificial Intelligence and Soft Computing*, ser. Lecture Notes in Computer Science, L. Rutkowski, M. Korytkowski, R. Scherer, R. Tadeusiewicz, L. Zadeh, and J. Zurada, Eds. Springer International Publishing, 2014, vol. 8467, pp. 267–278. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-07173-2_24

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 129–136
DOI: 10.15439/2016F30
ACSIS, Vol. 8. ISSN 2300-5963

Dispersed decision-making system with selected fusion methods from the measurement level—case study with medical data

Małgorzata Przybyła-Kasperek
University of Silesia
Institute of Computer Science
Będzińska 39, 41-200 Sosnowiec, Poland
Email: malgorzata.przybyla-kasperek@us.edu.pl

***Abstract*—In the paper issues related to the use of dispersed knowledge in medicine are discussed. The main aim of the article is to investigate the efficiency of inference of seven selected fusion methods in a dispersed decision-making system. The dispersed system was proposed by the author in previous papers. The examined fusion methods - the maximum rule, the minimum rule, the median rule, the sum rule, the probabilistic product method, the method that is based on the theory of evidence and the method that is based on decision templates - are well known from the literature. In the paper two medical data sets from the UCI repository were used. Based on the obtained results it was concluded that for one data set the maximum rule generates the best results, and for other data set better methods are the sum rule and the median rule.**

I. Introduction

IN THE paper issues related to the use of dispersed knowledge are considered. The use of dispersed knowledge is particularly important in medicine because many medical centers independently collect information on patients or studied cases. That means that we have access to dispersed knowledge. If this knowledge is from one field, for example, one disease entity, it is possible to use all of the collected information at the same time which should improve the efficiency of inference.

The use of dispersed knowledge was investigated by the author in the earlier papers [9], [10], [12], [13], [14]. In the paper [10] a dispersed decision-making system with dynamic structure was proposed. This system is used in this article. The use of dispersed knowledge in medicine was also considered by the author [8], [11]. The novelty that is proposed in this article is to apply the seven fusion methods in a dispersed system.

The issue of combining classifiers is a very important aspect in the literature [1], [2], [3], [4], [5], [6], [17]. The aim of the issue is to improve the quality of the classification by combining the results of the predictions of the base classifiers. One of the basic questions is what combination rule to use. In this article different fusion methods are considered. These methods are very popular and are described in numerous papers [1], [4], [5], [6], [7], [15], [16]. In this article, seven selected fusion methods were tested in conjunction with a dispersed system. The obtained results were compared and conclusions were drawn.

The paper is organized as follows. The second section briefly describes the dispersed decision-making system. The third section describes the fusion method that are used. The fourth section shows a description and the results of experiments. The article concludes with a short summary in the fifth section.

II. A dispersed decision-making system - brief overview

In the article [10] a dispersed decision-making system was proposed by the author. This system is also used in the paper. The main assumptions of the system are very briefly described below. A detailed discussion is omitted because it is not the goal of this article. A detailed description of the system can be found in the paper [10].

It was assumed that in the system the knowledge is available in a dispersed form. The dispersed form means that we have several decision tables. The set of local knowledge bases that contain data from one domain is pre-specified. One condition must be satisfied by the local knowledge bases. They must have common decision attributes. We assume that each local decision table $D_{ag} = (U_{ag}, A_{ag}, d_{ag})$ is managed by one agent, which is called a resource agent ag. We want to designate homogeneous groups of resource agents. The agents who agree on the classification for a test object into the decision classes will be combined in the group that is called a cluster. It is realized in a two step process with the negotiation stage. For more details, please refer to the paper [10]. For each cluster that contains at least two resource agents, a superordinate agent is defined, which is called a synthesis agent, as_j, where j is the number of cluster. The synthesis agent, as_j, has access to knowledge that is the result of the process of inference carried out by the resource agents that belong to its subordinate group. A formal definition of a dispersed decision-making system is as follows.

Definition 2.1: By a dispersed decision-making system with dynamically generated clusters we mean $WSD^{dyn}_{Ag} = \langle Ag, \{D_{ag} : ag \in Ag\}, \{As_x : x \text{ is a classified}$

object}, $\{\delta_x : x \text{ is a classified object}\}\rangle$ where Ag is a finite set of resource agents; $\{D_{ag} : ag \in Ag\}$ is a set of decision tables of resource agents; As_x is a finite set of synthesis agents defined for clusters dynamically generated for the test object x, $\delta_x : As_x \rightarrow 2^{Ag}$ is a injective function that each synthesis agent assigns a cluster generated due to classification of the object x.

On the basis of the knowledge of agents from one cluster, local decisions are taken. An important problem that occurs is to eliminate inconsistencies in the knowledge stored in different knowledge bases. In previous papers the approximated method of the aggregation of decision tables have been used to eliminate inconsistencies in the knowledge [9], [10]. In this paper, we also use this method. In the method for every cluster, a kind of combined information is determined. This combined information is in the form of aggregated decision table. Object of this table are constructed by combining relevant object from decision tables of the resource agents that belong to one cluster. Based on the aggregated decision tables global decisions are taken using the fusion method.

III. Fusion methods

In this study, seven fusion methods are used that belong to the measurement level group [1], [6]. In this group of methods each base classifier generates a vector that represents the probability of an observation belonging to different decision classes. Thus, for each synthesis agent, such a vector of probabilities is generated. This is realized in the following manner. A c-dimensional vector of values $[\mu_{j,1}(x), \ldots, \mu_{j,c}(x)]$ is generated for each j-th cluster, where c is the number of all of the decision classes. The value $\mu_{j,i}(x)$ determines the level of certainty with which the decision v_i is taken by agents belonging to the cluster j for a given test object x. This vector will be defined on the basis of relevant objects. From each aggregated decision table and from each decision class, the smallest set containing at least m_2 objects for which the values of conditional attributes bear the greatest similarity to the test object is chosen. The value of the parameter m_2 is selected experimentally. The value $\mu_{j,i}(x)$ is equal to the average value of the similarity of the test object to the relevant objects form j-th aggregated decision table, belonging to the decision class v_i. In this way, for each cluster the vector of probabilities is generated.

In the paper [5], it was proposed that the classifier outputs can be organized in a decision profile (DP) as the matrix. The decision profile is a matrix with dimensions $card\{As_x\} \times c$, where As_x is a finite set of synthesis agents defined for the test object x and c is the number of all of the decision classes. The decision profile is defined as follows

$$DP(x) = \begin{bmatrix} \mu_{1,1}(x) & \cdots & \mu_{1,c}(x) \\ \cdots & & \\ \mu_{card\{As_x\},1}(x) & \cdots & \mu_{card\{As_x\},c}(x) \end{bmatrix}$$

The j-th row of the matrix saves the output of j-th synthesis agents and the i-th column of the matrix saves support from agents As_x for decision class i.

The maximum rule and the minimum rule

The maximum rule and the minimum rule consist in the designation of the maximum or the minimum value of the probability values assigned to this class by each cluster for each decision class. The set of decisions taken by the dispersed system is the set of classes that have the maximum of these values. Thus, the sets of global decisions that are generated using these methods are defined as follows: the maximum rule

$$\hat{d}_{WSD_{Ag}^{dyn}}(x) = \arg\max_{i\in\{1,\ldots,c\}} \Big\{ \max_{j\in\{1,\ldots,card\{As_x\}\}} \mu_{j,i}(x) \Big\},$$

the minimum rule

$$\hat{d}_{WSD_{Ag}^{dyn}}(x) = \arg\max_{i\in\{1,\ldots,c\}} \Big\{ \min_{j\in\{1,\ldots,card\{As_x\}\}} \mu_{j,i}(x) \Big\}.$$

The median rule

In the median rule the median value of the probability values is determined for each decision class. The set of decisions taken by the dispersed system is the set of classes that have the maximum of these medians

$$\hat{d}_{WSD_{Ag}^{dyn}}(x) = \arg\max_{i\in\{1,\ldots,c\}} \Big\{ \mathrm{med}_{\mathrm{j}\in\{1,\ldots,\mathrm{card}\{\mathrm{As_x}\}\}}\ \mu_{\mathrm{j,i}}(\mathrm{x}) \Big\}.$$

The sum rule

The sum rule consists in the designation of the sum of the probability values assigned to this class by each cluster for each decision class. The set of decisions taken by the dispersed system is the set of classes that have the maximum of these sums. Thus, the set of global decisions that are generated using the sum rule is defined as follows

$$\hat{d}_{WSD_{Ag}^{dyn}}(x) = \arg\max_{i\in\{1,\ldots,c\}} \Big\{ \sum_{j=1}^{card\{As_x\}} \mu_{j,i}(x) \Big\}.$$

The probabilistic product method

The probabilistic product method was proposed in the paper [16]. For each decision v_i, the value is determined

$$\frac{1}{P(v_i)^{L-1}} \prod_{j=1}^{L} \mu_{j,i}(x), \tag{1}$$

where the probabilities $P(v_i)$ are estimates based on the training sets of the synthesis agents. $P(v_i) = \frac{N_i}{N}$, where $N = \sum_{as\in As_x} card\{U_{as}\}$ is the sum of the number of objects of the aggregated decision tables and $N_i = \sum_{as\in As_x} card\{X_{as}^{v_i}\}$ is the sum of the number of objects from the decision class v_i of the aggregated decision tables. The set of decisions taken by the dispersed system is the set of classes that have the highest value as calculated by Formula 1.

Method that is based on decision templates

The method that is based on decision templates was proposed in the paper [5]. The decision templates of each class are defined in this method. The decision template for class v_i is the average of the decision profiles of the objects of the training set labelled in class v_i. In the dispersed decision-making system the decision templates of the synthesis agents are constructed based on the decision templates of the resource agents that belong to its subordinate cluster. Therefore, the

decision profiles of the resource agents for the training objects were calculated

$$DP^{Ag}(x) = \begin{bmatrix} \bar{\mu}_{1,1}(x) & \cdots & \bar{\mu}_{1,c}(x) \\ & & \cdots \\ \bar{\mu}_{j,1}(x) & \cdots & \bar{\mu}_{j,c}(x) \\ & & \cdots \\ \bar{\mu}_{card\{Ag\},1}(x) & \cdots & \bar{\mu}_{card\{Ag\},c}(x) \end{bmatrix}$$

the values $\bar{\mu}$ are defined as follows:

$$\bar{\mu}_{j,i}(x) = \frac{\sum_{y \in U^{rel}_{ag_j} \cap X^{v_i}_{ag_j}} s(x,y)}{card\{U^{rel}_{ag_j} \cap X^{v_i}_{ag_j}\}},$$

where $U^{rel}_{ag_j}$ is the subset of relevant objects selected from the decision table D_{ag_j} of a resource agent ag_j and $X^{v_i}_{ag_j} = \{x \in U_{ag_j} : d_{ag_j}(x) = v_i\}$ is the decision class of the decision table of resource agent ag_j; $s(x,y)$ is the measure of similarity between objects x and y. Note that in order to construct the decision profiles of the resource agents for training objects, the same sets of objects must be included in the decision tables of the resource agents. Thus, the assumptions of a dispersed system must be narrowed slightly when we use this method, and therefore the system will no longer be so general. Based on the decision profiles of the resource agents, the decision templates of the resource agents are determined

$$DT^{Ag}_{v_i} = \frac{1}{card\{Z_{v_i}\}} \sum_{x \in Z_{v_i}} DP^{Ag}(x),$$

where Z_{v_i} is the set of objects from the training set that belong to the class v_i. The training process consists in determining the decision templates of the synthesis agents for each class $DT^{As_x}_{v_i}$, $i \in \{1, \ldots, c\}$. The decision templates of the synthesis agents are determined based on the decision templates of the resource agents in the following way. The j-th row of the decision template should save the output of the j-th synthesis agent. The j-th row of the decision template is calculated as the average of the rows of the decision templates of the resource agents that correspond to the resource agents that belong to the cluster that is subordinate to the j-th synthesis agent

$$DT^{As_x}_{v_i} = \begin{bmatrix} \frac{\sum_{ag_p \in \delta_x(as_1)} DT^{Ag}_{v_i}(p,1)}{card\{\delta_x(as_1)\}} & \cdots & \frac{\sum_{ag_p \in \delta_x(as_1)} DT^{Ag}_{v_i}(p,c)}{card\{\delta_x(as_1)\}} \\ \cdots & & \\ \frac{\sum_{ag_p \in \delta_x(as_k)} DT^{Ag}_{v_i}(p,1)}{card\{\delta_x(as_k)\}} & \cdots & \frac{\sum_{ag_p \in \delta_x(as_k)} DT^{Ag}_{v_i}(p,c)}{card\{\delta_x(as_k)\}} \end{bmatrix}$$

where $As_x = \{as_1, \ldots, as_k\}$ and $DT^{Ag}_{v_i}(p,l)$ is an element at the p-th row and the l-th column of the matrix $DT^{Ag}_{v_i}$.

The next step is to calculate the similarity measure between the decision profile of the test object and the decision templates $DT^{As_x}_{v_i}$ of each class $i \in \{1, \ldots, c\}$. Four different similarity measures were used in this study:

1) The similarity measure that uses the normalised Euclidean distance

$$s(DP(x), DT^{As_x}_{v_i}) = 1 - \frac{1}{card\{As_x\} \cdot c} \cdot \sum_{m=1}^{card\{As_x\}} \sum_{l=1}^{c} \left(DP_{m,l}(x) - DT^{As_x}_{v_i}(m,l)\right)^2,$$

where $DP_{m,l}(x)$ and $DT^{As_x}_{v_i}(m,l)$ is an element at the m-th row and the l-th column of the matrix $DP(x)$ or $DT^{As_x}_{v_i}$ respectively.

2) The similarity measure that uses the symmetric difference defined by the Hamming distance

$$s(DP(x), DT^{As_x}_{v_i}) = 1 - \frac{1}{card\{As_x\} \cdot c} \cdot \sum_{m=1}^{card\{As_x\}} \sum_{l=1}^{c} \left|DP_{m,l}(x) - DT^{As_x}_{v_i}(m,l)\right|$$

3) The Jaccard similarity coefficient

$$s(DP(x), DT^{As_x}_{v_i}) = \frac{\sum_{m=1}^{card\{As_x\}} \sum_{l=1}^{c} \min\{DP_{m,l}(x), DT^{As_x}_{v_i}(m,l)\}}{\sum_{m=1}^{card\{As_x\}} \sum_{l=1}^{c} \max\{DP_{m,l}(x), DT^{As_x}_{v_i}(m,l)\}}$$

4) The similarity measure that uses the symmetric difference

$$s(DP(x), DT^{As_x}_{v_i}) = 1 - \frac{1}{card\{As_x\} \cdot c} \cdot \sum_{m=1}^{card\{As_x\}} \sum_{l=1}^{c} \max\Big\{ \min\{DP_{m,l}(x), 1 - DT^{As_x}_{v_i}(m,l)\}, \min\{1 - DP_{m,l}(x), DT^{As_x}_{v_i}(m,l)\}\Big\}$$

All of these measures were also used by the authors of the method based on decision templates in the paper [5]. The set of decisions taken by the dispersed system is defined by selecting the decision that has the maximum value of similarity.

Method that is based on the theory of evidence

The method that is based on the theory of evidence was proposed in the paper [15]. In this method as in the decision templates method, the decision templates $DT^{As_x}_{v_i}, i \in \{1, \ldots, c\}$ are designated from the data. And like in the previous method, the same sets of objects must be included in the decision tables of the resource agents, which means that the assumptions of the system are a bit narrow. Instead of calculating the similarity between the decision template $DT^{As_x}_{v_i}$ and the decision profile $DP(x)$, the Dempster-Shafer theory is used in this method and the belief is calculated. The following steps are performed in the Dempster-Shafer algorithm:

1) Let $DT^{As_x}_{v_i}(m,\cdot)$ denote the m-th row of the decision template for class v_i and $DP_{m,\cdot}(x)$ denote the m-th row of the decision profile for the object x. The proximity between the prediction of the m-th synthesis agent $DP_{m,\cdot}(x)$ and the m-th row of the decision template for every class $v_i, i \in \{1, \ldots, c\}$ and for every synthesis agent $m \in \{1, \ldots, card\{As_x\}\}$ is calculated

$$\phi_{i,m}(x) = \frac{(1 + \|DT_{v_i}(m,\cdot) - DP_{m,\cdot}(x)\|^2)^{-1}}{\sum_{k=1}^{c}(1 + \|DT_{v_k}(m,\cdot) - DP_{m,\cdot}(x)\|^2)^{-1}}$$

where $\|\cdot\|$ is the norm. The Euclidean norm was applied in this study.

2) For every class $v_i, i \in \{1, \ldots, c\}$ and for every synthesis agent $m \in \{1, \ldots, card\{As_x\}\}$ the following belief degrees are calculated

$$Bel_i(DP_{m,\cdot}(x)) = \frac{\phi_{i,m}(x) \prod_{k \neq i}(1 - \phi_{k,m}(x))}{1 - \phi_{i,m}(x)[1 - \prod_{k \neq i}(1 - \phi_{k,m}(x))]}$$

3) The Dempster-Shafer membership degrees for every class $v_i, i \in \{1, \ldots, c\}$ are calculated

$$\mu_i(x) = K \prod_{m=1}^{card\{As_x\}} Bel_i(DP_{m,\cdot}(x))$$

where K is a constant that ensures that $\mu_i(x) \leq 1$.

The set of decisions taken by the dispersed system is defined by selecting the decision that has the maximum value of the Dempster-Shafer membership degrees.

IV. Experiments and results

In this section, experiments with two data sets from the medical field and seven fusion methods are described. The aim of the experiment is to compare the obtained results and, if possible, to choose the best fusion method from the methods that were examined. In the article data from the medical field are examined, as the use of dispersed knowledge is particularly important in this area. We consider a situation in which knowledge in the same field is stored independently in several medical centers (hospitals, laboratories). The use of all of this knowledge at the same time should improve the efficiency of inference.

A. Data

For the experiments the following data, which are in the UCI repository (archive.ics.uci.edu/ml/), were used: Lymphography data set, Primary Tumor data set. Both sets of data was obtained from the University Medical Centre, Institute of Oncology, Ljubljana, Yugoslavia (M. Zwitter and M. Soklic provided this data). Lymphography is a medical imaging technique in which a radiocontrast agent is injected, and then an X-ray picture is taken to visualize structures of the lymphatic system. In the Primary Tumor data set, on the basis of values of attributes such as histologic-type, supraclavicular etc. a decision is taken where (of 22 organs) the cancer cells are located. In order to determine the efficiency of inference each data set was divided into two disjoint subsets: a training set and a test set. A numerical summary of the data sets is as follows: Lymphography: # The training set - 104; # The test set - 44; # Conditional - 18; # Decision - 4; Primary Tumor: # The training set - 237; # The test set - 102; # Conditional - 17; # Decision - 22. The next step of data preparation consists in dispersion of datasets. The training set was divided into a set of decision tables. Divisions with a different number of decision tables were considered. For each of the data sets used, the decision-making system with five different versions (with 3, 5, 7, 9 and 11 resource agents) were considered. For these systems, we use the following designations: WSD^{dyn}_{Ag1} - 3 resource agents; WSD^{dyn}_{Ag2} - 5 resource agents; WSD^{dyn}_{Ag3} - 7 resource agents; WSD^{dyn}_{Ag4} - 9 resource agents; WSD^{dyn}_{Ag5} - 11 resource agents.

The dispersion of a data set proceeded in a random way but under certain conditions that were defined by the author of the paper. This process was carried out as follows. In the first step the cardinality of set of conditional attributes in each decision table of resource agent was determined, and the number of common conditional attributes of decision tables was defined. These values were defined by the author. Then the conditional attributes were assigned to the decision tables so that the conditions which were defined earlier were met. Each universe of decision tables includes all objects from the data set. However, after the dispersion the identifiers of objects are not stored in the decision tables, so it is not possible to reconstruct the original data set. The cardinalities of sets of conditional attributes of decision tables in the systems are given in Table I. Table II presents the cardinalities of all nonempty intersection of conditional attributes sets. The data set was dispersed in such a way to obtain a set of decision tables that could be collected independently by different medical centers. The author is aware that the results of experiments obtained for real data would be much more valuable, however, the author does not have access to such a data.

B. Quality measures and parameters

Some of the considered fusion methods have the property that the final decision may have ties. In order to analyze these properties the appropriate classification measures were applied, which are adapted to this situation. The measures of determining the quality of the classification are: *estimator of classification error* e in which an object is considered to be properly classified if the decision class used for the object belonged to the set of global decisions generated by the system; *estimator of classification ambiguity error* e_{ONE} in which object is considered to be properly classified if only one, correct value of the decision was generated to this object; *the average size of the global decisions sets* $\overline{d}_{WSD^{dyn}_{Ag}}$ generated for a test set.

In this article, the measures were applied that are adequate to the situation in which a set of decision instead of one decision is generated for a test object. Note that in the paper the classification problem is being considered but not in a standard version. Therefore, the standard measures, such as error rate, recall, precision and F-meaure are not appropriate. However the estimator of classification error and the estimator of classification ambiguity error can be considered as a modification of the standard error rate measure. If $\overline{d}_{WSD^{dyn}_{Ag}} = 1$ then e and e_{ONE} are equal to the standard error rate. During the experiments, the author tried to use measures such as precision and recall. The values were calculated based on the cases in which an unambiguous decision (only one decision) was generated by the system. Therefore, some test objects were not taken into account in these calculations. This caused that

TABLE I
THE CARDINALITIES OF SETS OF CONDITIONAL ATTRIBUTES

Data set, System	A_{ag_1}	A_{ag_2}	A_{ag_3}	A_{ag_4}	A_{ag_5}	A_{ag_6}	A_{ag_7}	A_{ag_8}	A_{ag_9}	$A_{ag_{10}}$	$A_{ag_{11}}$
Lymphography, WSD^{dyn}_{Ag1}	7	9	6	-	-	-	-	-	-	-	-
Lymphography, WSD^{dyn}_{Ag2}	5	5	5	4	4	-	-	-	-	-	-
Lymphography, WSD^{dyn}_{Ag3}	4	4	4	4	4	3	3	-	-	-	-
Lymphography, WSD^{dyn}_{Ag4}	3	3	3	3	3	2	2	2	2	-	-
Lymphography, WSD^{dyn}_{Ag5}	2	2	2	2	2	2	2	2	2	3	3
Primary Tumor, WSD^{dyn}_{Ag1}	7	9	5	-	-	-	-	-	-	-	-
Primary Tumor, WSD^{dyn}_{Ag2}	5	4	5	4	4	-	-	-	-	-	-
Primary Tumor, WSD^{dyn}_{Ag3}	4	4	4	4	4	2	2	-	-	-	-
Primary Tumor, WSD^{dyn}_{Ag4}	3	3	3	3	2	2	2	2	2	-	-
Primary Tumor, WSD^{dyn}_{Ag5}	2	2	2	2	2	2	2	2	2	2	3

TABLE II
THE CARDINALITIES OF ALL NONEMPTY INTERSECTION OF CONDITIONAL ATTRIBUTES SETS

Data set, System	# Intersection of conditional attributes sets
Lymphography, WSD^{dyn}_{Ag1}	$\lvert A_{ag_1} \cap A_{ag_2}\rvert = 2, \lvert A_{ag_2} \cap A_{ag_3}\rvert = 2$
Lymphography, WSD^{dyn}_{Ag2}	$\lvert A_{ag_1} \cap A_{ag_2}\rvert = 2, \lvert A_{ag_3} \cap A_{ag_5}\rvert = 1, \lvert A_{ag_4} \cap A_{ag_5}\rvert = 2$
Lymphography, WSD^{dyn}_{Ag3}	$\lvert A_{ag_1} \cap A_{ag_2}\rvert = 2, \lvert A_{ag_2} \cap A_{ag_3}\rvert = 2, \lvert A_{ag_4} \cap A_{ag_5}\rvert = 2, \lvert A_{ag_6} \cap A_{ag_7}\rvert = 2$
Lymphography, WSD^{dyn}_{Ag4}	$\lvert A_{ag_1} \cap A_{ag_2}\rvert = 1, \lvert A_{ag_3} \cap A_{ag_4}\rvert = 1, \lvert A_{ag_4} \cap A_{ag_5}\rvert = 1, \lvert A_{ag_6} \cap A_{ag_7}\rvert = 1, \lvert A_{ag_8} \cap A_{ag_9}\rvert = 1$
Lymphography, WSD^{dyn}_{Ag5}	$\lvert A_{ag_1} \cap A_{ag_2}\rvert = 1, \lvert A_{ag_3} \cap A_{ag_4}\rvert = 1, \lvert A_{ag_5} \cap A_{ag_6}\rvert = 1, \lvert A_{ag_6} \cap A_{ag_7}\rvert = 1, \lvert A_{ag_8} \cap A_{ag_9}\rvert = 1,$ $\lvert A_{ag_{10}} \cap A_{ag_{11}}\rvert = 1$
Primary Tumor, WSD^{dyn}_{Ag1}	$\lvert A_{ag_1} \cap A_{ag_2}\rvert = 2, \lvert A_{ag_2} \cap A_{ag_3}\rvert = 2$
Primary Tumor, WSD^{dyn}_{Ag2}	$\lvert A_{ag_1} \cap A_{ag_2}\rvert = 2, \lvert A_{ag_3} \cap A_{ag_5}\rvert = 1, \lvert A_{ag_4} \cap A_{ag_5}\rvert = 2$
Primary Tumor, WSD^{dyn}_{Ag3}	$\lvert A_{ag_1} \cap A_{ag_2}\rvert = 2, \lvert A_{ag_2} \cap A_{ag_3}\rvert = 2, \lvert A_{ag_4} \cap A_{ag_5}\rvert = 2, \lvert A_{ag_6} \cap A_{ag_7}\rvert = 1$
Primary Tumor, WSD^{dyn}_{Ag4}	$\lvert A_{ag_1} \cap A_{ag_2}\rvert = 1, \lvert A_{ag_3} \cap A_{ag_4}\rvert = 1, \lvert A_{ag_4} \cap A_{ag_5}\rvert = 1, \lvert A_{ag_6} \cap A_{ag_7}\rvert = 1, \lvert A_{ag_8} \cap A_{ag_9}\rvert = 1$
Primary Tumor, WSD^{dyn}_{Ag5}	$\lvert A_{ag_1} \cap A_{ag_2}\rvert = 1, \lvert A_{ag_3} \cap A_{ag_4}\rvert = 1, \lvert A_{ag_5} \cap A_{ag_6}\rvert = 1, \lvert A_{ag_6} \cap A_{ag_7}\rvert = 1, \lvert A_{ag_8} \cap A_{ag_9}\rvert = 1,$ $\lvert A_{ag_{10}} \cap A_{ag_{11}}\rvert = 1$

in some cases the calculation of precision and recall for a class was impossible because there was no test object from the class. For this reason, the calculation of micro-averaged and macro-averaged precision or recall would also be inappropriate. In the experiments the above-defined three measures were used.

In the description of the results of experiments for clarity some designations for parameters have been adopted: m_1 - which determines the number of relevant objects that are selected from each decision class of the decision table and are then used in the process of cluster generation; p - parameter which occurs in the definition of friendship, conflict and neutrality relations; m - parameter of the approximated method of the aggregation of decision tables; m_2 - parameter which determines the number of relevant objects that are used to generate decision of one cluster in the method of conflict analysis (the maximum rule, the minimum rule, the median rule, the sum rule, the probabilistic product method, the method that is based on the theory of evidence and the method that is based on decision templates).

C. Results

At the beginning of experiments the process of parameters optimization was carried out. A series of tests for different parameter values were performed: $m_1 \in \{1, 4, 7, 10, 13\}$, $m_2, m_3 \in \{1, \ldots, 10\}$ and $p \in \{0.05, 0.1, 0.15, 0.2\}$. From all of the obtained results, one was selected that guaranteed a minimum value of estimator of classification error (e), while maintaining the smallest possible value of the average size of the global decisions sets ($\overline{d}_{WSD^{dyn}_{Ag}}$). In tables presented below the best results, obtained for optimal values of the parameters, are given. In the tables the following information is given: the name of dispersed decision-making system (System); the selected, optimal parameter values (m_1/p/m_2/m_3); the three measures discussed earlier e, e_{ONE}, $\overline{d}_{WSD^{dyn}_{Ag}}$; the time t needed to analyse a test set expressed in minutes. In the tables below the best results in terms of the measures e and $\overline{d}_{WSD^{dyn}_{Ag}}$ are bolded.

The results of the experiments with the Lymphography data set are presented in Table III. Based on the results for the Lymphography data set it can be concluded that all of the examined methods generate almost unambiguous results - the average size of the global decision sets is very close to or equal to 1. On the basis of detailed analysis of vectors of probabilities generated by the individual classifiers, it was concluded that the reason of this situation is that for the Lymphography data there is very few dummy agents. That is undecided agents who assign the same probability value to many different decision values. For some test objects there

TABLE III
SUMMARY OF EXPERIMENTS RESULTS WITH THE LYMPHOGRAPHY DATA SET

System	Maximum rule		Minimum rule	
	$m_1/p/m_2/m_3$	$e/e_{ONE}/\bar{d}_{WSD_{Ag}^{dyn}}/t$	$m_1/p/m_2/m_3$	$e/e_{ONE}/\bar{d}_{WSD_{Ag}^{dyn}}/t$
WSD_{Ag1}^{dyn}	13/0.05/1/2	0.136/0.364/1.227/0.01	13/0.05/1/3	0.136/0.159/1.023/0.01
WSD_{Ag2}^{dyn}	7/0.05/5/5	**0.136**/0.136/**1**/0.02	1/0.05/1/2	0.182/0.250/1.068/0.01
WSD_{Ag3}^{dyn}	7/0.05/7/4	0.205/0.318/1.114/0.02	1/0.05/4/7	0.159/0.159/1/0.02
WSD_{Ag4}^{dyn}	1/0.05/4/4	0.273/0.614/1.341/0.04	1/0.05/1/3	0.159/0.295/1.136/0.02
WSD_{Ag5}^{dyn}	1/0.05/2/2	0.591/0.886/1.295/0.15	10/0.05/6/9	**0.205**/0.341/**1.136**/0.27
System	Median rule		Sum rule	
	$m_1/p/m_2/m_3$	$e/e_{ONE}/\bar{d}_{WSD_{Ag}^{dyn}}/t$	$m_1/p/m_2/m_3$	$e/e_{ONE}/\bar{d}_{WSD_{Ag}^{dyn}}/t$
WSD_{Ag1}^{dyn}	4/0.05/1/2	**0.136**/0.136/**1**/0.01	4/0.05/1/2	**0.136**/0.136/**1**/0.01
WSD_{Ag2}^{dyn}	13/0.05/3/7	0.136/0.182/1.045/0.01	4/0.05/6/6	**0.136**/0.136/**1**/0.02
WSD_{Ag3}^{dyn}	1/0.05/4/1	**0.114**/0.455/**1.341**/0.02	4/0.05/5/9	0.136/0.136/1/0.01
WSD_{Ag4}^{dyn}	10/0.05/7/8	0.159/0.182/1.023/0.02	7/0.05/5/9	**0.159**/0.159/**1**/0.03
WSD_{Ag5}^{dyn}	4/0.05/2/5	0.227/0.273/1.045/0.25	13/0.05/1/3	0.205/0.455/1.250/0.25
System	Probabilistic product		Method that is based on the theory of evidence	
	$m_1/p/m_2/m_3$	$e/e_{ONE}/\bar{d}_{WSD_{Ag}^{dyn}}/t$	$m_1/p/m_2/m_3$	$e/e_{ONE}/\bar{d}_{WSD_{Ag}^{dyn}}/t$
WSD_{Ag1}^{dyn}	7/0.05/1/2	0.318/0.318/1/0.01	1/0.05/2/5	0.182/0.182/1/0.01
WSD_{Ag2}^{dyn}	1/0.05/1/1	0.295/0.341/1.045/0.01	1/0.05/2/8	0.250/0.250/1/0.01
WSD_{Ag3}^{dyn}	7/0.05/1/1	0.318/0.386/1.068/0.02	1/0.05/3/8	0.250/0.250/1/0.02
WSD_{Ag4}^{dyn}	1/0.05/1/1	0.182/0.318/1.136/0.02	7/0.05/1/2	0.273/0.273/1/0.02
WSD_{Ag5}^{dyn}	13/0.05/1/1	0.205/0.455/1.250/0.24	1/0.05/1/1	0.341/0.341/1/0.13
System	Method that is based on decision templates			
	Euclidean distance		Hamming distance	
	$m_1/p/m_2/m_3$	$e/e_{ONE}/\bar{d}_{WSD_{Ag}^{dyn}}/t$	$m_1/p/m_2/m_3$	$e/e_{ONE}/\bar{d}_{WSD_{Ag}^{dyn}}/t$
WSD_{Ag1}^{dyn}	1/0.05/2/5	0.205/0.205/1/7.06	4/0.01/1/1	0.250/0.250/1/0.01
WSD_{Ag2}^{dyn}	1/0.05/2/8	0.250/0.250/1/0.01	10/0.05/1/1	0.250/0.250/1/0.02
WSD_{Ag3}^{dyn}	1/0.05/3/9	0.250/0.250/1/0.02	7/0.05/3/10	0.205/0.205/1/0.02
WSD_{Ag4}^{dyn}	7/0.05/1/3	0.318/0.318/1/0.02	7/0.05/3/5	0.227/0.227/1/0.03
WSD_{Ag5}^{dyn}	1/0.05/1/1	0.341/0.341/1/0.14	4/0.05/1/1	0.341/0.341/1/0.25
System	Jaccard similarity		Symmetric difference	
	$m_1/p/m_2/m_3$	$e/e_{ONE}/\bar{d}_{WSD_{Ag}^{dyn}}/t$	$m_1/p/m_2/m_3$	$e/e_{ONE}/\bar{d}_{WSD_{Ag}^{dyn}}/t$
WSD_{Ag1}^{dyn}	4/0.05/1/1	0.250/0.250/1/0.01	1/0.05/2/10	0.182/0.182/1/0.01
WSD_{Ag2}^{dyn}	10/0.05/1/1	0.250/0.250/1/0.01	7/0.05/4/6	0.318/0.318/1/0.02
WSD_{Ag3}^{dyn}	7/0.05/2/2	0.227/0.227/1/0.01	7/0.05/3/5	0.273/0.273/1/0.02
WSD_{Ag4}^{dyn}	7/0.05/3/5	0.227/0.227/1/0.02	10/0.05/3/9	0.295/0.295/1/0.03
WSD_{Ag5}^{dyn}	4/0.05/1/1	0.318/0.318/1/0.25	4/0.05/1/1	0.341/0.341/1/0.25

are dummy agents, therefore the maximum rule, the minimum rule and the median rule generate more ambiguous decisions. However, there is always a group of agents who assign different probabilities for decisions. Therefore, the sum rule generates unambiguous results. As can be seen, the analyzed methods can be divided into two groups due to the efficiency of inference. Better results are obtained by - the maximum rule, the minimum rule, the median rule and the sum rule, whereas worse results are obtained by - the probabilistic product method, the method that is based on the theory of evidence and the method that is based on decision templates. Among the first group of methods, definitely the sum rule produces better results than the minimum rule and the maximum rule and the median rule produces better results than the maximum rule. It is hard to choose the best method among the methods from the second group. Each of the methods at least once achieved the best result in this group.

The results of the experiments with the Primary Tumor data set are presented in Table IV. As can be seen, for the Primary Tumor data set only two from the analyzed methods produce unambiguous results - the method that is based on the theory of evidence and the method that is based on decision templates. Other methods generate results with the average size of the global decision sets that is close to 3. But it should be noted that the Primary Tumor data set has 22 decision classes and because of that even results with the average

TABLE IV
SUMMARY OF EXPERIMENTS RESULTS WITH THE PRIMARY TUMOR DATA SET

System	Maximum rule		Minimum rule	
	$m_1/p/m_2/m_3$	$e/e_{ONE}/\bar{d}_{WSD_{Ag}^{dyn}}/t$	$m_1/p/m_2/m_3$	$e/e_{ONE}/\bar{d}_{WSD_{Ag}^{dyn}}/t$
WSD_{Ag1}^{dyn}	1/0.1/2/1	**0.343**/0.863/**3.863**/0.01	1/0.15/1/3	0.382/0.775/2.569/0.01
WSD_{Ag2}^{dyn}	1/0.2/10/2	**0.304**/0.843/**3.176**/0.29	1/0.1/2/1	0.333/0.833/3.304/0.02
WSD_{Ag3}^{dyn}	1/0.2/2/3	**0.363**/0.843/**3.108**/0.08	1/0.15/3/2	0.412/0.824/3.098/0.12
WSD_{Ag4}^{dyn}	1/0.15/2/2	**0.314**/0.882/**3.765**/0.18	1/0.05/5/2	0.392/0.873/3.206/1.19
WSD_{Ag5}^{dyn}	1/0.05/3/3	0.392/0.863/3/1.47	1/0.05/3/3	**0.392**/0.853/**2.941**/1.47
System	Median rule		Sum rule	
	$m_1/p/m_2/m_3$	$e/e_{ONE}/\bar{d}_{WSD_{Ag}^{dyn}}/t$	$m_1/p/m_2/m_3$	$e/e_{ONE}/\bar{d}_{WSD_{Ag}^{dyn}}/t$
WSD_{Ag1}^{dyn}	1/0.15/3/1	0.382/0.794/2.627/0.02	4/0.2/2/1	0.382/0.784/2.618/0.01
WSD_{Ag2}^{dyn}	1/0.05/2/1	0.333/0.843/3.333/0.03	10/0.1/7/1	0.333/0.804/3.245/0.02
WSD_{Ag3}^{dyn}	1/0.15/3/2	0.402/0.824/3.137/0.12	1/0.15/3/2	0.412/0.814/3.098/0.12
WSD_{Ag4}^{dyn}	1/0.05/5/2	0.373/0.882/3.245/1.19	1/0.05/5/2	0.392/0.873/3.206/1.19
WSD_{Ag5}^{dyn}	1/0.05/3/3	0.392/0.862/2.990/1.47	1/0.05/3/3	**0.392**/0.853/**2.941**/1.47
System	Probabilistic product		Method that is based on the theory of evidence	
	$m_1/p/m_2/m_3$	$e/e_{ONE}/\bar{d}_{WSD_{Ag}^{dyn}}/t$	$m_1/p/m_2/m_3$	$e/e_{ONE}/\bar{d}_{WSD_{Ag}^{dyn}}/t$
WSD_{Ag1}^{dyn}	1/0.2/1/3	0.500/0.833/2.451/0.01	7/0.15/2/2	0.696/0.696/1/0.01
WSD_{Ag2}^{dyn}	1/0.05/1/2	0.647/0.931/2.520/0.01	1/0.15/4/1	0.696/0.696/1/0.08
WSD_{Ag3}^{dyn}	1/0.05/1/2	0.520/0.951/3.676/0.02	7/0.15/3/9	0.657/0.657/1/0.07
WSD_{Ag4}^{dyn}	1/0.05/1/3	0.431/0.922/3.784/0.04	4/0.1/4/4	0.627/0.627/1/0.12
WSD_{Ag5}^{dyn}	1/0.05/2/3	0.441/0.882/2.922/1.03	4/0.2/1/1	0.725/0.725/1/0.45
System	Method that is based on decision templates			
	Euclidean distance		Hamming distance	
	$m_1/p/m_2/m_3$	$e/e_{ONE}/\bar{d}_{WSD_{Ag}^{dyn}}/t$	$m_1/p/m_2/m_3$	$e/e_{ONE}/\bar{d}_{WSD_{Ag}^{dyn}}/t$
WSD_{Ag1}^{dyn}	10/0.2/8/8	0.696/0.696/1/0.03	4/0.2/8/10	0.706/0.706/1/0.03
WSD_{Ag2}^{dyn}	1/0.15/4/1	0.686/0.686/1/0.08	1/0.15/2/1	0.667/0.667/1/0.03
WSD_{Ag3}^{dyn}	4/0.05/7/10	0.667/0.667/1/0.08	7/0.15/3/8	0.696/0.696/1/0.07
WSD_{Ag4}^{dyn}	4/0.1/4/4	0.647/0.647/1/0.12	4/0.1/2/8	0.667/0.667/1/0.12
WSD_{Ag5}^{dyn}	4/0.2/1/1	0.725/0.725/1/0.46	4/0.05/3/9	0.745/0.745/1/0.45
System	Jaccard similarity		Symmetric difference	
	$m_1/p/m_2/m_3$	$e/e_{ONE}/\bar{d}_{WSD_{Ag}^{dyn}}/t$	$m_1/p/m_2/m_3$	$e/e_{ONE}/\bar{d}_{WSD_{Ag}^{dyn}}/t$
WSD_{Ag1}^{dyn}	4/0.2/8/8	0.716/0.716/1/0.04	4/0.05/6/8	0.794/0.794/1/0.03
WSD_{Ag2}^{dyn}	1/0.1/2/1	0.667/0.667/1/0.03	7/0.05/3/5	0.814/0.814/1/0.04
WSD_{Ag3}^{dyn}	4/0.05/5/5	0.696/0.696/1/0.07	13/0.1/2/5	0.784/0.784/1/0.07
WSD_{Ag4}^{dyn}	4/0.1/2/9	0.667/0.667/1/0.12	7/0.05/2/9	0.794/0.794/1/0.11
WSD_{Ag5}^{dyn}	4/0.05/3/10	0.745/0.745/1/0.46	7/0.1/1/1	0.824/0.824/1/0.36

number of global decisions sets less than 4 are interesting. On the basis of detailed analysis of vectors of probabilities generated by the individual classifiers, it was concluded that for some test objects there is a lot of dummy agents. Therefore, the maximum rule, the minimum rule, the median rule and the sum rule generate so ambiguous results. However, for about a third of the test objects unambiguous results are generated by these methods. Like before there are two groups of methods - those that generate better results (the maximum rule, the minimum rule, the median rule and the sum rule), and those that generate poorer results (the probabilistic product method, the method that is based on the theory of evidence and the method that is based on decision templates). In the first group of methods the best method is the maximum rule with the median rule in second place. The minimum rule and the sum rule obtain very similar results. In the second group of methods the best method is the probabilistic product method. From all of the analyzed similarity measures in the method that is based on decision templates the best results are generated by the similarity measure that uses the normalized Euclidean distance.

In conclusion, for both data the methods: the maximum rule, the minimum rule, the median rule and the sum rule produce significantly better results than the methods: the probabilistic product method, the method that is based on the theory of evidence and the method that is based on decision templates.

It is hard to say which method is the best, because for one data set the sum rule and the median rule generate better results, and for other data set the best method is the maximum rule. The reason for such results in the case of the probabilistic product method is that the method assigns greater weight to the smaller decision classes. The data sets that were analyzed have very diverse number of objects in the decision classes. Therefore, in this fusion method, the smaller decision classes are more awarded, which resulting in lower efficiency of inference. Therefore, the conclusion can be drawn that this is not the best method for data sets with very diverse, in terms of the number of objects, decision classes. For the method that is based on the theory of evidence and the method that is based on decision templates poor results are obtained, probably because a certain approximation was adopted during the training process. Due to the high computational complexity the decision templates of the synthesis agents are constructed based on the decision templates of the resource agents.

V. Conclusion

In this article, a significant problem that concerns the use of dispersed knowledge in medicine was considered. By dispersed knowledge in medicine we mean the set of knowledge bases that are accumulated independently in different medical centers. The knowledge base may contain information about various objects (patients), and may include various attributes (research methods). The use of dispersed knowledge will increase capabilities and efficiency in decision-making process.

In the paper a dispersed decision-making system with dynamic structure in conjunction with the seven fusion methods was considered. Dispersed medical data were used in the experiments: Lymphography data set and Primary Tumor data set. The conclusions, that were reached based on the results of experiments are as follows. The median rule, the sum rule, the maximum rule and the minimum rule generate the best results from the methods that were examined.

References

[1] Gatnar, E.: Multiple-model approach to classification and regression. PWN, Warsaw, 2008 (in Polish)

[2] Jakubczyc, J., Owoc, M.: Support of Contextual Classifier Ensemble Building, Proceedings of the 2015 Federated Conference on Computer Science and Information Systems, 2015, pp. 1683–1689, http://dx.doi.org/10.15439/2015F353

[3] Kalisch, M.: Supervised Context Classification Methods for an Industrial Machinery, Proceedings of the 2015 Federated Conference on Computer Science and Information Systems, 2015, pp. 1667–1672, http://dx.doi.org/10.15439/2015F292

[4] Kittler, J., Hatef, M., Duin, R.P.W., Matas, J.: On combining classifiers. IEEE Transactions on Pattern Analysis and Machine Intelligence, 20(3), 1998, pp. 226–239, http://dx.doi.org/10.1109/34.667881

[5] Kuncheva, L., Bezdek, J.C., Duin, R.P.W.: Decision templates for multiple classifier fusion: an experimental comparison. Pattern Recognition, 34(2), 2001, pp. 299–314, http://dx.doi.org/10.1016/S0031-3203(99)00223-X

[6] Kuncheva, L.: Combining pattern classifiers methods and algorithms. John Wiley & Sons, 2004.

[7] Littlestone, N., Warmuth, M.: The Weighted Majority Algorithm. Inf. Comput., 108(2), 1994, pp. 212–261, http://dx.doi.org/10.1006/inco.1994.1009

[8] Przybyła-Kasperek, M., Wakulicz-Deja, A.: Global decisions taking on the basis of dispersed medical data, Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing, Lecture Notes in Computer Science Volume 8170, 2013, pp. 355-365, http://dx.doi.org/10.1007/978-3-642-41218-9_38

[9] Przybyła-Kasperek, M., Wakulicz-Deja, A.: Global decision-making system with dynamically generated clusters. Information Sciences, 270, 2014, pp. 172–191, http://dx.doi.org/10.1016/j.ins.2014.02.076

[10] Przybyła-Kasperek, M., Wakulicz-Deja, A.: A dispersed decision-making system - The use of negotiations during the dynamic generation of a systems structure. Information Sciences, 288, 2014, pp. 194–219, http://dx.doi.org/10.1016/j.ins.2014.07.032

[11] Przybyła-Kasperek, M.: Global Decisions Taking Process, Including the Stage of Negotiation, on the Basis of Dispersed Medical Data, S. Kozielski et al. (Eds.): BDAS 2014, CCIS Communications in Computer and Information Science 424, 2014, pp. 290–299, http://dx.doi.org/10.1007/978-3-319-06932-6_28

[12] Przybyła-Kasperek, M.: The Borda Count, the Intersection and the Highest Rank Method in a Dispersed Decision-Making System. Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing - 15th International Conference, RSFDGrC 2015, Tianjin, China, November 20-23, 2015, Proceedings, Lecture Notes in Computer Science, 2015, pp. 298–309, http://dx.doi.org/10.1007/978-3-319-25783-9_27

[13] Przybyła-Kasperek, M., Wakulicz-Deja, A.: Global decision-making in multi-agent decision-making system with dynamically generated disjoint clusters. Applied Soft Computing, 40, 2016, pp. 603–615, http://dx.doi.org/10.1016/j.asoc.2015.12.016

[14] Przybyła-Kasperek, M., Wakulicz-Deja, A.: The strength of coalition in a dispersed decision support system with negotiations. European Journal of Operational Research, 2016, pp. 947–968, http://dx.doi.org/10.1016/j.ejor.2016.02.008

[15] Rogova, G. L.: Combining the results of several neural network classifiers. Neural Networks, 7(5), 1994, pp. 777–781, http://dx.doi.org/10.1016/0893-6080(94)90099-X

[16] Tax, D.M.J., Duin, R.P.W., Breukelen, M.: Comparison between product and mean classifier combination rules. In Proc. Workshop on Statistical Pattern Recognition, Prague, Czech, 1997.

[17] Zagorecki, A.: Feature Selection for Naive Bayesian Network Ensemble using Evolutionary Algorithms, Proceedings of the 2014 Federated Conference on Computer Science and Information Systems, 2014, pp. 381–385, http://dx.doi.org/10.15439/2014F498

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 137–140
DOI: 10.15439/2016F232
ACSIS, Vol. 8. ISSN 2300-5963

Hybrid Fuzzy-Genetic Algorithm Applied to Clustering Problem

Krzysztof Pytel
Faculty of Physics and Applied Informatics
University of Lodz, Poland
Email: kpytel@uni.lodz.pl

***Abstract*—Clustering is a task of grouping a set of objects in such a way that objects in the same group (called a cluster) are similar to each other and dissimilar to objects belonging to other groups (clusters). The article presents the idea of the hybrid Fuzzy Logic-Genetic Algorithm (FLGA) system that supports solving clustering problems. The Genetic Algorithm (GA) realizes the process of multi-objective optimization - it aims at optimal distribution of clusters and correctly assigns each object to a cluster. The Fuzzy Logic Controller (FLC) is used for setting the number of clusters. The FLC uses additional fuzzy logic criteria obtained from experts. Experiments show that the proposed algorithm is an efficient tool for the clustering problem. The algorithm can be also used for solving similar optimization problems.**

I. Introduction

CLUSTERING (or cluster analysis) is the problem of classifying an unlabeled set of objects into groups of similar objects, called clusters. Each cluster consists of objects that are similar to one another and dissimilar to objects belonging to other clusters. Clustering is often based on similarity or dissimilarity measure. This measure is problem-dependent. The similarity or dissimilarity between the objects is usually computed, based on the distance between objects. The most popular distance measure is the Euclidean distance, but other measures, such as the Manhattan or Minkowski distances, could also be used. Clustering can be formally considered as a kind of NP-hard grouping problem. The main difficulty in a clustering problem is that it is an unsupervised task, so usually we do not know the number and the distribution of clusters, the shape of clusters or association of objects to clusters. Clustering is not one specific algorithm, but a general task. A classical clustering method is the k-means [1]. A k-means algorithm is sensitive to the choice of an initial partition, and this step can have a significant impact on the performance of algorithm. The algorithm could converge to a local minimum. A k-means algorithm needs determining a number of clusters in all data sets (parameter k), an inappropriate parameter k may yield poor results.

The Genetic Algorithm is an optimization method that simulates the process of natural evolution. They usually search for approximate solutions for composite optimization problems in a large search space. A characteristic feature of genetic algorithms is that in the process of evolution they do not use the knowledge specific for a given problem, except for the fitness function assigned to all individuals. Genetic algorithms can be used for solving wide range of optimization problems.

Clustering can be formulated as a multi-objective optimization problem. It consists of three different objective functions: looking for an appropriate number of centroids, optimal placeing of centroids in a given area, and assigning each object to a cluster, represented by the centroid, to minimize the distance between the objects in the same cluster.

II. Problem formulation

A clustering problem is one of practical examples of multi-objective optimization problems. The clustering problem can be defined as: let us consider a data set $X = \{x_1, x_2, ..., x_N\}$ be a set of N objects. Each object $x_i = [x_{i1}, x_{i2}..., x_{id}]$ has d dimensions. The goal of the clustering algorithm is to find k clusters $C_1, C_2, ..., C_k$ so that objects belonging to the same cluster are more similar to each other than to the objects belonging to other clusters. The Euclidean distance between each pair of objects is typically used as a similarity measure. Clustering must comply assigning constrains: each cluster must contain at least one object, and the object can be assigned to one cluster only.

$$\begin{cases} min \quad f_1(x_1, x_2, ..., x_N) \\ opt \quad f_2(n) \\ min \quad f_3(n, x_1, x_2, ..., x_N) \\ \text{subject to: } assigning\ constraints \end{cases} \tag{1}$$

where: f_1 - is the function representing the distance between the objects assigned to the same cluster,

f_2 - is the function representing the number of clusters,

f_3 - is the function assigning an object to a cluster,

$(x_1, , x_N)$ - are objects in d dimentional space,

$1 \leq n \leq n_{max}$ - is the number of clusters.

More information about cluster analysis can be found in publications [1][2].

III. Proposed Fuzzy Logic-Genetic Algorithm

The clustering problem is discussed in literature, eg. [8][9]. There are a lot of publications concerning different methods of solving this problem, for example k-means or genetic algorithms, but these methods work well if a number of clusters is known before running an algorithm. The proposed Fuzzy Logic-Genetic Algorithm (FLGA) consists of two modules: the Genetic Algorithm (GA) and the Fuzzy Logic Controller (FLC). The GA seeks for an optimal placement of centroids

and assigns objects to clusters. It delivers information concerning a current state of optimization to the FLC after a fixed number of generations. The FLC looks for an optimal number of clusters. The FLC is engaged between generations of the GA in fixed intervals of generations. The FLC modifies the number of clusters in dependence on information delivered from the GA. If the FLC changes the number of clusters, it sends information to the GA and modifies the individuals' genes to comply with the new constraints. The system is able to find the optimal number of clusters simultaneously with an optimization executed by the GA, so we do not need to know the number of clusters before running the algorithm.

In the proposed FLGA the individuals' genes (potential solutions) are encoded by the means of a composite data structure consisting of:

- the table describing the position of centroids by geographical coordinates - coordinates (x, y) of centroids are represented by real numbers, the number of genes in a table is equal to the number of clusters,
- the table describing an association of objects to clusters - association of objects to clusters are coded by integer numbers, eg. number i in position k means the association of object k to cluster i, the number of genes in a table is equal to the number of objects. This method of gene coding ensures assigning every object to one cluster only.

The value of the fitness function of an individual is calculated as a total distance between objects and centroids. Because clustering is a problem of minimization of the distance, we introduced additional constant C to transform the increasing fitness function into the decreasing function optimized. The value of constant C was chosen experimentally for every solved task. In our experiment different types of crossing-over of chromosomes were used. The standard one-point crossing is used when the number of clusters does not change. We introduce two new crossing-over operators, used when the number of clusters changes:

- CR1 - is used when the descendant's length of genotype is greater than the genotype's length of its parents. The descendant's genotype is obtained by copying all the genes from its first parent and lacking the genes of its second parent beginning from the end of the genotype.
- CR2 - is used when the descendant's length of genotype is smaller than the genotype's length of its parents. The number of the genes copied from every parent is diminished in proportion to the genotype's length to obtain the required length of the descendant's genotype.

Genetic Algorithms can be used for solving multi-objective optimization problems. They can be used in hybrid systems with other methods inspired by observation of nature. For example, the Fuzzy Logic Controller can effectively direct the process of evolution in the Genetic Algorithm toward a desired area of the search space [4][5].

A basic task of the FLC in the proposed system is evaluation of the solutions found till now. The FLC uses experts' knowledge and the knowledge collected by the GA and transferred to the FLC in fixed intervals of GA's generations. The FLC optimizes the number of clusters and is engaged in fixed intervals of GA's generations - it makes decisions about the diminution or the enlargement of the clusters' number. The FLC calculates the change of the clusters' number, based on two parameters:

- the relation of the distance between the centroids to the distance between the centroids and the objects assigned to these clusters,

$$rd_1 = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} d(i,j)}{\sum_{i=1}^{n}\sum_{k=1}^{m} d(i,k)} \tag{2}$$

  where:
  - rd_1 - the relation of the distance between the centroids to the distance between the centroids and the objects assigned to these clusters,
  - $\sum_{i=1}^{n}\sum_{j=1}^{n} d(i,j)$ - the distance between the centroids,
  - $\sum_{i=1}^{n}\sum_{k=1}^{m} d(i,k)$ - the distance between the centroids and the objects assigned to these clusters,
  - n - the number of clusters (centroids),
  - m - the number of objects.

  This parameter lets us determine a suitable number of clusters. The low value of this parameter can be due to a small number of clusters with relation to the number of objects. The large value of this parameter can be due to a big number of clusters with relation to the number of objects.
- the relation of the distance between the centroids and the objects assigned to these clusters to the distance between the centroids and the objects not assigned to these clusters,

$$rd_2 = \frac{\sum_{i=1}^{n}\sum_{k=1}^{m1} d(i,k)}{\sum_{i=1}^{n}\sum_{l=1}^{m2} d(i,l)} \tag{3}$$

  where:
  - rd_2 - the relation of the distance between the centroids and the objects assigned to these clusters to the distance between the centroids and the objects not assigned to these clusters,
  - $\sum_{i=1}^{n}\sum_{k=1}^{m1} d(i,k)$ - the distance between centroids and objects assigned to these clusters,
  - $\sum_{i=1}^{n}\sum_{l=1}^{m2} d(i,l)$ - the distance between the centroids and the objects not assigned to these clusters,
  - n - the number of clusters (centroids),
  - $m1$ - the number of the objects assigned to these clusters,
  - $m2$ - the number of the objects not assigned to these clusters,

  This parameter lets us determine a suitable density of objects in the clusters. The low value of this parameter can be due to a big number of clusters with relation to the number of objects. The large value of this parameter can be due to a small number of clusters with relation to the number of objects. The value of this parameter is 0

TABLE I
FUZZY VALUES OF CLUSTERS' NUMBER CHANGE

		rd_1		
		Small	**OK**	**Large**
rd_2	**Large**	Enlarge	Enlarge	Not change
	OK	Enlarge	Not change	Diminish
	Small	Not change	Diminish	Diminish

when every object belongs to its own cluster, the value 1 is when all objects belongs to one cluster only.

The knowledge of experts is expressed by the following rules:

- enlarge the number of clusters if the relation of the distance between the centroids to the distance between the centroids and the objects assigned to these clusters (rd1) is small and the relation of the distance between the centroids and the objects assigned to these clusters to the distance between the centroids and the objects not assigned to these clusters (rd2) is large,
- do not change the number of clusters if the relation of the distance between the centroids to the distance between the centroids and the objects assigned to these clusters (rd1) is suitable and the relation of the distance between the centroids and the objects assigned to these clusters to the distance between the centroids and the objects not assigned to these clusters (rd2) is suitable,
- diminish the number of clusters if the relation of the distance between the centroids to the distance between the centroids and the objects assigned to these clusters (rd1) is large and the relation of the distance between the centroids and the objects assigned to these clusters to the distance between the centroids and the objects not assigned to these clusters (rd2) is small.

As the result from the FLC we accepted:

- signal to enlarge the number of clusters (+1),
- signal to do not change the number of clusters (0),
- signal to diminish the number of clusters (-1).

The knowledge base (rule base) of FLC is shown in Table 1 (fuzzy values of clusters' number change).

Figure 1 show membership functions of the relation of the distance between centroids to the distance between the centroids and the objects assigned to these clusters, the relation of the distance between the centroids and the objects assigned to these clusters to the distance the between centroids and the objects not assigned to these clusters and the value of the clusters' number change respectively. The shape of the membership functions was established experimentally and user can adapt them to solved problem. The FLC uses the center of gravity as a defuzzyfication method. Similar systems were successfully applied to other multiobjective optimalization problems, such as the Connected Facility Location Problem (ConFLP) [6] or the Wireless Access Points Placement Problem (WAPP) [7].

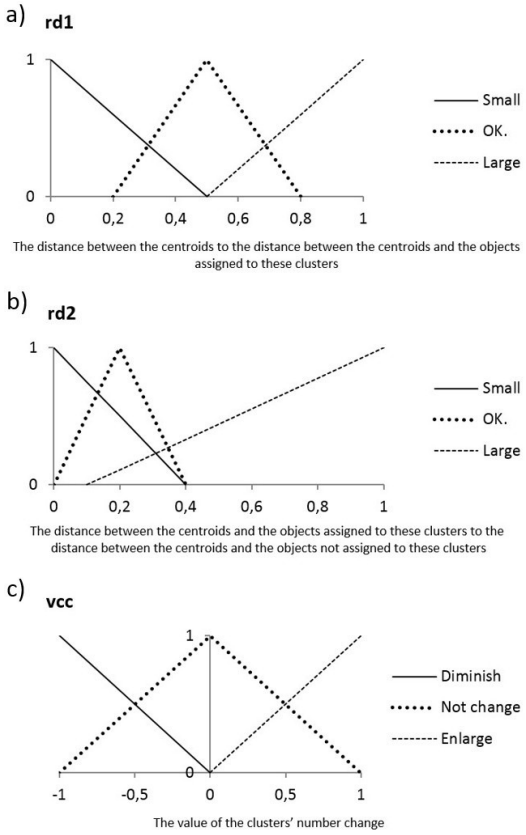


Fig. 1. Membership functions: a) the relation of the distance between the centroids to the distance between the centroids and the objects assigned to these clusters, b) the relation of the distance between the centroids and the objects assigned to these clusters to the distance between the centroids and the objects not assigned to these clusters, c) the value of the clusters' number change

IV. COMPUTATIONAL EXPERIMENT

The goal of our experiments is verification of the idea of the hybrid fuzzy-genetic algorithm to solving a clustering problem. In experiments we verify the ability of the FLC to optimize of the number of clusters, basing on experts' knowledge and data originated in the GA. Optimization of centroids' positions and optimal assigning of objects to clusters is realized by a genetic algorithm. For tests we used a set of data from "The Fundamental Clustering Problems Suite" (FCPS) [12] as a benchmark. We have chosen 4 two-dimensional problems from 400 to 4096 objects and from 2 to 3 clusters. Figure 2 show the distribution of objects in space and known a priori classifications in problems selected for our tests.

All tasks were solved by a k-means algorithm (we used the k-means method from the "rattle" library in R programming language [11]), proposed a hybrid genetic algorithm with the fuzzy logic (FLGA) and the simple genetic algorithm (SGA) - an algorithm proposed in [3], and modified by me to solve a clustering problem. The correct value clusters' number was used in a k-means and the SGA algorithms, the

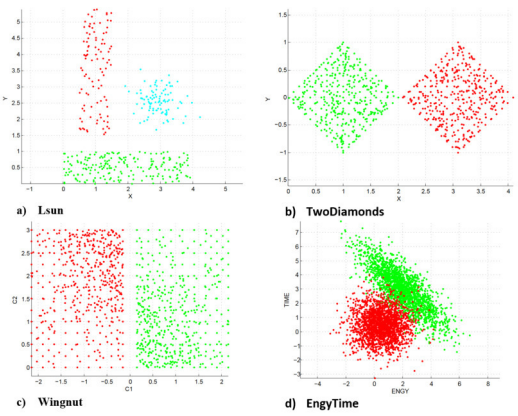


Fig. 2. The distribution of objects in space and known a priori classifications in problems selected for tests: a) Lsun, b) TwoDiamonds, c) Wingnut, d) EngyTime

TABLE II
THE DISTANCE BETWEEN OBJECTS AND CENTROIDS

		The distance between objects and centroids		
The problem Name	**The number of objects**	**k-means**	**SGA**	**FLGA**
Lsun	400	838	924	1112
TwoDiamonds	800	1645	1653	1654
Wingnut	1070	1815	1855	1821
EngyTime	4096	7912	9454	10093

FLGA was started from an incorrect number of clusters. Each algorithm was executed 10 times. In Table 2 and 3 there is the best distance between objects and centroids and the number of correctly assigned objects to clusters obtained by all algorithms. Figure 3 shows the distribution of objects in space obtained by the k-means algorithm and the FLGA algorithm.

V. CONCLUSIONS

The proposed Fuzzy Logic-Genetic Algorithm was able to find a solution near the optimum. However, looking at charts in Figure 3 presenting the distribution of objects after an optimization, it is easy to notice that improvement of this result is still possible. In Lsun task, the assignment of objects to clusters in the FLGA algorithm is more similar to known a priori classification, than in the k-means algorithm.

In all tasks proposed, the FLC correctly qualified the number of clusters.

The time operation of the FLGA on a PC computer did not exceed 10 minutes for the task of optimization of 4096 objects.

TABLE III
THE NUMBER OF CORRECTLY ASSIGNED OBJECTS

		The number of correctly assigned objects		
The problem Name	**The number of objects**	**k-means**	**SGA**	**FLGA**
Lsun	400	391	187	324
TwoDiamonds	800	800	567	767
Wingnut	1070	981	676	913
EngyTime	4096	4010	2128	2945

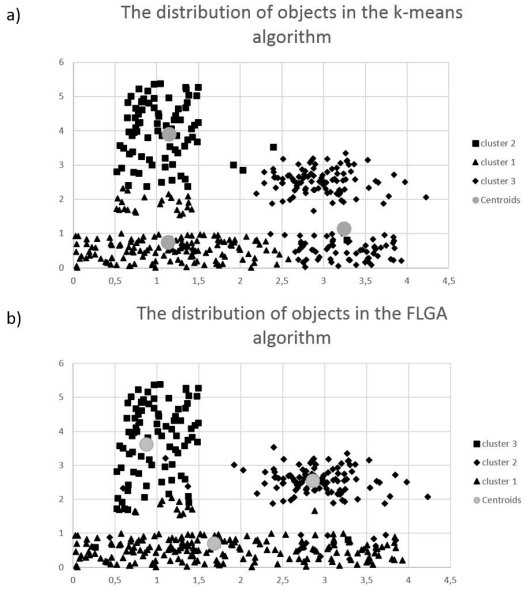


Fig. 3. The distribution of objects in space obtained by: a) the k-means algorithm, b) the FLGA algorithm

In tasks with a large number of objects, the time of calculations can be considerably longer. The parameters of an algorithm, eg. the number of generations, can be changed to fulfil the users' needs and reach a required accuracy of calculations.

The proposed algorithm is an efficient tool for solving clustering problems, where the number of clusters cannot be pre-determined. The proposed algorithm can be used for solving similar problems of multi-objective optimization.

REFERENCES

[1] Berkhin, P., "Survey of clustering data mining techniques." *Technical report*, Accrue Software, San Jose, CA, 2002.
[2] Maimon, O., Rokach, L., "Data Mining and Knowledge Discovery Handbook", Springer. DOI: 10.1007/978-0-387-09823-4
[3] Michalewicz Z., "Genetic Algorithms + Data Structures = Evolution Programs", *Springer Verlag*, Berlin (1992).
[4] Pytel K., Nawarycz T., "Analysis of the Distribution of Individuals in Modified Genetic Algorithms" *[in] Rutkowski L., Scherer R., Tadeusiewicz R., Zadeh L., Zurada J., Artificial Intelligence and Soft Computing*, Springer-Verlag Berlin Heidelberg (2010).
[5] Pytel K., "The Fuzzy Genetic Strategy for Multiobjective Optimization", *Proceedings of the Federated Conference on Computer Science and Information Systems*, Szczecin, (2011).
[6] Pytel, K, Nawarycz, T., "A Fuzzy-Genetic System for ConFLP Problem", *Advances in Decision Sciences and Future Studies, Vol. 2*, Progress & Business Publishers, Krakow 2013.
[7] Pytel, K., Nawarycz, T. "The Fuzzy-Genetic System for Multiobjective Optimization", *[in] Rutkowski L., Korytkowski M, Scherer R., Tadeusiewicz R., Zadeh L., Zurada J., Swarm and Evolutionary Computation*, Springer-Verlag Berlin Heidelberg 2012.
[8] Tan, P. N., Steinbach, M., Kumar, V., "Introduction to Data Mining" *Parson*, 2006
[9] Jiang, D., Tang, C., Zhang, A., "Cluster analysis for gene expression data: a survey", *IEEE Transactions on Knowledge and Data Engineering (Volume:16 , Issue: 11. pp. 1370 - 1386*, 2004
[10] Zitzler E., "Evolutionary Algorithms for Multiobjective Optimization: Methods and Applications", Zurich (1999).
[11] Rattle: A Graphical User Interface for Data Mining using R http://rattle.togaware.com/
[12] The Fundamental Clustering Problems Suite https://www.uni-marburg.de/fb12/datenbionik/data/

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 141–145
DOI: 10.15439/2016F183
ACSIS, Vol. 8. ISSN 2300-5963

Deep Evolving GMDH-SVM-Neural Network and its Learning for Data Mining Tasks

Galina Setlak
Rzeszow University of Technology
12 Al. Powstancow Warszawy, 35-959, Rzeszow, Poland
Email: gsetlak@prz.edu.pl

Yevgeniy Bodyanskiy
Kharkiv National Univesity of Radio Electronics
14 Nauky Ave., Kharkiv, 61166, Ukraine
Email: yevgeniy.bodyanskiy@nure.ua

Olena Vynokurova
Kharkiv National Univesity of Radio Electronics
14 Nauky Ave., Kharkiv, 61166, Ukraine
Email: olena.vynokurova@nure.ua

Iryna Pliss
Kharkiv National Univesity of Radio Electronics
14 Nauky Ave., Kharkiv, 61166, Ukraine
Email: iryna.pliss@nure.ua

***Abstract*—In the paper, the deep evolving neural network and its learning algorithms (in batch and on-line mode) are proposed. The deep evolving neural network's architecture is developed based on GMDH approach (in J. Schmidhuber's opinion it is historicaly first system, which realizes deep learning) and least squares support vector machines with fixed number of the synaptic weights, which provide high quality of approximation in addition to the simlicity of implementation of nodes with two inputs. The proposed system is simple in computational implementation, characterized by high learning speed and allows processing of data, which are sequentially fed in on-line mode. The proposed system can be used for solving a wide class of Dynamic Data Mining tasks, which are connected with non-stationary, nonlinear stochastic and chaotic signals. The computational experiments are confirmed the effectiveness of the developed approach.**

I. Introduction

Nowadays, artificial neural networks (ANNs) are widely used for solving a lot of Data Mining tasks. In these tasks initial information is presented in the form of both "object-properties" table and multivariate time series, which are generated by stochastic or chaotic nonstationary nonlinear objects. The advantages of these computational intelligence systems are, first of all, their universal approximation properties and learning abilities using real experimental data [1], [2].

Recent years Computational Intelligence specialists are interested in deep neural networks (DNN) [3], [4], [5], [6], [7]. The deep neural networks comparatively with conventional ANNs, also called shallow neural networks (SNNs), provide much higher quality of information processing. However, these networks are essentially tedious with relation to computational implementation, also are subjected to overfitting in case of short training samples and demand of high operation time and computational resources especially when operated with Big Data [8]. Both the standard neurons (which form set of layers) and shallow neural networks can be used as the basic elements of deep neural networks.

One of the most effective representatives of the shallow neural networks are support vector machines (SVM) [9], [10], [11], [12], [13]. The tuning of support vector machines is provided by using both lazy learning (the activation functions' centers tuning) and optimization procedures (the synaptic weights tuning). However, if the process of the lazy learning is implemented immediately, then optimization tasks solving using support vector machines with big training set is enough complex. In this connection deep neural networks, which are implemented using support vector machines [14], [15], [16], providing high quality of information processing are essentially tedious from the computational point of view.

It can be noticed, that tuning process of support vector machines can be essentially speed up if the least-squares support vector machines (LS-SVM) [17] are used instead of a conventional approach. The learning process of least-squares support vector machines reduces to a solution of the set of Karush-Kuhn-Tucker equations and the result of this learning can be written in an analytical form.

Among a great number of possible deep neural networks' architectures, the deep networks based on GMDH are one of the most effective networks, as it was mentioned in [6]. These networks are based on the group method of data handling [18], [19], [20], which allows automatically increasing a number of layers for information processing to achieve the required accuracy of the results. A combination of the GMDH approach with ANNs have led to synthesis of wide range of computational intelligence systems [21], [22], [23], [24], [25], [26], [27], [28] where different type of artificial neurons are used as nodes.

In this case unlimited increasing of layers in the sistem (using the GMDH paradigm) and simplicity of learning LS-SVM with two inputs (using their universal approximation properties) allow to efficiently information processing in on-line mode of the deep learning.

In the connection with mentioned above, it seems appropriate to develop deep neural networks' architecture based on GMDH and LS-SVM and its learning algorithm. The proposed approach is characterized by simplicity of a computational implementation and high speed learning for the solution of

978-83-60810-90-3/$25.00

wide range of Data Mining tasks, which are described by both short and large volume data set.

II. The architecture of deep evolving GMDH-SVM-neural network

An architecture of the proposed deep evolving GMDH-SVM-neural network is shown in Fig.1.

A $(n \times 1)$-dimensional vector of the input signals $x = (x_1, x_2, \ldots, x_n)^T \in R^n$ is fed to the zero (receptive) layer of the GMDH-SVM-neural network. Further, this input vector is passed to the first layer, which consists of $n_1 = C_n^2 = 0.5n(n-1)$ nodes that are the conventional LS-SVM with two inputs. It is obvious that learning process of the LS-SVM with two inputs has not any problem both in relation of a computational implementation and with regard for requirements to a volume of training set. The output signals $\hat{y}_l^{[1]}$ $(l = 1, 2, \ldots, n_1)$ are formed by the nodes' outputs of the $SVM^{[1]}$ of first hidden layer.

Further these signals are fed to the selection block $SB^{[1]}$ of the first hidden layer. This selection block SB^1 selects $n_1^*(n_1^* \leq n_1)$ signals from a range of signals $\hat{y}_l^{[1]}$. The selected signals are the best in the sense of accepted criterion, in more cases it is the mean square error $\sigma^2_{\hat{y}_l^{[1]}}$, but any other accuracy criterion can be used in relation to reference signal $y(k)$ $(k = 1, 2, \ldots, N$ is an observation number in a training set or a current discrete time, when a learning process and data processing take place in on-line mode).

From the n_1^* best output's signals of the first hidden layer $\hat{y}_l^{[1]*}$ (using the conventional GMDH approach) are formed $n_2(n \leq n_2 \leq 2n)$ pairwise combination of signals $\hat{y}_l^{[1]*}$, $\hat{y}_p^{[1]*}$, which are fed to the inputs of the second hidden layer. The second hidden layer is formed by $SVM^{[2]}$ nodes, which are similar to the elements of the first hidden layer.

From the output's signals $\hat{y}_l^{[2]}$ of this layer, the selection block $SB^{[2]}$ of the second hidden layer selects only that signals $\hat{y}_l^{[2]*}$, whose accuracy is better than the best signal $\hat{y}_l^{[1]*}$ of the first hidden layer. The third hidden layer with selection block $SB^{[3]}$ forms the signals, which have accuracy better than the best signal $\hat{y}_l^{[2]*}$ of the second layer.

In such way, the network's architecture is formed during learning process likewise evolving computational intelligence systems [29], [30].

The architecture's evolution process takes place until the selection block $SB^{[n-1]}$ forms only two signals $\hat{y}_1^{[s-1]*}$ and $\hat{y}_2^{[s-1]*}$ in its output. Just these two signals are fed to the single output nodes of $SVM^{[3]}$ where the output system's signal $\hat{y}^{[s]}$ is computed.

III. The learning of deep evolving GMDH-SVM-neural network

As it was previously noted an each node of the proposed system is the LS-SVM with single output and two inputs. Hence, two-dimensional vector $x_{ij}(k) = (x_i(k), x_j(k))^T$ $(i = 1, 2, \ldots, n; j = 1, 2, \ldots, n; i \neq j)$ is fed to the input of the first hidden layer and output of the each node is a scalar signal $\hat{y}_l^{[1]}$ $(l = 1, 2, \ldots, n_1)$. Therefore, the LS-SVM is the hybrid system, which combines a learning based on both an optimization and a memory [1], [2], [6], [7], [17], and implements minimization of an empirical risk criterion. It is necessary to notice, that the SVMs are the most effective under short data set conditions and are not subject to an overfitting and proved a high quality of approximation.

The mapping, which implements standard LS-SVM SNN, can be written in the form

$$\hat{y}_l^{[1]} = (w_l^{[1]})^T \varphi_l^{[1]}(x) + w_{0l}^{[1]} \tag{1}$$

where $w_l^{[1]} = (w_{1l}^{[1]}, w_{2l}^{[1]}, \ldots, w_{Nl}^{[1]})^T$, $\varphi_l^{[1]}(x) = (\varphi_{1l}^{[1]}, \varphi_{2l}^{[1]}, \ldots, \varphi_{Nl}^{[1]})^T$. The learning process reduces to setting the centers of activation functions (usually Gaussians) in the point, which are determined by a training sample x_{ij} $(k = 1, 2, \ldots, N)$ and minimization of squared criterion simultaneously in the form

$$E_l^{[1]}(N) = \frac{1}{2}\|w_l^{[1]}\|^2 + \frac{\gamma}{2}\sum_{k=1}^{N}(e_l^{[1]}(k))^2 \tag{2}$$

in the presence of N equality-constraints in the form

$$\begin{cases} y(1) & = (w_l^{[1]})^T \varphi_l^{[1]}(x_{ij}(1)) + w_{0l}^{[1]} + e_l^{[1]}(1), \\ \vdots & \\ y(N) & = (w_l^{[1]})^T \varphi_l^{[1]}(x_{ij}(N)) + w_{0l}^{[1]} + e_l^{[1]}(N) \end{cases} \tag{3}$$

where $\gamma > 0$ is a regularization parameter and

$$\begin{aligned} e_l^{[1]}(k) &= y(k) - \hat{y}_l^{[1]}(k) = \\ &= y(k) - (w_l^{[1]})^T \varphi_l^{[1]}(x_{ij}(k)) - w_{0l}^{[1]}. \end{aligned} \tag{4}$$

In this way, LS-SVM learning task is reduced to finding of a saddle point of Lagrange function

$$\begin{aligned} L_l^{[1]}(w_l^{[1]}, w_{0l}^{[1]}, e_l^{[1]}, \lambda_l^{[1]}(k)) &= E_l^{[1]}(N) + \\ &+ \sum_{k=1}^{N} \lambda_l^{[1]}(k)\Big(y(k) - (w_l^{[1]})^T \varphi_l^{[1]}(x_{ij}(k)) - \\ &- w_{0l}^{[1]} - e_l^{[1]}(k)\Big). \end{aligned} \tag{5}$$

This saddle point can be found by solving the Karush-Kuhn-Tucker equations set. In this case, besides $N + 1$ synaptic weights $w_l^{[1]}$, $w_{0l}^{[1]}$ the N indefinite Langrange multipliers $\lambda_l^{[1]}(k)$ have to be found.

The main disadvantage of SVM in the system under consideration is necessity of adding new synaptic weights in each nodes with rising of a training set volume. Therefore, if it is necessary to process Big Data than proposed system becomes too tedious. To avoid this problem it is possible by limiting a number of synaptic weights in each node by using, so-called, "sliding window" data processing. Such "sliding window" contains only h last observations.

Introducing the Lagrange function with "sliding window" instead of the expression (5) in the form

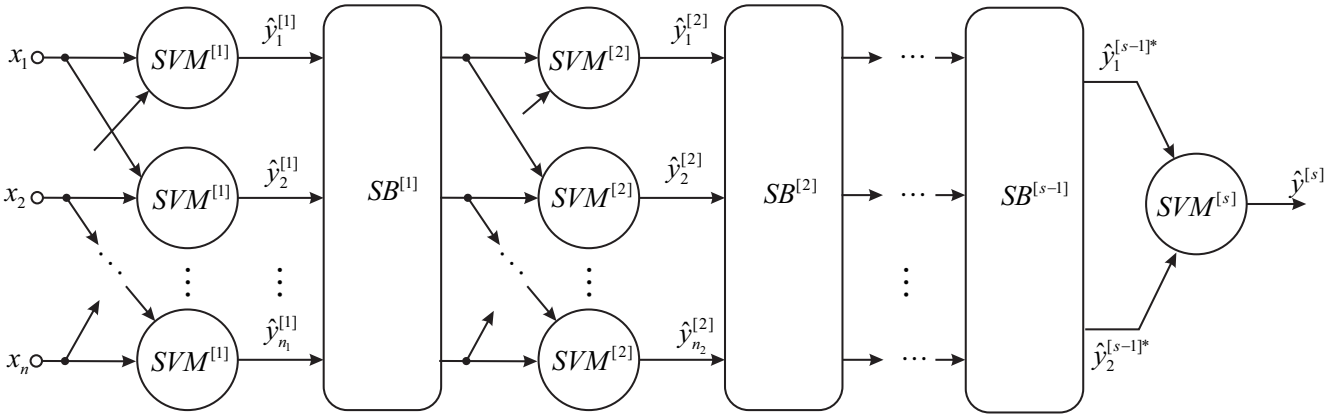


Fig. 1. The architecture of proposed deep evolving GMDH-SVM-neural network

$$
\begin{aligned}
&L_l^{[1]}(w_l^{[1]}, w_{0l}^{[1]}, e_l^{[1]}, \lambda_l^{[1]}(k), h) = \\
&= \frac{1}{2}\,\|w_l^{[1]}\|^2 + \frac{\gamma}{2}\sum_{\tau=k-h+1}^{k}(e_l^{[1]}(\tau))^2 + \\
&+ \sum_{\tau=k-h+1}^{k}\lambda_l^{[1]}(\tau)\Big(y(\tau) - (w_l^{[1]})^T\varphi_l^{[1]}(x_{ij}(\tau)) - \\
&- w_{0l}^{[1]} - e_l^{[1]}(\tau)\Big)
\end{aligned}
\tag{6}
$$

and solving Karush-Kuhn-Tucker equations set, we can write the result in the form

$$
\begin{pmatrix} 0 & I_h^T \\ I_h & \Omega(k) + \gamma^{-1} I_{hh} \end{pmatrix}
\begin{pmatrix} w_{0l}^{[1]} \\ \Lambda(k) \end{pmatrix} =
\begin{pmatrix} 0 \\ Y(k) \end{pmatrix} \tag{7}
$$

where $\Lambda(k) = (\lambda(k-h+1), \ldots, \lambda(k))^T$, I_h is a $(h \times 1)$ unity vector, I_{hh} is a $(h \times h)$ unity matrix, $Y(k) = (y(k-h+1), \ldots, y(k))$, $\Omega(k) = \{\Omega_{\tau l} = \varphi_l^{[1]T}(x_{ij}(\tau))\varphi_l^T(x_{ij}(t)) = K(x_{ij}(\tau), x_{ij}(t))\}$, $\tau = k-h+1, \ldots, k$, $t = k-h+1, \ldots, k$, $K(x_{ij}(\tau), x_{ij}(t))$ is the kernel function, which is satisfied to the conditions of Mercer theorem [17], and usually it is Gaussian function in the form

$$
K(x_{ij}(\tau), x_{ij}(t)) = \exp\left(-\frac{\|x_{ij}(\tau) - x_{ij}(t)\|^2}{2\sigma^2}\right). \tag{8}
$$

In this case, the transformation (1), which is implemented by the LS-SVM node, can be rewritten in the form

$$
\hat{y}_l^{[1]} = \sum_{\tau=k-h+1}^{k}\lambda_l^{[1]}(\tau)K(x_{ij}(\tau), x_{ij}(t)) + w_{0l}^{[1]} \tag{9}
$$

where parameters $\lambda_j^{[1]}(\tau)$, $w_{0l}^{[1]}$ can be defined from the system (7) in the form

$$
\begin{aligned}
&\begin{pmatrix} w_{0l}^{[1]} \\ \Lambda(k) \end{pmatrix} = \\
&= \begin{pmatrix} 0 & I_h^T \\ I_h & \Omega(k) + \gamma^{-1} I_{hh} \end{pmatrix}^{-1}
\begin{pmatrix} 0 \\ Y(k) \end{pmatrix}.
\end{aligned}
\tag{10}
$$

As a result, the learning of nodes in the proposed system reduces to a solving of linear equations set with fixed number of variables. It should be noticed, the "sliding window" learning allows managing the deep neural network tuning process, when information is fed to the system's input in on-line mode in the form of data stream. The nodes of second and next hidden layers are tuned absolutely likewise the expression (10).

IV. Simulation results

A. Identification of the mechanical system signal

Efficiency of proposed deep evolving GMDH-SVM-neural network was examined based on different benchmark data including the identification task using real data from mechanical system. Data is taken from a flexible robot arm. The arm is installed on an electrical motor. [We are grateful to Hendrik Van Brussel and Jan Swevers of the laboratory of Production Manufacturing and Automation of the Katholieke Universiteit Leuven, who provided us with these data, which were obtained in the framework of the Belgian Programme on Interuniversity Attraction Poles (IUAP-nr.50) initiated by the Belgian State - Prime Minister's Office - Science Policy Programming. http://homes.esat.kuleuven.be/ smc/daisy/daisydata.html].

Inputs number of deep evolving GMDH-SVM-neural network were taken as $n = 5$, that for input vector in the form $u(k-2), y(k-2), u(k-1), y(k-1), u(k)$ for identification value $y(k)$ where u is a reaction torque of the structure, y is an accelaration of the flexible arm . Node of deep evolving GMDH-SVM-neural network was training by proposed learning algorithm during 100 iterations. Initial parameters values of kernel functions were taken $\sigma = 0.1$. After 100 iterations the training process was stopped, and the next 50 points for $k = 101 \ldots 150$ we have used as the testing data

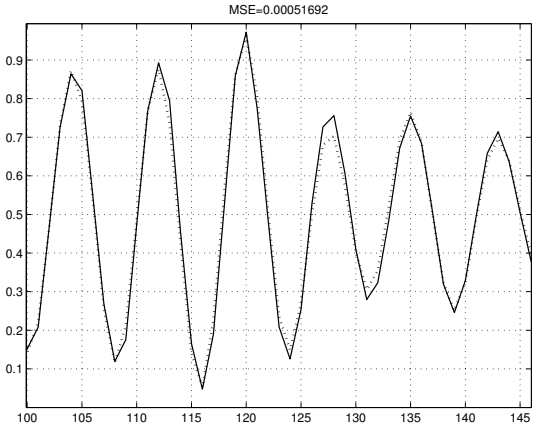


Fig. 2. Results of signal identification

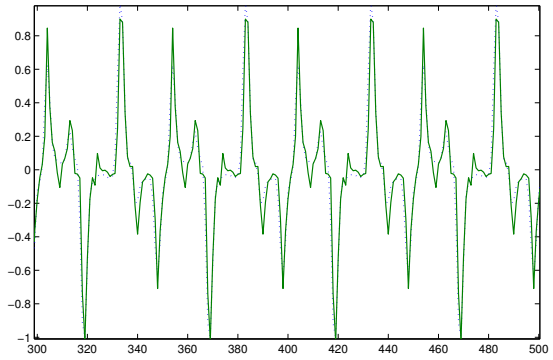


Fig. 3. Results of non-linear non-stationary system identification

set to compute forecast. Initial values of synaptic weights were taken equal to 0. As the quality criterion of forecasting root mean squared error (MSE) was used. Fig.2 shows the results of signal identification. The two curves, representing the actual (dot line) and identification (solid line) values, are almost indistinguishable.

Table I cotains comparative analysis of the signal identification based on different approaches.

Thus as it can be seen from experimental results the proposed approach provides the best quality of prediction in comparison with conventional GMDH-neural networks.

B. The identification of nonlinear nonstationary signal

Simulation of deep evolving cascaded GMDH-SVM-neural network was performed in the process of identification of nonlinear signal, which is described by equation in the form [31]

$$y(k+1) = \frac{y(k)}{1+y^2(k)} + u^3(k) \tag{11}$$

where $u(k) = sin(2\pi k/25) + sin(2\pi k/10)$ is control signal.

The inputs number of evolving cascaded GMDH-neural network were taken as $n = 4$, which correspond to the input vector $x(k-3), x(k-2), x(k-1), x(k)$ for the value $x(k+1)$.

LS-SVM-neuron was trained based on proposed procedures for 400 iterations (400 training samples for $k = 1 \ldots 400$). After 400 iterations the training process was stopped, and the next 100 points for $k = 401 \ldots 500$ we have used as the testing data set to compute signal value. Initial values of synaptic weights were taken equal to 0. As the identification quality criterion mean squared error (MSE) was used.

Fig. 3 shows the results of signal identification. The two curves, representing the actual (dot line) and identification (solid line) values, are very close. Table II shows the comparative analysis of nonlinear non-stationary signal identification based on different approaches.

V. CONCLUSIONS

In the paper, the deep evolving neural network and its learning algorithms are proposed. The architecture of the deep evolving neural network is developed based on GMDH and least squares support vector machines with fixed number of the synaptic weights are used as nodes. The proposed system is simple in computational implementation, characterized by high learning speed and allows processing of data, which are fed sequentially in on-line mode. The combining, in the context of the common deep learning system, the GMDH paradigm with unlimited increasing of the layers number and LS-SVM nodes with fixed sinaptic weights number allow to predetermine an on-line deep learning in Dynamic Data Mining tasks. The computational experiments are confirmed the effectiveness of developed approach.

REFERENCES

[1] S. Haykin, *Neural Networks and Learning Machines*. Upper Saddle River, New Jersey: Pearson, Prentice Hall, 2009.

[2] K.-L. Du and M. Swamy, *Neural Networks and Statistical Learning*. Springer-Verlag London, 2014. [Online]. Available: http://dx.doi.org/10.1007/978-1-4471-5571-3

[3] G. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Computation*, vol. 18, no. 7, pp. 1527–1554, May 2006. doi: 10.1162/neco.2006.18.7.1527. [Online]. Available: http://dx.doi.org/10.1162/neco.2006.18.7.1527

[4] I. Arel, D. Rose, and T. Karnowski, "Deep machine learning - a new frontier in artificial intelligence research," *IEEE Computational Intelligence Magazine*, vol. 5, no. 4, pp. 13–18, Nov. 2010. doi: 10.1109/MCI.2010.938364. [Online]. Available: http://dx.doi.org/10.1109/MCI.2010.938364

[5] Y. Bengio, Y. LeCun, and G. Hinton, "Deep learning," *Nature*, no. 521, pp. 436–444, May 2015. doi: 10.1038/nature14539. [Online]. Available: http://dx.doi.org/10.1038/nature14539

[6] J. Schmidhuber, "Deep learning in neural networks: an overview," *Neural Networks*, no. 61, pp. 85–117, Jan. 2015. doi: 10.1016/j.neunet.2014.09.003. [Online]. Available: http://dx.doi.org/10.1016/j.neunet.2014.09.003

[7] J.Kacprzyk and W. Pedrycz, Eds., *Springer Handbook of Computational Intelligence*. Springer-Verlag Berlin Heidelberg, 2015. [Online]. Available: http://dx.doi.org/10.1007/978-3-662-43505-2

[8] W.Pedrycz and S.-M. Chen, Eds., *Information Granularity, Big Data, and Computational Intelligence*. Springer International Publishing Switzerland, 2015. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-08254-7

TABLE I
THE COMPARATIVE ANALYSIS OF THE SIGNAL IDENTIFICATION

Neural network /Learning algorithm	Number of layers	Number of inputs into nodes	MSE
Deep evolving GMDH-SVM-neural network/Proposed learning algorithm	3	2	0.00051
Hybrid multilayer GMDH-neural network [28] / Proposed learning algorithm	3	3	0.00098
Hybrid multilayer GMDH-neural network [28] / Proposed learning algorithm	3	2	0.00102
GMDH-neural network / Recurrent least squares method	3	2	0.0563

TABLE II
THE COMPARATIVE ANALYSIS OF SIGNAL IDENTIFICATION RESULTS

Neural network	Number of layers	MSE
Deep evolving GMDH-SVM-neural network	3	0.00018
Hybrid multilayer GMDH-neural network [28]	3	0.00123
Evolving cascaded GMDH neural network based on W-neurons	3	0.0009
Evolving cascaded GMDH neural network based on R-neurons with Gaussian functions	3	0.0023
Evolving cascaded GMDH neural network based on R-neurons with Epanechnikov functions	3	0.0015
Multilayer GMDH-neural network with polynomial functions	4	0.0201

[9] V. Vapnik, *The Nature of Statistical Learning Theory*. Springer-Verlag New York, 2000. [Online]. Available: http://dx.doi.org/10.1007/978-1-4757-3264-1

[10] C. Cortes and V. Vapnik, "Support vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, Sep. 1995. doi: 10.1023/A:1022627411411. [Online]. Available: http://dx.doi.org/10.1023/A:1022627411411

[11] M. Grochowina and L. Leniowska, "Comparison of svm and k-nn classifiers in the estimation of the state of the arteriovenous fistula problem," in *Proceedings of the 2015 Federated Conference on Computer Science and Information Systems*, ser. Annals of Computer Science and Information Systems, M. Ganzha, L. Maciaszek, and M. Paprzycki, Eds., vol. 5. IEEE, 2015. doi: 10.15439/2015F194 pp. 249–254. [Online]. Available: http://dx.doi.org/10.15439/2015F194

[12] B. Krawczyk, "Combining one-class support vector machines for microarray classification," in *Proceedings of the 2013 Federated Conference on Computer Science and Information Systems*, M. P. M. Ganzha, L. Maciaszek, Ed. IEEE, 2013, pp. pages 83–89.

[13] M. Ochab and W. Wajs, "Bronchopulmonary dysplasia prediction using support vector machine and libsvm," in *Proceedings of the 2014 Federated Conference on Computer Science and Information Systems*, ser. Annals of Computer Science and Information Systems, M. P. M. Ganzha, L. Maciaszek, Ed., vol. 2. IEEE, 2014. doi: 10.15439/2014F111 pp. pages 201–208. [Online]. Available: http://dx.doi.org/10.15439/2014F111

[14] S. Kim, , S. Kavuri, and M. Lee, *Neural Information Processing: 20th International Conference, ICONIP 2013, Daegu, Korea, November 3-7, 2013. Proceedings, Part I*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, ch. Deep Network with Support Vector Machines, pp. 458–465. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-42054-2_57

[15] S. Kim, Y. Choi, and M. Lee, "Deep learning with support vector data description," *Neurocomputing*, vol. 165, no. 1-2, pp. 111–117, Oct. 2015. doi: 10.1016/j.neucom.2014.09.086. [Online]. Available: http://dx.doi.org/10.1016/j.neucom.2014.09.086

[16] S. Kim, Z. Yu, R. Kil, and M. Lee, "Deep learning of support vector machines with class probability output networks," *Neural Networks*, vol. 64, pp. 19–28, Apr. 2015. doi: 10.1016/j.neunet.2014.09.007. [Online]. Available: http://dx.doi.org/10.1016/j.neunet.2014.09.007

[17] J. Suykens, T. Gestel, J. Brabanter, B. Moor, and J. Vanderwalle, *Least Squares Support Vector Machines*. World Scientific, 2002. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-08254-7

[18] A. Ivakhnenko, "The group method of data handling - a rival of the method of stochastic approximation," *Soviet Automatic Control*, vol. 13, no. 3, pp. 43–55, 1968.

[19] ——, "The group method of data handling - a rival of the method of stochastic approximation," *Automatica*, vol. 6, no. 2, pp. 207–219, 1970.

[20] ——, "Polynomial theory of complex systems," *IEEE Transaction on Systems, Man, Cybernetics*, vol. 1, no. 4, pp. 364–378, Oct. 1971. doi: 10.1109/TSMC.1971.4308320. [Online]. Available: http://dx.doi.org/10.1109/TSMC.1971.4308320

[21] A. Ivakhnenko, G. Ivakhnenko, and A. Muller, "Self-organization of the neural networks with active neurons," *Pattern Recognition and Image Analysis*, vol. 4, no. 2, pp. 177–178, 1994.

[22] A. Ivakhnenko, "Self-organization of neuro net with active neurons for effects of nuclear test explosions forecasting," *System Analysis Modeling Simulation*, vol. 20, no. 1-2, pp. 107–116, 1995.

[23] T. Kondo, "Gmdh neural network algorithm using the heuristic self-organization method and its application to the pattern identification problem," in *Proceedings of the 37th SICE Annual Conference, Japan, Tokyo*, Jul. 1998. doi: 10.1109/SICE.1998.742993 pp. 1143–1148. [Online]. Available: http://dx.doi.org/10.1109/SICE.1998.742993

[24] ——, "Identification of radial basis function networks by using revised gmdh-type neural networks with a feedback loop," in *Proceedings of the 41st SICE Annual Conference, Japan, Osaka*, vol. 5, Aug. 2002. doi: 10.1109/SICE.2002.1195514 pp. 2882–2887. [Online]. Available: http://dx.doi.org/10.1109/SICE.2002.1195514

[25] Y. Bodyanskiy, I. Pliss, and O. Vynokurova, "Hybrid gmdh-neural network of computational intelligence," in *Proceedings of 3rd International Workshop on Inductive Modelling, Poland, Krynica*, Sep. 2009, pp. 100–107.

[26] Y. Bodyanskiy, Y. Zaychenko, E. Pavlikovskaya, M. Samarina, and Y. Viktorov, "The neo-fuzzy neural network structure optimization using the gmdh for the solving forecasting and classification problems," in *Proceedings of 3rd International Workshop on Inductive Modelling, Poland, Krynica*, Sep. 2009, pp. 77–89.

[27] Y. Bodyanskiy, O. Vynokurova, and N. Teslenko, "Cascade gmdh-wavelet-neuro-fuzzy network," in *Proceedings of International Workshop Inductive Modelling, Ukraine, Kyiv*, Sep. 2011, pp. 22–30.

[28] Y. Bodyanskiy, O. Vynokurova, A. Dolotov, and O. Kharchenko, "Wavelet-neuro-fuzzy-network structure optimization using gmdh for the solving forecasting tasks," in *Proceedings of International Workshop Inductive Modelling, Ukraine, Kyiv*, Sep. 2013, pp. 61–67.

[29] N. Kasabov, *Evolving Connectionist Systems*. Springer-Verlag London, 2007. [Online]. Available: http://dx.doi.org/10.1007/978-1-84628-347-5

[30] P. Angelov, D. Filev, and N. Kasabov, *Evolving Intelligent Systems: Methodology and Applications*. John Willey and Sons, 2010.

[31] K. Narendra and K. Parthasarathy, "Identification and control of dynamical systems using neural networks," *IEEE Transaction on Neural Networks*, vol. 1, no. 1, pp. 4–26, Mar. 1990. doi: 10.1109/72.80202. [Online]. Available: http://dx.doi.org/10.1109/72.80202

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 147–152
DOI: 10.15439/2016F558
ACSIS, Vol. 8. ISSN 2300-5963

Analysis of time-frequency representations for musical onset detection with convolutional neural network

Bartłomiej Stasiak
Institute of Information Technology,
Lodz University of Technology,
ul. Wólczańska 215, 90-924
Poland
Email: bartlomiej.stasiak@p.lodz.pl

Jędrzej Mońko
Institute of Information Technology,
Lodz University of Technology,
ul. Wólczańska 215, 90-924
Poland
Email: render@widzew.net

***Abstract*—In this paper a convolutional neural network is applied to the problem of note onset detection in audio recordings. Two time-frequency representations are analysed, showing the superiority of standard spectrogram over enhanced autocorrelation (EAC) used as the input to the convolutional network. Experimental evaluation is based on a dataset containing 10,939 annotated onsets, with total duration of the audio recordings of over 45 min.**

I. Introduction

Onset detection is a well recognized and important problem in automatic music information retrieval. It directly addresses one of the most fundamental aspects of music – the time flow and novelty detection; abstracting from *what* and *how*, it concentrates on the *when* question and tries to answer it as precisely as possible. Interesting on its own, this problem is also fundamental in the analysis of many higher-level concepts, such as *rhytm*, *meter* or *tempo* [1]. In a broader context, sound attack analysis can also support other audio processing tasks, including i.a. audio to score alignment (score following), query-by-humming melody search, singing voice quality evaluation and speech analysis [2][3][4][5][6][7].

While trivial when looking at the musical score, note onset detection appears surprisingly complex when musical recordings with real instruments are considered, with all kinds of phenomena and effects like vibrato, glissando, varied dynamics, embouchure and articulation types, etc. As the result, the precise definition of the *onset time*, enabling to unambiguously locate it on the time axis may be difficult [8][1]. Various definitions, including Perceptual Onset Time (POT), Perceptual Attack Time (PAT), Acoustic Onset Time (AOT) and Note Onset Time (NOT) have been proposed [9][1] in order to highlight differences between the time when the onset is perceivable by a human listener, when it is measurable in the signal or when e.g. the *note-on* command is triggered by a MIDI synthesizer [10]. Presence of vibrato, glissandi or ornamentation, not to mention impulse noise or other distortions in low-quality recordings, may in fact render the problem ill-posed, which makes us resort to machine learning approaches for example-based definition of what an *onset* actually is.

This work was supported by the Faculty of Technical Physics, Information Technology and Applied Mathematics, Lodz University of Technology

II. Previous work

The classical approach to note onset detection is based on the *onset detection function* (ODF) constructed to detect novel events in the sound signal [8][11][12][13]. Typically, the signal waveform $x(t)$ is first split into a series of consecutive, usually overlapping time frames $x_n(t)$ with a windowing function applied to each of them:

$$x_n(t) = x(t)w(St - nh) \tag{1}$$

where $w(St - nh)$ is a windowing function stretched by a factor of frame size S and shifted by an integer, n-th multiple of the hop-size h between the consecutive frames. Discrete Fourier transform (DFT) is then computed, and the ODF construction may be based on either its magnitude spectrum [8], the phase spectrum [11] or both [12]. Obviously, the difference between the consecutive frames is considered, such as in the following simple example (ODF based on the *spectral flux* [1][14]):

$$ODF_{sf}(n) = \sum_k H(|X_k(n)| - |X_k(n-1)|) \tag{2}$$

where

$$H(x) = \frac{x + |x|}{2} \tag{3}$$

$$X_k(n) = \mathrm{DFT}(x_n(t))(k) \tag{4}$$

and where the half-wave rectifier function H is used to consider only positive differences, indicating new spectral components appearing in the signal. The onsets may be then easily detected by thresholding the ODF with a fixed threshold T or – more frequently – with a threshold based on moving mean or moving median.

978-83-60810-90-3/$25.00

It should be noted that some onsets (e.g. percussive ones) may be reliably detected also in the time domain by simply monitoring the signal energy. However, sound signal is generally better described in the frequency domain, as opposed to e.g. image processing, where the frequency domain methods have usually more limited and specialized applications [15][16]. For sound, spectral analysis if far more flexible and it opens possibilities of the construction of many specialized algorithms where the signal may be easily split into frequency bands, often distributed logarithmically according to human perception of the pitch. For example, Böck and Widmer proposed an onset detection algorithm with vibrato suppression called SuperFlux, where the input data is filtered with a bank of varying-length frequency domain triangular filters spaced equally in musical scale and where the maximum filter is applied to the resulting spectrograms in order to ignore minor pitch fluctuations [17]. It has been shown [18] that this approach enhances onset detection for bowed instruments playing both with and without the vibrato technique.

In contrary to classical onset detection methods, many recent works involve machine learning techniques – most notably the neural networks [19][20][21], although other data-driven techniques, such as Support Vector Machines (SVM) have also been applied [22]. The input data usually consists of a time-frequency representation of the sound signal, mapped non-linearly in the frequency domain according to a perceptual model. Böck *et al* [21] used a bank of triangular filters positioned at critical bands of the Bark scale to filter the STFT magnitude spectra, computed with three different window lengths in parallel. In this way the redundancy resulting from unnecessarily high frequency resolution of the STFT in the upper frequency range may be avoided. Hertz to Mel scale mapping [23] and constant-Q transform [20] have also been applied for similar reasons.

Several approaches have been proposed in which the *fusion* of many onset detection functions is applied. This is accomplished either on the feature-level by a set of pre-defined rules or a linear combination of ODFs [24], or in the form of the *score-level fusion* in which the decisions are taken on the basis of the already computed onsets [25][24]. Quintela *et al* [25] apply i.a. KNN- and SVM-based classifiers to the lists of pre-computed onset candidates and their locations in time. Recently, Stasiak *et al* [10] proposed to simultaneously use several ODF functions as the input to a multilayer perceptron with one output, playing *de facto* the role of a new "integrated" onset detector. In this way the neural network learns to merge the onset-related information from various sources, while not being forced to extract it explicitly from raw spectral data.

On the other hand, the recent progress in theory and practical applications of deep neural architectures enabled to successfully use the solutions developed by the image processing community also to directly process audio spectrograms, transforming the onset detection task into a problem similar to that of texture recognition. Apart from bidirectional long short-term memory neural networks (LSTM) [23] and recurrent neural networks (RNN) [21], the convolutional neural networks (CNNs) [26][27] proved to be especially useful here.

In this work we adopted the approach proposed in [27] to test the effectiveness of a convolutional network in the onset detection task using two different signal representations, namely the logarithmically scaled spectrogram and enhanced autocorrelation (EAC).

III. The proposed approach

A. Neural network architecture

The input to our network is a spectrogram fragment in the form of an image with 15 columns, representing 15 consecutive time frames and 80 rows, corresponding to 80 logarithmically distributed frequency bands (up to 16kHz). The initial audio files are sampled 44100Hz and the spectrogram parameters are: window size $N = 2048$, hop-size $K = 512$ samples, which yields time resolution of ca. 11.6 ms. The target is composed of a single value, indicating the distance of the onset from the middle frame of the current input image, similarly as in [10] (Fig. 1). If more onsets are present within the fragment, only the closest one is considered. In this way the network has to solve *regression* problem instead of binary classification (onset absent/present in the middle of the fragment). Preliminary experiments showed that it enhances the results significantly.

The network structure is as follows:

- Convolutional layer with ten rectangular filters of size: w×h = 7×3 with ReLU (Rectified Linear Unit) activation function and stride value of 1 in both directions (full overlap). Note that for input size of w×h=15×80 it yields w×h=9×78 output.
- Max-pooling layer with non-overlapping kernels of size: w×h = 1×3 (output size w×h = 9×26).
- Convolutional layer with twenty square filters of size w×h = 3×3 and stride value of 1 in both directions (output size w×h = 7×24).
- Max-pooling layer with non-overlapping kernels of size: w×h = 1×3 (output size w×h = 7×8).
- Inner product (i.e. fully connected) layer with 256 hidden neurons and ReLU activation function.
- Inner product (i.e. fully connected) layer with one output neuron and tanh (hyperbolic tangent) activation function.

The neural architecture basically follows the scheme proposed by Schlüter and Böck in [27] with some modifications concerning – apart from the aforementioned regression, replacing classification – mostly the type of nonlinearity of the layers. We agree with [27] that the rectified linear units in the first convolutional layer may play the role of the half-wave rectifier H function (cf. Eq. 3) helping to detect onset-related energy increases. Additionally, we use the same nonlinearity type for the fully connected layer, instead of sigmoidal units which proved to positively influence the learning process in our tests. We also change the unipolar sigmoid into tanh function in the output neuron, which leads to increasing the output range from $[0, 1]$ into $[-1, 1]$ (cf. Fig. 1, the top plot).

The last change influences the threshold which is applied to the output of the network in order to find the onset positions.

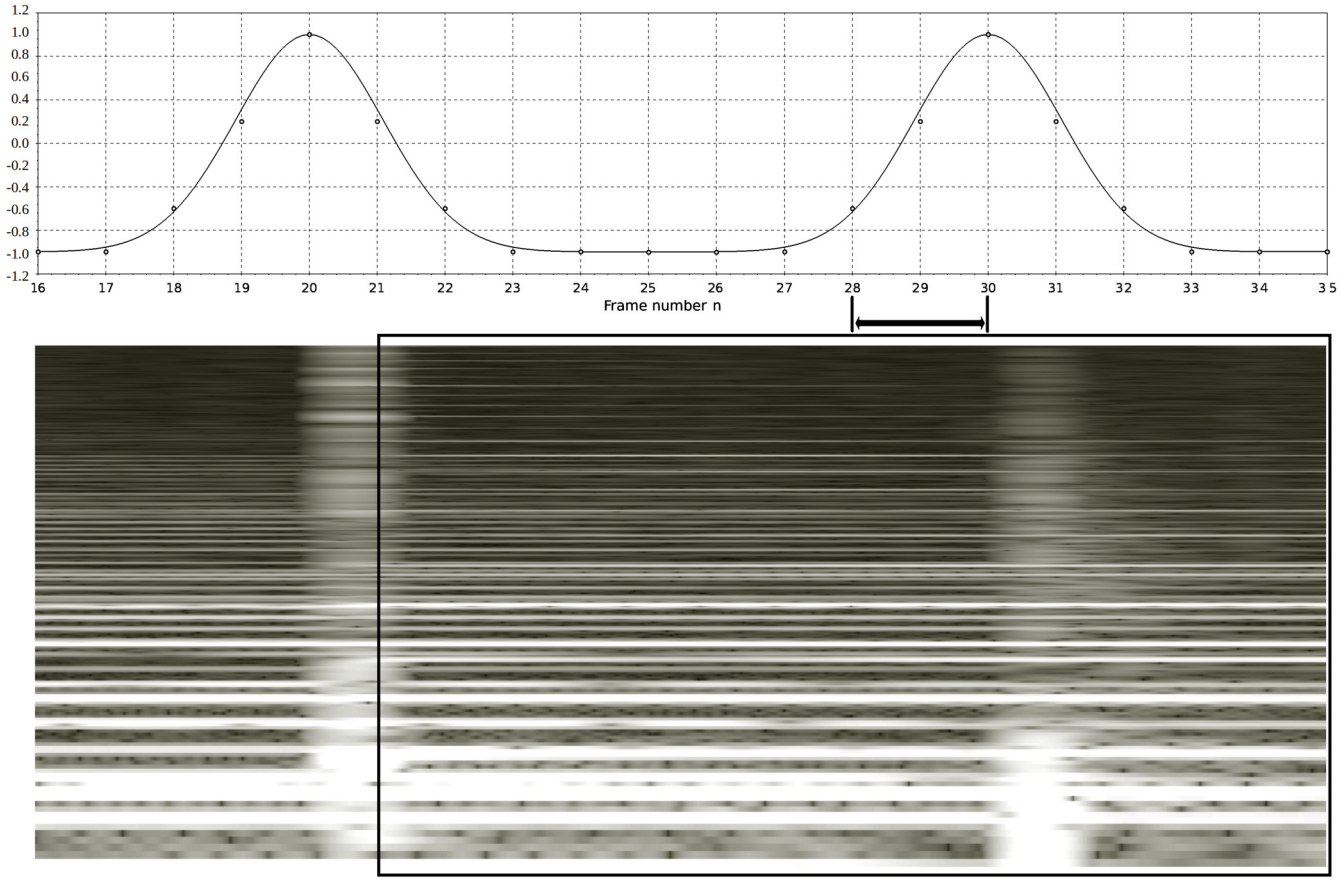


Fig. 1. *Spectrogram fragment enlarged (in a black box, lower plot) and the associated target values (top plot). The middle of this fragment (frame 28) is two frames apart from the onset (frame 30), so the target value for this fragment is -0.6 (top plot). Note, that the actual resolution of the image representing this fragment, that is fed to the network input is much lower (w×h=15×80 pixels)*

It should be mentioned here, that the output of the trained network may be treated as a classical ODF, with the difference, that a fixed threshold T may be used instead of moving mean or moving median, due to general lack of dependence on the signal energy. For the tanh activation function the optimal threshold value T_{opt} determined in our tests, i.e. the value maximizing the F-measure [27][10] was always lower than zero. After the thresholding, the peak-picking procedure is applied and peaks found within the range of 50ms relative to the actual onsets are treated as the properly detected ones.

Having denoted the correctly located onsets by **TP** (true positives), the assessment of the quality of the onset detection may be expressed in terms of *precision*, defined as the ratio: **TP/(TP+FP)**, and *recall*, defined as: **TP/(TP+FN)**. In our experiments we use the harmonic mean of precision and recall, known as the *F-measure*, as a "balanced" result of the onset detection procedure [10].

B. Audio material and data preparation

The dataset used in our experiments is a collection merged from several sources, including [8][28][29][20][30]. The total duration of all the audio files in our collection is over 45 min. and it contains 10,939 annotated onsets. The dataset has been divided at random into the train, test and validation subsets, containing 6236, 2520 and 2183 onsets, respectively. Complete files are assigned to either of the subsets (they are not split between the subsets).

For training, the spectrograms are cut into overlapping fragments which are then selected so that the obtained set is balanced, i.e. the number of "onset fragments" (for which the target value is non-zero) is equal to the number of "non-onset fragments". For testing, all possible fragments are used in an ordered sequence.

The main time-frequency representation (TFR) used in the experiments is the spectrogram, computed as explained in the previous section. In a separate test we use also enhanced autocorrelation (EAC) correlogram, calculated frame-by-frame in a similar way. EAC is an intermediate representation for the task of pitch estimation, thus also suitable for supporting onset detection with a degree of additional information on the input audio features related to melodic content. The procedure itself had been developed by Tolonen and Karjalainen [31] and (as the name implies) it is an extension of standard autocorrelation method. In our research we use EAC implementation operating in frequency domain for each frame of input signal using the

following processing scheme:

1) Transform a frame to frequency domain with Fourier transform. In this step, we use the same parameters as for the spectrogram computation – frame size of 2048 samples and frame step size (hop-size) of 512 samples.
2) Compute signal power
3) Take cube root of the resulting transform to compress magnitude in a non-linear manner. For normal autocorrelation the spectral coefficients are raised to the power of 2, however using the factor of 1/3 (cube root) of "generalised autocorrelation" is more suitable for the task of periodicity detection.
4) Clip all values below zero
5) Create a stretched copy (by factor of 2) of the values derived, and subtract it from the original (at step 4)
6) Clip all values below zero
7) Transform back to time domain with inverse Fourier transform

Steps 4-6 are performed for the purpose of peak pruning to improve pitch representation clarity, following Tolonen and Karjalainen's method. The procedure is applied to the signal frame-wise, yielding a correlogram, which can be processed further in a similar way as an ordinary spectrogram (Fig. 2).

C. Experimental evaluation

Caffe framework [32] has been used for training and testing the convolutional neural network presented in Sect. III-A. Separate validation set was used to determine the optimal model and to avoid overfitting. Stochastic gradient descent with momentum was used as the optimization strategy with mini-batch size of 1000 input spectrogram fragments. We used fixed momentum parameter of 0.4 and variable step size.

In the initial experiments we tested the influence of the activation function type on the results, as discussed in Sect. III-A. The results are presented in Table I.

TABLE I
THE RESULTS OF THE ONSET DETECTION TESTS

Experiment	F-measure
Original architecture based on [27]	82.13%
Our version with ReLU in the hidden layer and tanh in the output layer	83.35%
Our version trained on EAC correlograms instead of the spectrograms	73.10%

Although our modification enhanced the result by over one percent point, yet the EAC correlogram appeared definitely inferior to the spectrogram-based TFR. In the search of the potential reasons we conducted a series of additional tests in which we compensated for the potentially different annotation procedures in our heterogeneous dataset, by artificially shifting the onset positions by several multiples of the hop-size (from $-4 \times K$ to $4 \times K$). All the onsets in a given file were naturally shifted by the same displacement, but the displacement for each file was determined independently. Due to the latter fact, the figures presented in Table II obviously cannot be treated as the final objective results achieved by our network – they are rather indicators of its theoretical capabilities if some strict, uniform rules were applied for annotating the input files. They may also be used for a comparison of the spectrogram- and correlogram-based representations which again shows definite superiority of the first one. Figure 3 demonstrates the F-measure changes for varying values of the threshold T for the spectrogram-based input.

TABLE II
THE RESULTS OF THE ONSET DETECTION TESTS WITH ONSET SHIFTING

Experiment	F-measure
Our version with ReLU in the hidden layer and tanh in the output layer	88.62%
Our version trained on EAC correlograms instead of the spectrograms	81.13%

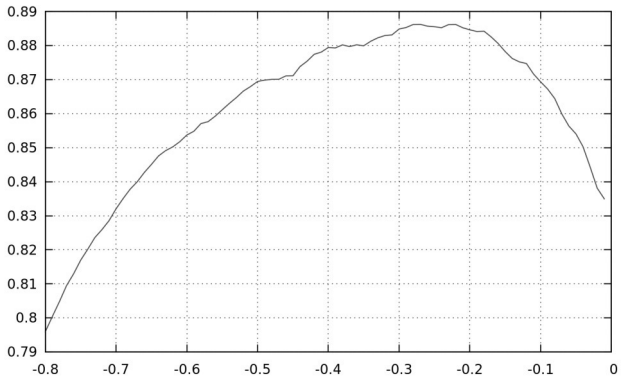


Fig. 3. *F-measure for varying values of the threshold T*

IV. DISCUSSION AND FUTURE WORKS

In this paper a convolutional neural network has been applied in the note onset detection problem. The obtained results demonstrate the superiority of the standard, spectrogram-based representation of the audio signal over the EAC correlogram. This observation confirms the potential of convolutional neural networks which are able to successfully extract useful information from the lower-level audio representation (spectrogram). The EAC correlogram, on the other hand, may be seen as a result of some more sophisticated processing, yielding more directly interpretable information related to pitch and melody content. However, this processing, although potentially useful from the human perspective, inevitably removes some information, which in consequence limits the potential of the convolutional neural network, eventually impairing the results.

The obtained results for the spectrogram-based input are satisfactory in terms of absolute onset detection rate. Enhancements might be searched for in increasing the precision of onset location (the annotations in the database used in the experiments should be manually checked and corrected to obtain more consistent annotation style [14]). Also, combining several time-frequency representations in a single spectrogram fragment, possibly computed with varied window size

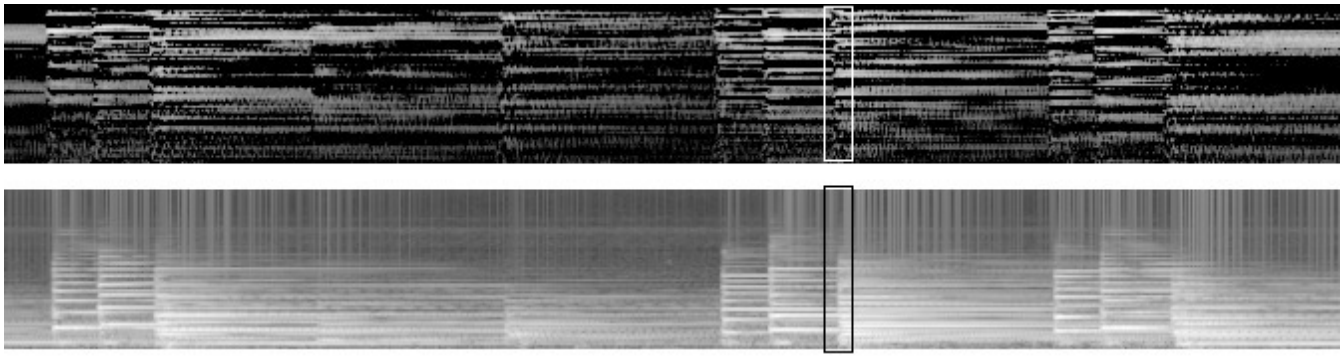


Fig. 2. *EAC correlogram (top) and spectrogram (bottom) of the same input file, with a bounding box around a fragment with an onset in the middle*

as proposed in [27], would probably lead to some further improvements.

ACKNOWLEDGMENT

We are truly grateful to Juan Pablo Bello, Sebastian Böck and Andre Holzapfel for making the annotated audio datasets available for our experiments.

REFERENCES

[1] A. Lerch, *An Introduction to Audio Content Analysis: Applications in Signal Processing and Music Informatics*. Wiley-IEEE Press, 2012.

[2] B. Thoshkahna and K. R. Ramakrishnan, "An onset detection algorithm for query by humming (QBH) applications using psychoacoustic knowledge." in *Proc. of 17th European Signal Processing Conference, EUSIPCO 2009*. IEEE, 2009, pp. 939 – 942.

[3] B. Stasiak, "Query by Singing/Humming (MIREX 2015). The Tune Follower," 2015. [Online]. Available: http://www.music-ir.org/mirex/abstracts/2015/BS2.pdf

[4] M. Purgina, A. Kuznetsov, and E. Pyshkin, "An approach for developing a mobile accessed music search integration platform," in *Proc. of Federated Conference on Computer Science and Information Systems, FedCSIS 2013*, M. Ganzha, L. Maciaszek, and M. Paprzycki, Eds. IEEE, 2013, pp. 267–273.

[5] E. Półrolniczak and M. Kramarczyk, "Analysis of the sound attack in context of computer evaluation of the singing voice quality," in *Proc. of Federated Conference on Computer Science and Information Systems, FedCSIS 2015*, M. Ganzha, L. Maciaszek, and M. Paprzycki, Eds., vol. 5. IEEE, 2015. doi: 10.15439/2015F240 pp. 889–894.

[6] B. Stasiak and K. Rychlicki-Kicior, "Fundamental frequency extraction in speech emotion recognition," *Communications in Computer and Information Science*, vol. 287, pp. 292 – 303, 2012. doi: 10.1007/978-3-642-30721-8-29

[7] H. Wang and L. Wang, "Onset detection algorithm in voice activity detection for Mandarin," in *Proc. of Int. Conf. on Computer Science and Network Technology (ICCSNT)*. IEEE, 2013. doi: 10.1109/ICCSNT.2013.6967305 pp. 1148 – 1151.

[8] J. Bello, L. Daudet, S. Abdullah, C. Duxbury, M. Davies, and M. Sandler, "A Tutorial on Onset Detection in Music Signals," *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 5, pp. 1035–1047, September 2005. doi: 10.1109/TSA.2005.851998

[9] B. H. Repp, "Patterns of note onset asynchronies in expressive piano performance," *Journal of the Acoustical Society of America (JASA)*, vol. 100, no. 6, pp. 3917–3932, 1996. doi: 10.1121/1.417245

[10] B. Stasiak, J. Mońko, and A. Niewiadomski, "Note onset detection in musical signals via neural-network-based multi-ODF fusion," *International Journal of Applied Mathematics and Computer Science*, vol. 26, no. 1, pp. 203 – 213, 2016. doi: 10.1515/amcs-2016-0014

[11] P. Bello and M. Sandler, "Phase-based note onset detection for music signals," in *Proceedings of IEEE Conference on Acoustics, Speech, and Signal Processing ICASSP*, vol. 5, 2003. doi: 10.1109/ICASSP.2003.1200001 pp. 441–444.

[12] C. Duxbury, J. Bello, M. Davies, and M. Sandler, "Complex Domain Onset Detection For Musical Signals," in *Proceedings of the 6th International Conference on Digital Audio Effects (DAFx-03)*, September 2003.

[13] J. Laroche, "Efficient Tempo and Beat Tracking in Audio Recordings," *Journal of the Audio Engineering Society (JAES)*, vol. 51, no. 4, pp. 226–233, 2003.

[14] S. Böck, F. Krebs, and M. Schedl, "Evaluating the online capabilities of onset detection methods." in *Proceedings of the 11th International Society for Music Information Retrieval Conference (ISMIR), 2012.*, 2012.

[15] V. Korzhik, G. Morales-Luna, A. Kochkarev, and I. Shevchuk, "Fingerprinting system for still images based on the use of a holographic transform domain," in *Proc. of Federated Conference on Computer Science and Information Systems, FedCSIS 2013*, M. Ganzha, L. Maciaszek, and M. Paprzycki, Eds. IEEE, 2013, pp. 585–590.

[16] B. Stasiak and M. Yatsymirskyy, *Frequency Domain Methods for Content-Based Image Retrieval in Multimedia Databases*. Springer Berlin Heidelberg, 2009, pp. 137 – 166. ISBN 978-3-642-02196-1. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-02196-1_6

[17] S. Böck and G. Widmer, "Maximum filter vibrato suppression for onset detection," in *Proceedings of the 16th International Conference on Digital Audio Effects (DAFx-13)*, Maynooth, Ireland, September 2013, pp. 55–61.

[18] B. Stasiak and J. Mońko, "Analysis of Onset Detection with a Maximum Filter in Recordings of Bowed Instruments," in *Proceedings of the 138th Audio Engineering Society Convention*, May 2015. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=17695

[19] M. Marolt, A. Kavcic, and M. Privosnik, "Neural networks for note onset detection in piano music," in *Proceedings of the International Computer Music Conference*, 2002.

[20] A. Lacoste and D. Eck, "A Supervised Classification Algorithm for Note Onset Detection." *EURASIP Journal of Advanced Signal Processing*, pp. 153–153, 2007. doi: 10.1155/2007/43745

[21] S. Böck, A. Arzt, F. Krebs, and M. Schedl, "Online Real-time Onset Detection with Recurrent Neural Networks," in *Proceedings of the 15th International Conference on Digital Audio Effects (DAFx 2012)*, September 2012.

[22] M. Davy and S. J. Godsill, "Detection of abrupt spectral changes using support vector machines. An application to audio signal segmentation." in *ICASSP*. IEEE, 2002. doi: 10.1109/ICASSP.2002.1005992 pp. 1313–1316.

[23] F. Eyben, S. Böck, B. Schuller, and A. Graves, "Universal Onset Detection with Bidirectional Long Short-Term Memory," in *Neural Networks, 11 th International Society for Music Information Retrieval Conference (ISMIR 2010)*, 2010, pp. 589–594.

[24] M. Tian, G. Fazekas, D. A. A. Black, and M. Sandler, "Design and Evaluation of Onset Detectors Using Different Fusion Policies," in *15th International Society of Music Information Retrieval (ISMIR) Conference*, 2014, pp. 631–636.

[25] N. D. Quintela, A. P. Giménez, and S. T. Guijarro, "A Comparison of Score-level Fusion Rules for Onset Detection in Music Signals." in *Proceedings of 10th International Society for Music Information Retrieval Conference ISMIR09*, October 2009, pp. 117–121.

[26] J. Schlüter and S. Böck, "Musical Onset Detection with Convolutional Neural Networks," in *6th International Workshop on Machine Learning and Music (MML)* , 2013.

[27] J. Schlüter and S. Bock, "Improved Musical Onset Detection with Convolutional Neural Networks," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2014)*, Florence, Italy, 2014. doi: 10.1109/ICASSP.2014.6854953

[28] L. Daudet, G. Richard, and P. Leveau, "Methodology and Tools for the evaluation of automatic onset detection algorithms in music." in *ISMIR*, 2004, pp. 72–75.

[29] A. Holzapfel, Y. Stylianou, A. C. Gedik, and B. Bozkurt, "Three dimensions of pitched instrument onset detection." *IEEE Trans. Audio, Speech & Language Processing*, vol. 18, no. 6, pp. 1517–1527, 2010. doi: 10.1109/TASL.2009.2036298

[30] J. Glover, V. Lazzarini, and J. Timoney, "Real-time detection of musical onsets with linear prediction and sinusoidal modeling." *EURASIP J. Adv. Sig. Proc.*, vol. 2011, p. 68, 2011. doi: 10.1186/1687-6180-2011-68

[31] T. Tolonen and M. Karjalainen, "A computationally efficient multipitch analysis model," *IEEE Transactions on Speech and Audio Processing*, vol. 8, no. 6, pp. 708–716, Nov 2000. doi: 10.1109/89.876309

[32] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," *arXiv preprint arXiv:1408.5093*, 2014.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 153–160
DOI: 10.15439/2016F562
ACSIS, Vol. 8. ISSN 2300-5963

iQbees: Interactive Query-by-example Entity Search in Semantic Knowledge Graphs

Marcin Sydow
Grzegorz Sobczak
Institute of Computer Science,
Polish Academy of Sciences, Warsaw, Poland
msyd@ipipan.edu.pl
grzegorz.sobczak0@gmail.com

Ralf Schenkel
University of Passau, Germany
rschenkel@acm.org

Krzysztof Mioduszewski
Polish-Japanese Academy of
Information Technology, Warsaw, Poland
krzysztof.mioduszewski@pjwstk.edu.pl

***Abstract*—We present IQBEES, a novel prototype system for similar entity search by example on semantic knowledge graphs that is based on the concept of maximal aspects. The system makes it possible for the user to provide positive and negative relevance feedback to iteratively refine the information need. The maximal aspects model supports diversity-aware results.**

***Index Terms*—entity search; relevance feedback; interactive retrieval; aspect model; diversity**

I. Introduction

IN THIS paper we present *iQbees*, a prototype system for *Interactive Query-By-Example Entity Search* on semantic knowledge graphs. The system solves the following list completion problem: given one or more example entities (e.g. actors, movies, etc.) as representatives of a group of entities, find other entities in that group. Our system solves this task through a sequence of *interactive query refinement steps* based on positive or negative user feedback. The results in each step are selected according to our own model based on *maximal aspects* that supports diversity awareness of the results.

The target of our interactive prototype *iQbees* system are non-expert users searching for a well-defined group of entities, for example European scientists who won a Nobel prize in physics. We assume that the users are unaware of the structure of the knowledge base and are not familiar with powerful structured query languages like SPARQL, thus cannot formulate a precise query that would satisfy their need. For such users, it is much simpler to provide one or a few examples of entities within the target group; in our example, such entities could be Maria Skłodowska-Curie or Max Planck. Our system will then, in a sequence of steps, present candidate result entities, for some of which the user will be able to give positive or negative feedback, refining the initial query and moving closer to her search goal. Our system thus combines techniques from entity list completion and relevance feedback and interactive search. The *iQbees* is built on the model previously applied in our (off-line) QBEES system [1], equipped with the interactive approach for providing positive and negative feedback by the user.

The first author is also with Polish-Japanese Academy of Information Technology, Warsaw, Poland

An obvious issue in such a scenario is the *ambiguity* of the actual user information need that is imperfectly represented by the given set of example entities. This problem is particularly difficult when just a single example entity is given which may represent a large number of possible entity sets; more example entities make this somewhat easier, but still far from trivial to solve. In such a scenario, it is almost impossible to immediately retrieve the correct set of entities in a single step, as standard list completion systems would do. The QBEES system, which is used as the backbone of our interactive *iQbees* system, does not provide user interaction and thus, as other list completion systems too, does not always retrieve the results that match the user information need.

We model possible user information needs by the concept of *aspects* of an entity, i.e., subsets of the facts and types in the semantic knowledge graph that characterize the entity. This model is designed so that it naturally supports *diversity* of the retrieved entities in order to cover as many as possible potential aspects (and thus possible information needs). A user will therefore likely find at least one relevant entity among the results for which she can give positive feedback, refining her query for the next iteration. The result diversity is guaranteed by the properties of *maximal aspects* that will be shortly described later. The *iQbees* system relaxes the strict aspect-based retrieval model used in QBEES, focusing more on popular entities that a user may know instead of obscure entites that perfectly match a possible information need.

iQbees and its underlying rankers are based on the structural and statistical properties of the underlying semantic knowledge graph. Such approach is orthogonal to any other approaches using external data or textual features of the entities. Hence, our approach can be easily enriched or combined with other text-based approaches that have been havily researched before. Our model assumes the correctness and completeness of the data, treating the issue of missing and wrong facts as out of scope of this paper.

II. The QBEES Aspect-based Model

Our system is based on the *aspect model* that will be very shortly described in this section. The full description of the model can be found in [1].

978-83-60810-90-3/$25.00

A. Knowledge graph

A Knowledge Graph KG is a directed multi-graph that consists of three basic components, a *Fact Graph* FG, an *Ontology Tree* O, and a set of type assignment arcs TA connecting the two. Arcs in KG are labelled. We will use the notation `relation(arg1,arg2)` for any directed arc with label `relation` in KG that points from node `arg1` to `arg2`.

The *Fact Graph* $FG = (E, F)$ is a directed multigraph where nodes in E represent *entities* (e.g. `Fryderyk_Chopin`, `Warsaw`) and edges in F represent *facts* about the entities. For example, an arc `bornIn(Fryderyk_Chopin, Żelazowa_Wola)` represents the fact that a Polish composer Fryderyk Chopin was born in Żelazowa Wola or `hasWonPrize(Max_Planck,Nobel_Prize_in_Physics)` means that the German scientist Max Planck is a Nobel Prize awardee in Physics.

The *Ontology Tree* $O = (C, S)$ is a graph where each node (class) $c \in C$ represents some *type* of entities (e.g. person). The class nodes are connected by directed arcs labelled as `subClassOf`. For instance, `subClassOf(composer,musician)` indicates that every composer is also a musician. Because the only relation present in this ontology tree in our current setting is the `subClassOf` relation, it can be also viewed as just a *taxonomy* of types.

The *Type Assignment* TA is a set of arcs labelled `hasType` which connect entities from the Fact Graph and classes from the Ontology Tree. For example the arc `hasType(Chopin, composer)` means that "Chopin is a composer".

B. Basic aspects

A basic aspect represents an "atomic property" that describes a specific entity q. In our most general setting, we distinguish three types of basic aspects: *fact aspects*, *relational aspects* and *type aspects*.[1]

1) Fact aspects: are created from the arcs incident with the entity in the fact graph FG that represent a fact concerning this entity. For example: the arc `bornIn(Chopin,Poland)` induces the fact aspect `bornIn(.,Poland)`.

2) Relational aspects: are predicates obtained from arcs that represent facts in the fact graph FG concerning the entity but the remaining argument of a factual aspect is replaced by a free variable ?. For example, `actedIn(.,?)` indicates that an entity acted in at least one movie.

3) Type aspects: are obtained similarly as relational aspects but inside the type assignment TA set of edges. It is obtained by replacing the particular entity q in an arc that represents a type with a free variable. Intuitively it represents the type of the entity under consideration. For example, a type arc `hasType(Chopin,composer)` naturally induces predicates of the form `hasType(.,composer)` that represents the "basic property" of this entity of "being a composer".

The three kinds of basic aspects can be viewed as forming a 3-level hierarchy, from the most general type aspects to more specific relational aspects (since particular kinds of relations between entities can be bounded only to particular types of entities) to the most specific factual aspects (as being realisations of relational aspects by substituting the free variable with a particular value referring to other entity).

[1] In some particular practical tasks the general framework can be simplified by reducing the considered types of basic aspects to only fact aspects, for example.

C. Compound aspects

A set of basic aspects is called a *compound aspect*. For example, a property "being a composer born in Poland", which consists of two basic aspects - "being a composer" and "being born in Poland", is represented by a set of two basic aspects, i.e. a compound aspect: {`bornIn(.,Poland)`, `hasType(.,composer)`}.

It is easy to see that each entity can be characterised by its set of basic aspects and vice versa – each set of basic aspects represents some set of entities that share it.

For an entity $e \in E$ we will henceforth use A_e to denote a *set of all basic aspects of* e.

We will say that "entity e *satisfies* a compound aspect A" whenever $A \subseteq A_e$.

D. Entity set of an aspect

For each basic aspect $a_q \in A_q$, of some entity q, we naturally define its entity set $E(a_q)$ as the set of all entities $e \in E$ that share this aspect with q, i.e. contain a_q in their set of basic aspects A_e:

$$E(a_q) = \{e \in E : a_q \in A_e\}$$

We extend the above definition of entity set $E(a_q)$ (for a basic aspect a_q) to the concept of entity set $E(A)$ of a compound aspect A as the set of all entities that share *all* basic aspects in A.

E. Maximal aspects

Let q be a query example and A_q be its set of basic aspects. For any $e \in E, e \neq q$ consider the set of all basic aspects of e common with q, that is $A'_e = A_e \cap A_q$.

A compound aspect A_q for entity q is called a *maximal aspect* if and only if it satisfies the following two conditions:

1) it is satisfied by q and *at least one entity other than* q
2) A_q is maximal wrt inclusion (i.e., extending this set of basic aspects with any more basic aspect of q would violate the first condition).

In other words, the maximal aspects of q are maximal compound aspects satisfied by q and any other entity.

We assume that the concept of maximal aspect is defined with respect to the current content of the underlying semantic knowlege graph.

There may exist multiple different maximal aspects for a given entity.

The definition of maximal aspect for a single entity q extends to a set Q of query entities as follows. For a query entity set Q we say that a compound aspect A_Q is *a maximal aspect for* Q if and only if it satisfies the following two conditions:

1) it is satisfied by all entities in Q and at least one entity outside Q
2) A_Q is maximal wrt inclusion

We introduce the denotation of $E(A_Q)$ as the natural extension of the concept of $E(A_q)$ corresponding to a single entity q to the set of entities Q.

To illustrate the maximal aspect concept we will use the following example. Assume an entity set $Q = \{\texttt{Schwarzenegger}, \texttt{Stallone}\}$ is given. Then, consider two compound aspects for Q:

$A_1 = \{\texttt{hasType(.,ActionMovieActor)}, \texttt{livesIn(.,USA)}\}$

$A_2 = \{\texttt{hasType(.,ActionMovieActor)}, \texttt{livesIn(.,USA)}, \texttt{hasType(.,MovieDirector)}\}$

If, in the underlying knowledge base, there is at least one entity that is an action movie actor and director living in USA, other than Schwarzenegger and Stallone the first condition of the maximality definition holds. Moreover, if adding any other basic aspect of Q (e.g. additionally being a body-builder) would make the compound aspect satisfied only by the two mentioned entities, the set A_2 is a maximal aspect for Q while A_1 is obviously not since it subsumes A_2.

F. Diversity-awareness of Maximal Aspects

Maximal aspects have the following two crucial properties:

- (RELEVANCE) any entity that satisfies a maximal aspect is *maximally similar* (in terms of sharing maximally many basic aspects with the target entities)
- (DIVERSITY) each maximal aspect A_Q represents a *different* set of entities, i.e. they do not intersect except Q.

The last property can be expressed in the form of the following theorem:

Theorem: *Let Q be a (query) set of entities and $A_Q \neq B_Q$ be two different (non-empty) maximal aspects of Q. Then, $E(A_Q)$ and $E(B_Q)$ do not share any entities, except those in Q (i.e. $(E(A_Q) \cap E(B_Q) \setminus Q) = \emptyset$).*

Proof: *Assume, $e \in E(A_Q) \cap E(B_Q)$ for some entity $e \notin Q$. This implies that e shares all the basic aspects from both A_Q and B_Q. Let us introduce denotation $C_Q = A_Q \cup B_Q$. Thus, e shares all basic aspects from C_Q which implies that $e \in E(C_Q)$. But, since A_Q and B_Q are different and non-empty, C_Q strictly contains both A_Q and B_Q which would contradict the maximality property of them.*

Due to the above theorem, the set of candidate similar entities to be returned is *partitioned* by the entity sets of maximal aspects. Thus, the maximal aspects model supports *diversity* of the results, since they are non-redundant (i.e. different maximal aspects imply non-intersecting sets of entities).

G. Searching with basic QBEES approach

Given a set of query examples, the approach calculates all maximal aspects and generates results from them based on a two-step ranking scheme that first selects the most promising maximal aspect and then the most promising entity that satisfies this maximal aspect. This process is repeated until k results are retrieved. Entities are ranked based on various factors including graph properties (e.g. random-walk based) and importance (popularity). Maximal aspects are ranked based on their size and the popularity of their entities, hence retrieving (and thereby removing) an entity from a maximal aspect changes its score. We adapted the QBEES ranking system that is out of scope of this paper and is described in [1].

III. Interactive Search with *iQbees*

With *iQbees*, the static search procedure used by QBEES is extended with interactive feedback cycles.

Initially, the user provides an example entity.[2]. The system then returns an initial list of result entities using the ranking approach. If this answer is not yet perfect, the user can use the returned entities as *refinement suggestions* and select some of them as *positive* hints, some of them as *negative* hints. The system will then exploit this feedback and produce new, hopefully better, results. This interaction cycle proceeds until the user is satisfied with the results.

A. Positive feedback and Negative feedback

To shortly explain the mechanism of positive and negative feedback, let's introduce some ancillary concepts. Let *domain* be defined as the intersection of compound aspects of all initial entities and all positive feedback entities. Let *candidate aspects* be defined as the set of compound aspects that are contained in the *domain*. In the basic setting they are exactly *maximal aspects* contained in the *domain*. In the *relaxed* variant, that will be described below, *candidate aspects* are any compound aspects contained in the *domain*.

Filtered candidates are exactly those *candidate aspects* that are *not satisfied* by any entity from the negative hint set.

Then, we run the algorithm using *filtered candidates* only, i.e. only the entities that satisfy filtered candidates can be returned.

B. Relaxed Variant of Aspect Selection

The maximal aspect concept proved to perform well in the basic QBEES algorithm [1] where the user provided a set of entities in a "one-shot" mode.

However, in the interactive *iQbees* approach, if we applied the basic model described above, one can observe the following phenomenon. The more positive feedback entities are provided by the user, the more entities can be potentially returned which is counter-intuitive to the concept of *query refinement*. This phenomenon is due to the fact, that the more positive examples are provided, the smaller is their aspect intersection (*domain*). As a result, more entities can potentially satisfy such maximal aspect being the intersection of more entities, since it is easier to contain a smaller aspect set than a bigger one. This is counter-intuitive, since providing more positive examples should *refine* the query intent rather than make it more vague.

[2]Actually, our model can be easily extended to allow for any number of initial query entities. This can be also easily simulated by providing one entity and marking the others as positive feedback. The extension to multi-entity input is planned as a close future work

Hence, to "correct" such undesirable behaviour, in *iQbees* we "relax" the original QBEES mechanism by considering (as *candidates*) *all* compound aspects contained in the *domain* instead of considering only *maximal aspects* contained in the *domain* as *candidates*. Such "relaxed", non-maximal compound aspects are ranked using the base QBEES approach to promote the maximal or near-maximal aspects and only top-k are considered. After this "relaxation" the above phenomenon is removed, i.e. the more positive hints are provided, the fewer entities potentially satisfy (contain) any relaxed compound aspect.

IV. An example search session

A working demo application was built as a preliminary proof of concept, and as an experimental platform and to illustrate the approach studied in this paper. It is currently using YAGO [2] as the underlying semantic knowledge base[3]. By using the demo application a user, interested in a particular entity present in the semantic knowledge graph, can query the graph to find similar entities. *User Interface* UI of the demo application allows to mark each search result (entity) as relevant or not. This information is used in subsequent queries and allows the user to iteratively refine his initial query until he receives satisfactory results.

A. Preliminary Demo UI

Demo application UI consists of three main parts - *Query Input Field*, *Search Parameters Section* and *Search Results Section*. *Query Input Field* is a typical text input, where the user can type in the entity name which will be the subject of the query. *Search Parameters Section* contains UI controls which allow to change the number of expected search results and specific settings of the query algorithm.

Search Results Section, visible after performing the initial and subsequent queries, contains similar entities found in the previously executed query and *Feedback Subsection*. Each result entity has two buttons labeled "+" and "-". These buttons are used to mark a particular result entity as positive or negative feedback i.e. relevant or not to the initial user information need. Pressing one of these buttons results in moving the entity to the *Feedback Subsection* and marking it accordingly as "positive" or "negative" feedback for future searches. Each feedback entity has an additional "show debug" button which displays the set of all *basic aspects* satisfied by it. Likewise each search result entity has a "show debug" button which displays *maximal aspect* of query subject satisfied by this particular entity.

The "Calculate" button triggers query execution. The "Fresh Start" button restarts the searching session, clears the initial entity and previous search feedback. Screenshots of available "work in progress UI" are visible in Figures 1 and 2.

[3]Other semantic knowledge graphs are considered to be used as the underlying databes in our future work

B. Example usage scenario

Application user is interested in Arnold Schwarzenegger and entities similar to him in terms of political career. Notice the particular *ambiguity* of this entity, since it represents also many other aspects, e.g. famous body-builders, action movie actors, directors, etc. Assume the user applies the default search parameters visible in Figure 1 and inputs "Arnold Schwarzenegger"[4] in *Query Input Field* and clicks the "Calculate" button. The system verifies if "Arnold Schwarzenegger" entity is present in the underlying knowledge graph (demo application requires exact match of the entity name) and starts computing similar entities. After certain amount of time the results are visible in *Search Results Section*. In the results, among others, the user can see:

- Gray Davis: a former governor of California.
- Dany DeVito: an American actor who acted in a move "Junior" with Schwarzenegger.

User marks Gray Davies as positive feedback and clicks "Calculate" once again. The new result set contains only politicians. The user can further search for similar entities by providing new feedback.

If user eventually decides that entities similar to Arnold Schwarzenegger in terms of political career are not interesting to her, she can start a new searching session by clicking the "Fresh Start" button. This will clear previous search results and return *Search Parameters Section* settings to default values. User once more inputs "Arnold Schwarzenegger" in *Query Input Field* and *Search Results Section* gets populated by previously visible results. This time user marks Gray Davies as negative entity by clicking the "-" button next to it. Once the search is finished a new result set contains mainly actors and movie producers and does not contain any politicians. In the next step user chooses Sandahl Bergman as a positive entity. Sandahl Bergman is an actress who starred in "Conan the Barbarian" and "Red Sonya" together with Schwarzenegger. The final result set contains actors and actresses among whom three actors played in "Conan the Barbarian" (Gerry Lopez, James Earl Jones) and one actress starred in "Red Sonya" (Brigitte Nielsen).

V. Experiments

The prototype system has been implemented and is being tested. Besides experimenting with the on-line prototype, we made a preliminary experimental evaluation aiming at an objective comparison of the described variants of our system on publicly available benchmark data.

We used the YAGO semantic knowledge graph [2] as the underlying database. We used data from the INEX 2007 entity track to build a 163-query one-entity gold-standard dataset by mapping Wikipedia pages to YAGO and using list completion topics, following the approach sketched in [1]. Each topic

[4]In our current working demo the input is not preprocessed so that the user has to type the exact string as it is represented in the database. Obviously, in future we plan to add semi-automatic query correction using methods proposed in [3] or [4] to make the user interface more user-friendly

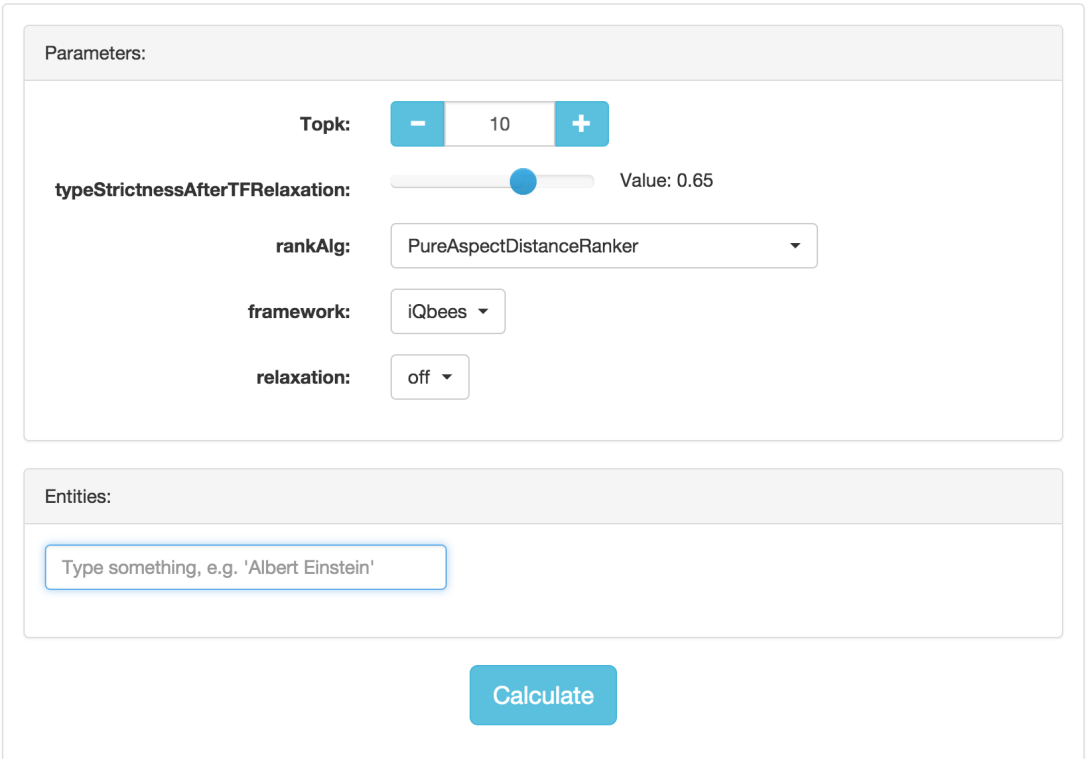


Fig. 1. Initial application view with default search parameters.

is provided with a set of relevant entities (called *ground truth*). As the basic back-end, we reused the existing QBEES engine that was extended by positive and negative feedback functionality and ranking all the candidate aspects (not only maximal aspects).

We simulated three simple 2-step strategies of the user. In each scenario the "simulated" user, in step 0, inputs one query entity to the system and then, in step 1, marks one (randomly) selected result entity that is relevant according to the ground truth as "positive" (p), or one (randomly) selected result entity that is irrelevant according to the ground truth as "negative" (n), or both (p+n). We simulate this by marking 1 *random* relevant or irrelevant entity, respectively.

Each combination of one of the three mentioned strategies (p, n, p+n) and one of the two extensions described above (FEEDBACK for positive and negative feedback, RELAX for additional consideration of non-maximal aspects) was run on each of the 163 one-entity queries. For all settings, we computed average MAP and MNDCG in step 0 and step 1 to measure and compare the performance of the system in each step and each setting. For each query and each setting the procedure of computation of measures is as follows: we run two steps; we remove the selected entieties from the ground truth and from both result list; then we calculate measures (MAP, MNDCG) for each step.

The presented experimental evaluation results indicate that in the two examined settings: marking positive (p) and positive and negative (p+n) entities observably improves the quality of the results in the next step (Tables I and III). Interestingly, we also found out that marking only 1 negative entity did not improve the results in the next step (Table II), in this particular experimental setting even if in our on-line experiments the negative feedback functionality seems to improve the results. This issue will be the matter of our further study.

Furthermore, one can observe that the relaxed variant of our approach (i.e. RELAX) performs consistently better than FEEDBACK in this setting.

VI. Previous and Related Work

The system is based on the *aspect model* described in detail in [1]. An early version of the *iQbees* system (without negative feedback functionality nor relaxation nor experimental evaluation) was presented in a short paper [5] that this paper is a substantial extension of.

Entity search has been considered extensively in the past, with a focus on finding related entities and list completion, and with extensive evaluation campaigns at TREC [6] and INEX [7]. We consider the specific scenario where entities

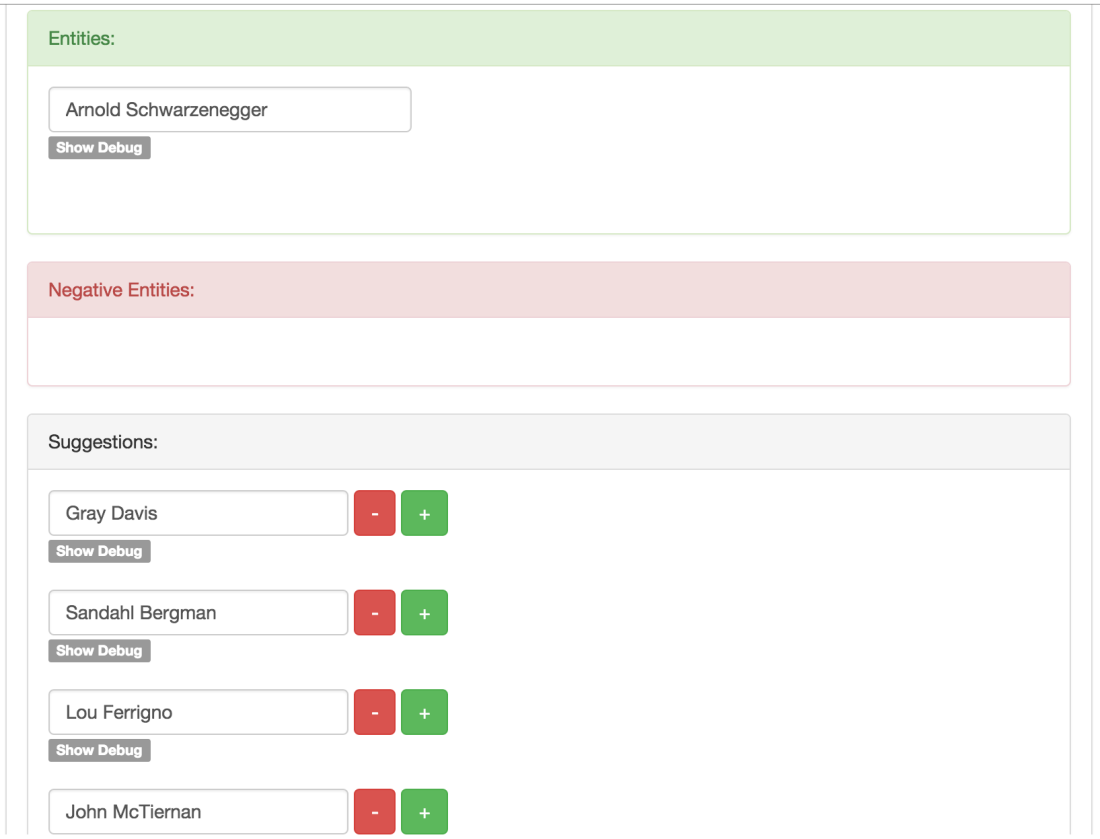


Fig. 2. Application view with search results visible.

TABLE I
THE RESULTS OF THE EXECUTED EXPERIMENTS FOR THE 1-POSITIVE HINT STRATEGY

Model	map		mndcg	
	step 0	step 1	step 0	step 1
FEEDBACK	0.0222	0.0404	0.0715	0.0941
RELAX	0.0386	0.0801	0.1015	0.1617

TABLE II
THE RESULTS OF THE EXECUTED EXPERIMENTS FOR THE 1-NEGATIVE HINT STRATEGY

Model	map		mndcg	
	step 0	step 1	step 0	step 1
FEEDBACK	0.0485	0.0441	0.1284	0.1173
RELAX	0.0650	0.0615	0.1563	0.1471

TABLE III
THE RESULTS OF THE EXECUTED EXPERIMENTS FOR THE MIXED (1-POSITIVE AND 1-NEGATIVE) STRATEGY

Model	map		mndcg	
	step 0	step 1	step 0	step 1
FEEDBACK	0.0212	0.0400	0.0679	0.0920
RELAX	0.0396	0.0567	0.0997	0.1289

from a knowledge graph are searched. Existing systems usually build on entity similarity measures, exploiting the graph structure (e.g., SimRank [8]), the context of entities in the graph (e.g., Albertoni and De Marino [9]), or additional context outside the graph (e.g., Bron et al. [10], which combines a term-based language model with a simple structural model).

The problem of example-based entity search has been actively studied recently. Yu et al. [11] solve a slightly different problem where entities similar to a single query entity are computed, exploiting a small number of example results. Focusing on heterogeneous similarity aspects, they propose to use features based on so-called meta paths between entities and several path-based similarity measures, and apply learning-to-rank methods for which they require labelled test data. Wang and Cohen [12] present a set completion system retrieving candidate documents via keyword queries based on the entity examples. Using an extraction system additional entities are then extracted from semi-structured elements, like HTML-formatted lists.

Mottin et al. [13], [14] introduce the concept of exemplar queries. Similar to our setting, an example result is used instead of a query. However, the setting in their XQ system is strictly different since it considers examples in the form of a connected subgraph of entities, not single entities, and determines result subgraphs based on their similarity to the query graph. The problem is therefore in some sense easier, as more information can be exploited for identifying query results.

The GQBE system by Jayaram et al. [15] is similar to XQ, but does not use connected subgraphs, but just entities that form a query result as input; the meaningful connections between those entities are explored by the system. Again, the main difference to our system is that we consider only single entities as results, not combinations, and hence have less information for identifying relevant results.

Relevance feedback has seen surprisingly little use in entity search on knowledge bases. Only very recently, Su et al. [16] proposed exploiting relevance feedback for improving results of searching a knowledge graph, but not for entity search. For entity ranking using text or semi-structured information, relevance feedback has been more popular [17].

Diversity-aware entity summarization was originally proposed in [18] and further studied in [19], but we are not aware of any work on entity search that takes diversity into account.

VII. Conclusions and Further Work

We presented *iQbees* – a prototype interactive approach to the problem of entity list completion based on semantic knowledge graphs. This is an extension of the previously presented QBEES system [1] by adding positive and negative interactive feedback functionality. The prototype of the approach was preliminarily implemented as a proof of concept and demonstration available online. We also presented experimental results that indicate that the proposed approach outperforms on a publicly available benchmark the basic (non-interactive) QBEES system without the feedback functionality. The future work includes better understanding and modeling of the feedback functionality, in particular negative feedback, and ranking strategies since the current experiments indicate that there is room for improvement in the current model. It is also envisaged to further work on developing the on-line demo in order to improve its functionality and optimise it and to perform more experimental evaluation on other semantic knowledge graphs, e.g. DBPedia.

VIII. Acknowledgments

This work is partially supported by the Polish National Science Centre grant 2012/07/B/ST6/01239. Grzegorz Sobczak was partially supported by the European Union under the European Social Fund Project PO KL "Information technologies: Research and their interdisciplinary applications", Agreement UDA-POKL.04.01.01-00-051/10-00 when working on this project during his research stay in the Institute of Computer Science, Polish Academy of Sciences, Warsaw, Poland. The authors would like to thank Steffen Metzger, who provided the substantial part of the code of the back-end of the earlier version of the system.

References

[1] S. Metzger, R. Schenkel, and M. Sydow, "Aspect-based similar entity search in semantic knowledge graphs with diversity-awareness and relaxation," in *WI-IAT*, pp. 60–69. [Online]. Available: http://dx.doi.org/10.1109/WI-IAT.2014.17

[2] Http://www.mpi-inf.mpg.de/yago-naga/yago.

[3] J. Piskorski and M. Sydow, *String Distance Metrics for Reference Matching and Search Query Correction*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 353–365. [Online]. Available: http://dx.doi.org/10.1007/978-3-540-72035-5_27

[4] J. Piskorski, K. Wieloch, and M. Sydow, "On knowledge-poor methods for person name matching and lemmatization for highly inflectional languages," *Information Retrieval*, vol. 12, no. 3, pp. 275–299, 2009.

[5] G. Sobczak, M. Chochół, R. Schenkel, and M. Sydow, "iQbees: Towards interactive semantic entity search based on maximal aspects," in *Foundations of Intelligent Systems*. Springer International Publishing, 2015, vol. 9384, pp. 259–264, 10.1007/978-3-319-25252-0-28. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-25252-0_28

[6] K. Balog, P. Serdyukov, and A. P. de Vries, "Overview of the TREC 2011 entity track," in *TREC*, 2011.

[7] G. Demartini, T. Iofciu, and A. P. de Vries, "Overview of the INEX 2009 entity ranking track," in *INEX*, 2009, pp. 254–264.

[8] G. Jeh and J. Widom, "SimRank: a measure of structural-context similarity," in *KDD*, 2002, pp. 538–543. [Online]. Available: http://doi.acm.org/10.1145/775047.775126

[9] R. Albertoni and M. D. Martino, "Asymmetric and context-dependent semantic similarity among ontology instances," *J. Data Semantics*, vol. 10, pp. 1–30, 2008. [Online]. Available: http://dx.doi.org/10.1007/978-3-540-77688-8_1

[10] M. Bron, K. Balog, and M. de Rijke, "Example based entity search in the web of data," in *ECIR*, 2013, pp. 392–403. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-36973-5_33

[11] X. Yu, Y. Sun, B. Norick, T. Mao, and J. Han, "User guided entity similarity search using meta-path selection in heterogeneous information networks," in *CIKM*, 2012, pp. 2025–2029. [Online]. Available: http://doi.acm.org/10.1145/2396761.2398565

[12] R. C. Wang and W. W. Cohen, "Language-independent set expansion of named entities using the web," in *ICDM*, 2007, pp. 342–350. [Online]. Available: http://doi.ieeecomputersociety.org/10.1109/ICDM.2007.104

[13] D. Mottin, M. Lissandrini, Y. Velegrakis, and T. Palpanas, "Exemplar queries: Give me an example of what you need," *PVLDB*, vol. 7, no. 5, pp. 365–376, 2014. [Online]. Available: http://dx.doi.org/10.14778/2732269.2732273

[14] ——, "Searching with XQ: the exemplar query search engine," in *SIGMOD*, 2014, pp. 901–904. [Online]. Available: http://doi.acm.org/10.1145/2588555.2594529

[15] N. Jayaram, A. Khan, C. Li, X. Yan, and R. Elmasri, "Querying knowledge graphs by example entity tuples," *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 10, pp. 2797–2811, 2015. [Online]. Available: http://doi.ieeecomputersociety.org/10.1109/TKDE.2015.2426696

[16] Y. Su, S. Yang, H. Sun, M. Srivatsa, S. Kase, M. Vanni, and X. Yan, "Exploiting relevance feedback in knowledge graph search," in *KDD*, 2015, pp. 1135–1144. [Online]. Available: http://doi.acm.org/10.1145/2783258.2783320

[17] T. Iofciu, G. Demartini, N. Craswell, and A. P. de Vries, "Refer: Effective relevance feedback for entity ranking," in *ECIR*, 2011, pp. 264–276. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-20161-5_26

[18] M. Sydow, M. Pikula, and R. Schenkel, "Diversum: Towards diversified summarisation of entities in knowledge graphs," *2014 IEEE 30th International Conference on Data Engineering Workshops*, vol. 0, pp. 221–226, 2010.

[19] M. Sydow, M. Pikuła, and R. Schenkel, "The notion of diversity in graphical entity summarisation on semantic knowledge graphs," *Journal of Intelligent Information Systems*, vol. 41, pp. 109–149, 2013. [Online]. Available: http://dx.doi.org/10.1007/s10844-013-0239-6

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 161–164
DOI: 10.15439/2016F409
ACSIS, Vol. 8. ISSN 2300-5963

Probabilistic 2D Cellular Automata Rules for Binary Classification

Miroslaw Szaban
Institute of Computer Science,
Siedlce University of Natural Sciences and Humanities, Poland
Email: mszaban@uph.edu.pl

***Abstract*—In this paper are presented classification methods with use of two-dimensional three-state cellular automata. This methods are probabilistic forms of cellular automata rule modified from wide known almost deterministic rule designed by Fawcett. Fawcetts rule is modified into two proposed forms partially and fully probabilistic. The effectiveness of classifications of these three methods is analysed and compared. The classification methods are used as the rules in the two-dimensional three-state cellular automaton with the von Neumann and Moore neighbourhood. Preliminary experiments show that probabilistic modification of Fawcett's method can give better results in the process of reconstruction (classification) than the original algorithm.**

I. Introduction

In a classification problem, we wish to determine to which class new observations belong, based on the training set of data containing observations whose class is known. The binary classification deals with only two classes, whereas in a multiclass classification observations belong to one of the several classes. The well-known classifiers are neural networks, support vector machines, k-NN algorithm, decision trees, and others. The idea of using cellular automata (CA) in the classification problem was described by Maji et al. [4], Povalej et al. [7] and recently by Fawcett [1]. Fawcett designed the heuristic rule based on the von Neumann neighborhood (so-called voting rule) moreover, tested its performance on different data sets. Recently, GA was considered as a tool to select CA rules with von Neumann neighbourhood for binary classification problem by Piwonska et al. [6].

Despite the fact that CAs have the potential to efficiently perform complex computations [9]; the main problem is a difficulty of designing CAs which would behave in the desired way. One must not only select a neighborhood type and size, but most importantly the appropriate rule (or rules). In some applications of CAs one can design an appropriate rule by hand (e.g. the GKL rule designed in 1978 by Gacs, Kurdyumov and Levin for density classification task [2]) or can use partial differential equations describing a given phenomenon [5]. Since the number of possible rules is usually huge, this is the extremely hard task, and it is not always possible to select them by hand. Therefore, in the 90-ties of the last century Mitchell et al. proposed to use GAs to find CAs rules able to perform one-dimensional density classification task [3].

In this paper are used rules based on methods of classification like the Fawcett's [1] method, and new proposed modifications into a probabilistic form of such method. Above rules are applied to the rectangular grid with both neighbourhood types i.e. von Neumann and Moore neighbourhood, and the effectiveness of the modified rules is compared with the effectiveness of the original Fawcett's rule.

This paper is organized as follows. Section 2 describes two-dimensional CAs and binary classification problem. In section 3 are proposed new CA-based classifier as a modification of the Favcett's rule. Experimental results are presented in Section 4. Last section contains conclusions and future works.

II. Two-Dimensional Cellular Automata and Binary Classification Problem

A two-dimensional CA considered in this paper is a rectangular grid of $N \times M$ cells, each of which can take on k possible states. After determining initial states of all cells (i.e. the initial configuration of a CA), each cell changes its state according to a rule - transition function TF which depends on states of cells in a neighborhood around it. In this paper is considered finite CA with the periodic boundary conditions. This is usually done synchronously, although asynchronous mode is used too. Two types of the neighborhood are commonly used: the von Neumann neighborhood (the four cells orthogonally surrounding the central cell) and can be described as: $a_{i,j}^{(t+1)}=TF[a_{i,j-1}^{(t)}, a_{i-1,j}^{(t)}, a_{i,j}^{(t)}, a_{i+1,j}^{(t)}, a_{i,j+1}^{(t)}]$, where $a_{i,j}^{(t)}$ denotes the state of a cell at position i, j in the two-dimensional cellular grid, at time step t, and also the Moore naighbourhood (the eight cells around the central cell) and can be described as: $a_{i,j}^{(t+1)}=TF[a_{i-1,j-1}^{(t)}, a_{i,j-1}^{(t)}, a_{i+1,j-1}^{(t)}, a_{i-1,j}^{(t)}, a_{i,j}^{(t)}, a_{i+1,j}^{(t)}, a_{i-1,j+1}^{(t)}, a_{i,j+1}^{(t)}, a_{i+1,j+1}^{(t)}]$.

The evolution of a CA is usually presented using so-called 'space-time diagrams' displaying grid of cells at subsequent time steps, with each state marked with different color.

The square state of the data space in classification problem should be i.e. $[0,1] \times [0,1]$. Suppose that $N \times M$ data-points $p_{(i,j)}=(x_i, y_j)$, where i=1, 2, ..., N and j=1, 2, ..., M are given as a training set from two classes: class 1 and class 2. When each of $p_{(i,j)}$ data-points is known as one of two classes then we have the classification. On the other hand, when even one of the data-points is not one of two known classes we have the classification problem. Moreover, to answer the question, what kind of class 1 or class 2 are unclassified data points it should be applied the classification method. In CA the data space of such problem should be mapped from $[0,1] \times [0,1]$ into the

grid of $N \times M$ cells (in this paper $N \times N$ for the simplicity). Each of cells can take one of 3 states, classified the state 1 (class 1) and state 2 (class 2) and also unclassified state (class 0). Classifier - the rule of CA will analyze the unclassified cells and changes its states into one of two known.

III. Proposed CA-based Classifier

The classification problem described in [1] is the base of this paper. The rule of CA known as $n4\_V1\_nonstable$ and presented there was the starting point to create a better classifier. The classification with use this rule is defined as:

- classify as $class$ 0, if $class$ 1 neighbors + $class$ 2 neighbors = 0,
- classify as $class$ 1, if $class$ 1 neighbors > $class$ 2 neighbors,
- classify as $class$ 2, if $class$ 1 neighbors < $class$ 2 neighbors,
- classify as $rand(\{class\ 1, class\ 2\})$, if $class$ 1 neighbors = $class$ 2 neighbors.

The Fawcett's rule is productive enough for classification problems in little CA grids. This rule was presented in [1] as better than other, but it was tested and compared for only 81×81 wide CA grid. Therefore, in this paper are proposed modifications of Fawcett's rule into two patterns: partially and fully probabilistic. A proposed modification should strengthen an original and more accurate classify binary data, specially for large CA grid. The partially probabilistic modification ($n4\_V1\_nonstable\_PP$) is proposed as follows:

- classify as $class$ 0, if $class$ 1 neighbors + $class$ 2 neighbors = 0,
- classify as $class$ 1, with probability $p(1)$,
- classify as $class$ 2, with probability $p(2)$,

where probability are calculated form classified neighbours in the neighbourhood, i.e.: $p(i)=\frac{class\ i\ neighbors}{class\ 1\ neighbors+class\ 2\ neighbors}$, where $i=\{1,2\}$. It means that unclassified cell will change in to one of classified state with probability calculated from known states in the neighbourhood and suitable state.

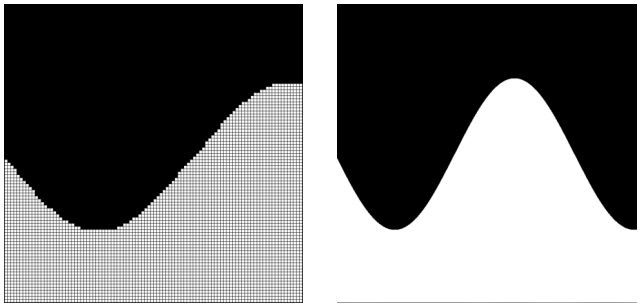
The full probabilistic modification ($n4\_V1\_nonstable\_FP$) is proposed as follows:

- classify as $class$ 0, with probability $p(0)$,
- classify as $class$ 1, with probability $p(1)$,
- classify as $class$ 2, with probability $p(2)$,

where probabilities are calculated form each class of neighbours in the neighbourhood, i.e.: $p(i)=\frac{class\ i\ neighbors}{\sum_{j=0}^{2} class\ j\ neighbors}$, where $i=\{0,1,2\}$. It means that unclassified cell could change in to one of classified state or stay unclassified with probability calculated from states of cells in neighbourhood.

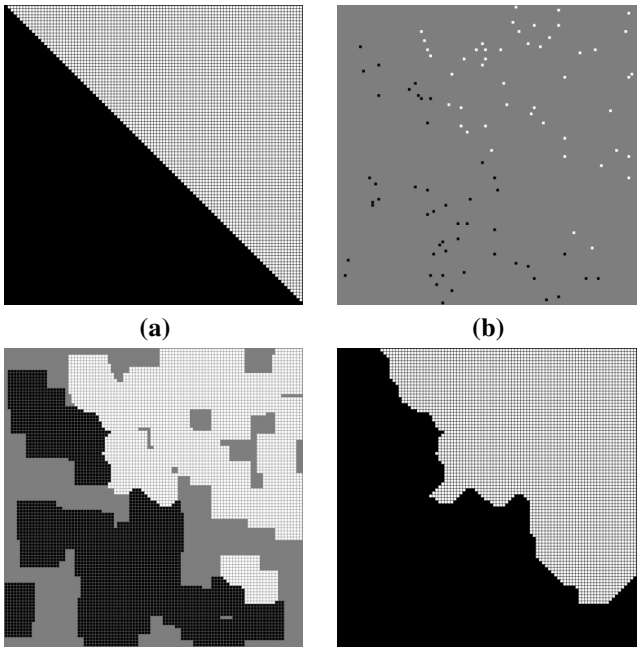
The $n4\_V1\_nonstable$ rule presented in [1] and newly proposed modifications will be examine on the sinusoidal testing sets (CA grids), shown in the Fig. 1.

In the Fig. 2(b, c and d) one can observe a classification process of the linear goal Fig. 2(a). In the classification was used $n4\_V1\_nonstable$ rule in CA size: 100×100. In the first step (see, Fig. 2(b)), 1% cells of known state (classified as class 1 - cells in black state or class 2 - cells in the white state) was randomly chosen from the linear goal. This initial configuration of CA was subject to the classification and the unclassified (cells in the gray state) with the use of $n4\_V1\_nonstable$ rule (temporary state of CA under classification process, see Fig. 2(c)). After the classification, the obtained result shows Fig. 2(d). The effectiveness of classification in presented example achieved level $93,38\%$ it means that 662 CA cells from 10000 CA cells have an incorrect classification.



(a) **(b)**

Fig. 1. Two-dimensional classifications of sinusoidal goals (examples): 100×100 **(b)** and 800×800 **(c)**.



(a) **(b)** **(c)** **(d)**

Fig. 2. An example of classification process with use of $n4\_V1\_nonstable$ rule in 2D CA with size 100×100 for linear goal **(a)**, initial configuration of CA - 1% of classified cells **(b)**, temporary CA state - classification in progress **(c)**, final CA state - after classification **(d)**.

Suppose, the $G_{N\times M}$ is the binary matrix of the classification goal. Also, the $F_{N\times M}$ is the binary matrix of the final configuration. The $E_{N\times M} = |G_{N\times M} - F_{N\times M}|$ is the absolute difference between two matrixes. Therefore, the Effectiveness (in %) is calculated by the formula: $Effectiveness = \frac{N*M-\sum_{i=1}^{N}\sum_{j=1}^{M} e_{(i,j)}}{N*M} * 100$, where $e_{(i,j)} \in E_{N\times M}$.

IV. Experimental Results

Proposed modifications of Fawcett's CA rule was tested and compared with original one. The results of analysis for the efficiency of classification are described above. During each test were used $N \times N$ CA, where $N \in \{100, 200, ..., 700, 800\}$, according to each of CA rules ($n4\_V1\_nonstable$, and newly proposed: $n4\_V1\_nonstable\_PP$, $n4\_V1\_nonstable\_FP$) with the von Neumann and Moore neighbourhood. In the goal, a 99% random states of CA cells was changed into the unclassified. So, only 1% CA cells stayed as classified. Such form of the goal was the initial configuration of CA to classification. After classification (reconstruction of goal) we obtained the finite CA state, and it was compared with the goal CA state; then the differences between both states (incorrect classified CA cells) and next ere calculated $Effectiveness$.

A. An analysis of the incorrect classifications for CA rules.

Classification in CA with the use of the Fawcett's rule, and new proposed probabilistic modification with von Neumann and Moore neighbourhood was realized in [8] for a linear goal during 1000 tests with above-presented conditions and assumptions. In [8] we can see that the effectiveness grows with the CA size, but it only depends on the fast growing area with the same class of data where the border with the different classes is increasing very slowly. The results showed that in general, the proposed modifications give better results in a classification for a linear goal than the original rule. For von Neumann neighbourhood, it could be observed for higher CA sizes (from 300 to 800) with both modifications in particular partially probabilistic one $n4\_V1\_nonstable\_PP$, where we can see that the effectiveness of new proposed modifications is better than for the original Fawcett's rule. Some kind anomaly one can see for CA size equal to 700×700 (see, Table 1 in [8]), where can be observed the best result of classification for the fully probabilistic classifier ($n4\_V1\_nonstable\_FP$), better than Fawcett's rule and much better than partially probabilistic one. For Moore neighbourhood both modification gives better results for each CA sizes except the size equal to 700×700 where Fawcett's rule classified effective (see, Table 2 in [8]).

In this paper is presented analysis mentioned above rules for sinusoidal goal (see, Fig. 1) with both von Neumann and Moore neighbourhood.

In the Tab. I one can see the results of classification for a sinusoidal goal in CA with the use of mentioned above three CA rules with Moore neighbourhood for 1000 tests. We can observe that the effectiveness of new proposed modifications is better than for the original Fawcett's rule for the sinusoidal goal. It could be observed for both modifications (marked in bold), in particular for fully probabilistic one ($n4\_V1\_nonstable\_FP$), except the tests for CA size equal to 400×400 and 600×600 (the underlined numbers mean the best results in general, from all analysed rules). We can also observe that effectiveness of classification with use of CA with Moore neighbourhood for a sinusoidal goal in much higher than for linear goal, it is easy to see in comparison the incorrect classification from Tab. I and Tab. 2 from [8].

TABLE I
CLASSIFICATION RESULTS OF GOAL SINUSOIDAL SET FOR CA RULES: $n4\_V1\_nonstable$, $n4\_V1\_nonstable\_PP$, $n4\_V1\_nonstable\_FP$ WITH MOORE NEIGHBOURHOOD. THE WORST EFFECTIVENESS AND (HIGHEST NUMBER OF INCORRECT CLASSIFIED CA CELLS) FROM 1000 TESTS.

CA size $N \times N$	**CA rule Fawcett's**	**CA rule partially probab.**	**CA rule fully probab.**
100×100	92,7% (730)	93,69% (**631**)	93,69% (**631**)
200×200	96,81% (1275)	97,1% (**1160**)	97,35% (**1062**)
300×300	98,21% (1613)	98,34% (**1493**)	98,4% (**1436**)
400×400	98,78% (1959)	98,82% (**1882**)	99,12% (1977)
500×500	99,04% (2388)	98,99% (2513)	99,12% (**2203**)
600×600	99,28% (**2606**)	99,22% (2817)	99,26% (2661)
700×700	99,37% (3076)	99,37% (3100)	99,4% (**2937**)
800×800	99,45% (3528)	99,47% (**3413**)	99,46% (**3465**)

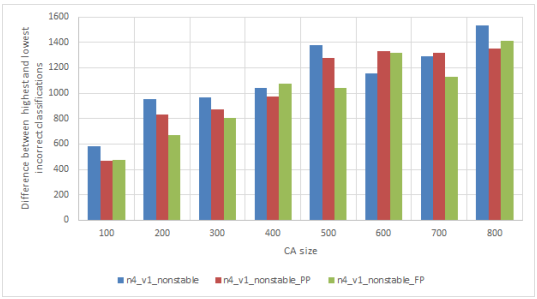
TABLE II
CLASSIFICATION RESULTS OF GOAL SINUSOIDAL SET FOR CA RULES: $n4\_V1\_nonstable$ (FAWCETT'S RULE), $n4\_V1\_nonstable\_PP$, $n4\_V1\_nonstable\_FP$ WITH VON NEUMANN NEIGHBOURHOOD. THE WORST EFFECTIVENESS AND (HIGHEST NUMBER OF INCORRECT CLASSIFIED CA CELLS) FROM 1000 TESTS.

CA size $N \times N$	**CA rule Fawcett's**	**CA rule partially probab.**	**CA rule fully probab.**
100×100	92,75% (725)	94,31% (**569**)	93,53% (**647**)
200×200	97,08% (1169)	97,43% (**1028**)	97,35% (**1061**)
300×300	98,44% (1403)	98,35% (1482)	98,44% (**1402**)
400×400	98,9% (1757)	98,92% (**1732**)	98,78% (1952)
500×500	99,13% (2184)	99,16% (**2102**)	99,08% (2310)
600×600	99,3% (**2507**)	99,28% (2588)	99,29% (2561)
700×700	99,41% (**2875**)	99,39% (2974)	99,41% (2913)
800×800	99,49% (3259)	99,49% (**3246**)	99,48% (3342)

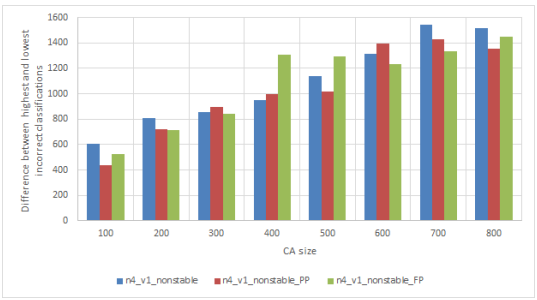
For the rule $n4\_V1\_nonstable\_FP$ the difference between number incorrect classification in linear and sinusoidal goal is equal to $\{210, 432, 436, 526, 726, 1008, 1171, 923\}$ for CA sizes $\{100, ..., 800\}$, and also for $n4\_V1\_nonstable\_PP$ differences are equal to $\{185, 167, 402, 564, 471, 617, 921, 1086\}$ respectively.

Similar results but not so high, are obtained for classification the sinusoidal goal with this three methods for CA with von Neumann neighbourhood. In the Tab. II one can see the results of classification for a sinusoidal goal in CA with the use of mentioned above three CA rules with von Neumann neighbourhood for 1000 tests. We can observe that the effectiveness of partially probabilistic modification ($n4\_V1\_nonstable\_PP$) is better than other rules for the sinusoidal goal (marked in bold and underlined). It can be observed for tests realised for most of analysed CA sizes (compare, Tab. 1 from [8] and Tab. II).

The observed highest effectiveness for modification of Fawcett's rule maybe is not so spectacular, but we should interpret it from the another point of view; it means the number of incorrectly classified CA cells should be analyzed. One can see that the modifications have in general the lowest number of incorrectly classified CA cells than the original rule (see, numbers in bold in Tab. II, Tab. I also Tab. 1 and Tab. 2 from [8]). For example, for the linear goal in the Tab. I for CA size 200×200, the fully probabilistic modi-



(a)



(b)

Fig. 3. Ranges of scattering for incorrect classifications for 1000 tests with a sinusoidal goal: **(a)** Moore neighbourhood, **(b)** von Neumann neighbourhood.

fication ($n4\_V1\_nonstable\_FP$) has 213 CA cells less than the original rule, and also partially probabilistic modification ($n4\_V1\_nonstable\_FP$) has 115 CA cells less.

So, the analysis of the worst effectiveness and highest number of incorrect classification for data presented above in Tables leads to the conclusion; if the worst results for classification with use of proposed modifications are better than for the original rule, then the new classifiers are more efficient because their the worst classifications are much better.

B. An analysis of the scattering ranges for incorrect classifications for CA rules.

The scattering range is the difference between the highest number of incorrect classifications and the lowest number of incorrect classification. The scattering ranges were calculated for each analysed CA rules for linear and sinusoidal goals with the use of both neighbourhoods von Neumann and Moore. The highest and lowest number of incorrect classifications were selected from 1000 tests for random initial configuration of CA.

We can observe in the Fig. 3 the scattering ranges for sinusoidal goal with Moore neighbourhood Fig. 3(a) and von-Neumann neighbourhood Fig. 3(b). One can see in Fig. 3(a) the scattering ranges for newly proposed modified CA rules are shorter in general than for original rule, in particular for fully probabilistic modification ($n4\_V1\_nonstable\_FP$) with the use of Moore neighbourhood. Moreover, remembering such rule has, in general, the lowest incorrect classifications we can conclude that fully probabilistic classifier shows up the more accurate and consistent with Moore neighbourhood for the sinusoidal goal.

Similarly, in Fig. 3(b) can be observed in general the shortest scattering ranges for proposed modification in particular for partially probabilistic modification ($n4\_V1\_nonstable\_FP$) of CA rule with the use of von Neumann neighbourhood.

For the linear goal, also we can observe the shortest scattering ranges in most cases of CA size, for the proposed new classifiers (see, [8] in particular the Moore neighbourhood).

V. Conclusions and Future Works

In the paper was presented problem of classification with use of three-state two-dimensional cellular automaton. Among classifiers were analysed the wide known Fawcett's rule and two proposed probabilistic modifications of such rule. The conducted experiments show the better effectiveness for classification applying a newly proposed classifiers (modifications) to reconstruction goals from state consisted of only 1% classified states. Moreover, the proposed modifications result in a much lower number of incorrectly classified CA cells.

Also, the analysis the data from Tab. I, Tab. II; Table 1 and Table 2 from [8] leads to the conclusion that seems to be a weak relationship between kind of modification and type of CA neighbourhood. For the von Neumann neighbourhood better classifications have obtained for partially probabilistic modification, and for Moore neighbourhood fully probabilistic modification are better.

Moreover, from Fig. 3 and Fig. 2 from [8] ensue in general the shortest scattering range for proposed modifications of Fawcett's rule than the original one. From that and above conclusions, we can understand that proposed probabilistic classifiers characterized by more accurate and consistent classification.

In the future works the newly proposed rules will also be examined with use of new testing goals like e.g.: parabolic, closed area (circular, square, concave boundary, ...), disjunctive and other. Also, will be analysed the effectiveness of classification depending on the number of unclassified states of CA cells initial configuration.

References

[1] T. Fawcett, "Data mining with cellular automata." *ACM SIGKDD Explorations Newsletter*, 10(1), pp. 32–39, 2008.
[2] P. Gacs, G. Kurdyumov, and L. Levin, "One dimensional uniform arrays that wash out finite islands." *Problemy Peredachi Informatsii*, 12, pp. 92–98, 1978.
[3] M. Mitchell, P. Hraber and J. Crutchfield, "Revisiting the edge of chaos: Evolving cellular automata to perform computations." *Complex Systems*, 7, pp. 89–130 1993.
[4] P. Maji, B. Sikdar, and P. Chaudhuri, "Cellular automata evolution for pattern classification." *Lecture Notes in Computer Science 3305* , pp. 660–669. Springer Verlag 2004.
[5] S. Omohundro, "Modelling cellular automata with partial differential equations." *Physica 10D*, 10(1-2), pp. 128–134, 1984.
[6] A. Piwonska, F. Seredynski, M. Szaban, "Learning Cellular Automata Rules for Binary Classification Problem," *The Journal of Supercomputing*, Volume 63, Issue 3, pp. 800–815, 2013.
[7] P. Povalej, M. Lenic, and P. Kokol, "Improving ensembles with classificational cellular automata." *Lecture Notes in Computer Science 3305*, pp. 242–249. Springer Verlag 2004.
[8] M. Szaban, "Probabilistic Binary Classification with Use of 2D Cellular Automata." *12th International conference on Cellular Automata for Research and Industry - ACRI 2016*, 2016 (in print).
[9] S. Wolfram: *A New Kind of Science*. Wolfram Media, 2002.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 165–174
DOI: 10.15439/2016F559
ACSIS, Vol. 8. ISSN 2300-5963

Generalized Majority Decision Reducts

Sebastian Widz* and Sebastian Stawicki†
*Systems Research Institute, Polish Academy of Sciences
ul. Newelska 6, 01-447 Warsaw, Poland
†Institute of Mathematics, University of Warsaw
ul. Banacha 2, 02-097 Warsaw, Poland
sebastian.widz@gmail.com, sebastian.stawicki@gmail.com

***Abstract*—We discuss several new methods for constructing approximate decision reducts from the rough set theory. We introduce generalized approximate majority decision reducts, which are an extension to standard approximate decision reducts known from literature but with improved calculation performance and complexity. We also discuss the relationship and differences of the new approximate decision reduct notion with so called decision bireducts – another type of approximate decision reducts.**

***Keywords*-Feature Subset Selection; Rough Sets; Approximate Decision Reducts; Decision Bireducts**

I. Introduction

ATTRIBUTE subset selection plays an important role in knowledge discovery [1]. It establishes the basis for more efficient classification, prediction and approximation models. It also provides the users with a better insight into data dependencies. In this paper, we concentrate on attribute subset selection methods originating from the theory of rough sets [2]. There are numerous rough-set-based algorithms aimed at searching for so called *reducts* – irreducible subsets of attributes that satisfy predefined criteria for keeping enough information about decisions. Those criteria are verified on the training data and, usually, they encode the risk of misclassification by if-then decision rules with their antecedents referring to the values of investigated attribute subsets and their consequents referring to decisions.

Original definition of a reduct is quite restrictive, requiring that it should determine decisions or, if data inconsistencies do not allow full determinism, provide the same level of information about decisions as the complete set of attributes. There are a number of approaches to formulate and search for approximate or inexact reducts, which *almost* preserve the decision information [3]. Approximate reducts are usually smaller than standard ones, providing the basis for learning more efficient classifiers [4], [5].

In our previous work [6] we compared approximate decision reducts based on feature subset quality functions [7], [8] with decision bireducts [9] – another extension to the rough set framework of decision reducts. In short, a decision bireduct is an irreducible feature subset which preserves the information about the decision but only on selected objects from the data set. The feature subset cannot be reduced as well as the object subset cannot be further expanded without violating the decision functional dependence.

In this article we show new methods for calculating approximate decision reducts based on generalized majority decision (GMD) function. The concept of GMD function might be considered as an modification to a very well known concept of generalized decision function known from the rough set theory. We also consider a few variations of decision rule-based classifiers induction from the discussed approximate decision reducts. We introduce the concept of *exception* decision rules which are created during the reduct calculation process. The use of *exceptions* can be helpful in case we need to construct simplified decision models [10] and preserve more information about the original data set. We also explain the relationship of the proposed methods to the notion of decision bireducts and propose a new algorithm for decision bireducts calculation based on GMD function.

The remainder of this paper is organized as follows. Section II briefly describes the concept of generalized decision function. In Section III we explain approximate decision reducts. In Sections IV and V we present a new definition of approximate decision reducts and show new algorithms in Section VI. In Section VII we introduce the concept of *exception* decision rules. In Section VIII we show relationship to the decision bireducts and a new way of decision bireduct calculation. Finally, we show experiment results in Section IX and conclude our work in Section X.

II. Decision Reducts and Generalized Decision

We use the standard notion of a decision system to represent data [2]. By a decision system we mean a tuple $\mathbb{A} = (U, A \cup \{d\})$, where U is a set of objects and A is a set of attributes and $d \notin A$ is a distinguished decision attribute. For simplicity, we refer to the elements of U using their ordinal numbers $i = 1, ..., |U|$, where $|U|$ denotes the cardinality of U. We treat attributes $a \in A$ as functions $a : U \rightarrow V_a$, V_a denoting a's value domain. The values $v_d \in V_d$ correspond to decision classes that we want to describe using the values of attributes in A. Each subset $B \subseteq A$ partitions the space U into *equivalence classes*. For such a division we get the partition space denoted as $U/B = \{E_1, \ldots, E_t\}$ where $E_t \subseteq U$. Each equivalence class is defined as $E_t = \{x \in U : B(x) = v_t\}$ where v_t is a vector of values on attributes B. We will further refer the set of objects with particular decision k as X_k.

Definition 1. *We say that $B \subseteq A$ is a decision reduct for $\mathbb{A} = (U, A \cup \{d\})$ if and only if it is an irreducible subset*

978-83-60810-90-3/$25.00

of features such that each pair $i, j \in U$ satisfying inequality $d(i) \neq d(j)$ is discerned by B.

Definition 1 works well in case of consistent decision tables, however in case of inconsistent data no attribute reduction is possible, in fact, even the whole attribute set A cannot be considered to be a decision reduct itself. There are many alternative definitions of decision reducts that could be applied to inconsistent decision tables, e.g., subsets of features that preserve the same positive region as the full set of attributes. Another example could be a decision reduct based on generalized decision function [2]:

Definition 2. *For a given decision table $\mathbb{A} = (U, A \cup \{d\})$, we say that a **generalized decision function** is a function $\partial_d : 2^U \to 2^{V_d}$ defined as follows:*

$$\partial_d(E) = \{k : X_k \cap E \neq \emptyset\} \tag{1}$$

The cardinality of a generalized decision may be used to express the level of inconsistency in describing decision attribute by subsets of features. In particular, if $|\partial_d([x]_A)| = 1$, for any $x \in U$, then the decision table $\mathbb{A} = (U, A \cup \{d\})$ is said to be *consistent*. Otherwise it is *inconsistent*. The example of a generalized decision is presented in Table I in column denoted as $\partial_d([x]_A)$.

Definition 3. *Let $\mathbb{A} = (U, A \cup \{d\})$ be given. We say that $B \subseteq A$ is a ∂-decision superreduct if and only if the following condition holds:*

$$\bigvee_{x \in U} \partial_d([x]_B) = \partial_d([x]_A) \tag{2}$$

We say that B is a ∂-decision reduct if and only if it is a ∂-superreduct and none of its proper subsets satisfy the above condition.

III. Foundations of Approximate Reducts

There are a variety of methods of searching for approximate decision reducts (e.g. [11], [12], [13]). The criteria usually include formulas for functions measuring degrees of decision information induced by subsets of features and thresholds for those functions' values specifying which subsets of attributes are *good enough*. The choice of functions may depend on the nature of particular data sets and methods of learning classifiers based on reduced sets of attributes. In order to follow the filter approach to feature subset selection, we need to design some measures that evaluate particular feature subsets in the selection process. From this point of view, the rough set literature may be regarded as a source of measures that draw correspondence between feature subsets and rule-based classifiers corresponding to those subsets, where each subset of attributes $B \subseteq A$ yields a set of decision rules based on all combinations of its values in U and the consecutive value is chosen according to some criteria.

Let us define a measure $F : P(A) \to \Re$ which evaluates the degree of influence $F(B)$ of subset $B \subseteq A$ in d. Then one can decide which features may be removed from A without significant loss of accuracy.

Definition 4. *Let $\mathbb{A} = (U, A \cup \{d\})$ and approximation threshold $\varepsilon \in [0, 1)$ be given. We say that $B \subseteq A$ is an (F, ε)-approximate decision reduct if and only if it satisfies the following condition:*

$$F(B) \geq (1 - \varepsilon)F(A) \tag{3}$$

and none of its proper subsets $C \subseteq B$ does it.

There are many examples how function F can be defined. Let us focus on two examples. The first is *Majority* measure (further denoted as M), proposed in [7]:

$$M(B) = \sum_{E \in U/B} \frac{|E|}{|U|} \max_{k \in V_d} \frac{|X_k \cap E|}{|E|} \tag{4}$$

The function M points to a decision value that appears the most frequently within a particular equivalence class E. It means that if we need to decide which decision should be attached to objects from a particular class, then we always choose the most frequent decision that was observed in the training data. Based on the measure M, one could generate decision rules for which the rules' right sides are the most frequent decisions for $E \in U/B$. Another example of F refers to a Bayesian extension of the classical rough set model, where rules induced by a given subset of attributes are pointing at the decision classes which become maximally frequent comparing to their overall occurrence in the data. This function called *Relative information gain* measure (further denoted as R) was proposed in [8].

$$R(B) = \frac{1}{|V_d|} \sum_{E \in U/B} \max_{k \in V_d} \frac{|X_k \cap E|}{|X_k|} \tag{5}$$

In our research the permutation based *REDORD* algorithm introduced in [14] is used as a baseline algorithm (see Algorithm 1).

Algorithm 1 Permutation-based (F, ε)-REDORD Algorithm

Input: $\mathbb{A} = (U, A \cup \{d\})$, $\varepsilon \in [0, 1)$,
$\sigma : \{1, ..., n\} \to \{1, ..., n\}$, $n = |A|$
Output: (F, ε) approximate decision reduct $B \subseteq A$

```
B ← A
for i = 1 → n do
    if F(B \ {a_σ(i)}) ≥ (1 − ε)F(A) then
        B ← B \ {a_σ(i)}
    end if
end for
return B
```

IV. Generalized Majority Decision Reducts

Approximate decision reducts calculation methods allow some level of interaction with users designing decision models. However, this interaction is only limited to a single parameter describing the degree of approximation that refers to the allowed overall misclassification rate of a decision model. What is more restricting is that the decision model based on

approximate decision reduct is always pointing to majority decision within particular equivalence classes. We would like to give the user the ability to decide about the decision classes distribution within each E. Based on the idea of generalized decision we formulate *generalized majority decision* (described in this section) and *generalized approximation majority decision* concepts. The approximate version is described in the next section.

Definition 5. *For a given decision table $\mathbb{A} = (U, A \cup \{d\})$, we say that a **generalized majority decision function** is a function $m_d : 2^U \rightarrow 2^{V_d}$ defined as follows:*

$$m_d(E) = \{k : |X_k \cap E| = \max_j(|X_j \cap E|)\} \tag{6}$$

Generalized majority decision function reflects the choice of most frequent decision classes within subsets $E \subseteq U$. In Table I we present an example of generalized decision and generalized majority decision attributes denoted $\partial_d([x]_A)$ and $m_d([x]_A)$ respectively.

Table I
DECISION TABLE AND GENERALIZED DECISION AND GENERALIZED MAJORITY DECISION ATTRIBUTES

Id	a_1	a_2	a_3	a_4	d	$\partial_d([x]_A)$	$m_d([x]_A)$
x_1	1	1	2	2	0	{0,1}	{0,1}
x_2	1	1	2	2	1	{0,1}	{0,1}
x_3	1	1	2	2	1	{0,1}	{0,1}
x_4	1	1	2	2	0	{0,1}	{0,1}
x_5	3	3	1	2	1	{0,1}	{0,1}
x_6	3	3	1	2	0	{0,1}	{0,1}
x_7	2	3	1	2	1	{1}	{1}
x_8	1	2	2	1	2	{0,1,2}	{1,2}
x_9	1	2	2	1	2	{0,1,2}	{1,2}
x_{10}	1	2	2	1	1	{0,1,2}	{1,2}
x_{11}	1	2	2	1	1	{0,1,2}	{1,2}
x_{12}	1	2	2	1	0	{0,1,2}	{1,2}
x_{13}	2	1	1	1	1	{1}	{1}
x_{14}	2	2	1	1	0	{0}	{0}

In case of consistent decision tables $|m_d([x]_A)| = 1$. In case of inconsistent decision tables the cardinality of $m_d([x]_A)$ will be greater than one only in a case of an equivalence class with more than one equally distributed decision values. However, let us notice that one could be interested in calculating $m_d([x]_B)$ for any subset of features $B \subseteq A$ containing the smaller number of attributes. In such a case the equivalence classes are relatively large, and equal frequencies of decision values are more common.

Definition 6. *Let $\mathbb{A} = (U, A \cup \{d\})$ be given. We say that $B \subseteq A$ is an $(m, =)$-decision superreduct if and only if the following condition holds:*

$$\bigwedge_{x \in U} m_d([x]_B) = m_d([x]_A) \tag{7}$$

We say that B is an $(m, =)$-decision reduct if and only if it is an $(m, =)$-superreduct and none of it proper subsets satisfy the above condition.

Proposition 1. *Let $\mathbb{A} = (U, A \cup \{d\})$ be given. If $B \subseteq A$ is an $(m, =)$-decision superreduct, then:*

$$\forall_{x,y \in U} m_d([x]_A) \neq m_d([y]_A) \Rightarrow \exists_{a \in B} a(x) \neq a(y) \tag{8}$$

Proof: First let us notice that we can transform the above equation to the following form:

$$\forall_{x,y}\Big(\forall_{a \in B} a(x) = a(y) \Rightarrow m_d([x]_A) = m_d([y]_A)\Big) \tag{9}$$

Let us consider any x, y such that $\bigwedge_{a \in B} a(x) = a(y)$. We need to show that $m_d([x]_A) = m_d([y]_A)$. Based on Definition (6) we know that $m_d([x]_A) = m_d([x]_B)$ and that $m_d([y]_A) = m_d([y]_B)$. Because $[x]_B = [y]_B$, then $m_d([x]_B) = m_d([y]_B)$, this also $m_d([x]_A) = m_d([y]_A)$. Let us take any $x \in U$. We need to show that $m_d([x]_B) = m_d([x]_A)$. Let us take any $v_k \in m_d([x]_A)$. We need to show that $v_k \in m_d([x]_B)$. If $v_k \in m_d([x]_A)$, then based on Equation (9) we have $v_k \in m_d([u]_A)$ for all $y \in [x]_B$. This means, that v_k has maximal frequency among all decision classes for all A-indiscernibility classes contained in $[x]_B$. This means that is also has maximum frequency in $[x]_B$. Thus:

$$m_d([x]_B) \subseteq m_d([x]_A) \tag{10}$$

always holds.
In Proposition (3) we will show that if the intersection of $m_d([y]_A)$ sets for all $y \in [x]_B$ for a given $x \in U$ is not empty, then the following holds:

$$m_d([x]_B) \subseteq m_d([x]_A) \tag{11}$$

Based on Equation (9) we know that this intersection is nonempty. Based on Equations (9) and (10) we have

$$m_d([x]_B) = m_d([x]_A) \tag{12}$$

■

Let us propose also another method of using the m_d function to define attribute reduction criteria, where we are not interested in the full discernibility with regard to the symbolic representation of generalized majority decision but we apply a weaker condition.

Definition 7. *Let $\mathbb{A} = (U, A \cup \{d\})$ be given. We say that $B \subseteq A$ is an $(m, \subseteq)$-decision superreduct if and only if the following condition holds:*

$$\bigwedge_{x \in U} m_d([x]_B) \subseteq m_d([x]_A) \tag{13}$$

We say that B is an $(m, \subseteq)$-decision reduct if and only if it is an $(m, \subseteq)$-superreduct and none of it proper subsets satisfy the above condition.

The following result emphasizes the importance of Definition 7.

Proposition 2. *Let $\mathbb{A} = (U, A \cup \{d\})$ be given. $B \subseteq A$ is $(m, \subseteq)$-decision superreduct if and only if the following condition holds:*

$$M(B) = M(A) \tag{14}$$

Let us further notice that Definition 7 is more flexible than Definition 6 and it usually allows more attribute reduction.

Proposition 3. *Let $\mathbb{A} = (U, A \cup \{d\})$ be given. $B \subseteq A$ is an $(m, \subseteq)$-decision superreduct if and only if:*

$$\bigwedge_{x \in U} \bigcap_{y \in [x]_B} m_d([y]_A) \neq \emptyset \tag{15}$$

Proof: Let us take any $x \in U$ and let us consider all objects $y \in [x]_B$. We know that:

$$\bigwedge_{y \in [x]_B} m_d([y]_B) \subseteq m_d([y]_A) \tag{16}$$

Because $y \in [x]_B$, then $[y]_B = [x]_B$. Thus:

$$\bigwedge_{y \in [x]_B} m_d([x]_B) \subseteq m_d([y]_B) \tag{17}$$

and therefore:

$$\bigwedge_{y \in [x]_B} m_d([x]_B) \subseteq \bigcap_{y \in [x]_B} m_d([y]_A) \tag{18}$$

Thus, $\bigcap_{y \in [x]_B} m_d([y]_A)$ cannot be empty.
Let us take any $x \in U$ and let us consider all objects $y \in [x]_B$. Because, we know that:

$$\bigcap_{y \in [x]_B} m_d([y]_A) \neq \emptyset \tag{19}$$

we also know that there exists such v that $v \in m_d([y]_A)$ for every $y \in [x]_B$. By Definition (5) we know that v must be one of the most frequent decision values on all equivalence classes $[y]_A$ such that $y \in [x]_B$. This is because $[x]_B$ is a sum of some disjoint blocks $E \in U/A$, which are represented by $[y]_A$ For that reason it must be also one of the most frequent values in the whole $[x]_B$ i.e. $v \in m_d([x]_B)$. For any other decision v' contained in $m_d([x]_B)$, it would have to be as frequent in $[x]_B$ as the decision value v. But if would require for v' to be as frequent as v in all blocks $E_A \in U/A, E_A \subseteq E_B$. Thus, it would be also $v' \in m_d(E_A)$ for all such blocks. In particular we would have:

$$m_d([x]_B) \subseteq m_d([x]_A) \tag{20}$$

■

Proposition 4. *Let $\mathbb{A} = (U, A \cup \{d\})$ be given. $B \subseteq A$ is an $(m, \subseteq)$-decision superreduct if and only if:*

$$\bigwedge_{x \in U} \bigcap_{y \in [x]_B} m_d([y]_A) = m_d([x]_B) \tag{21}$$

Let us illustrate the difference between $(m, \subseteq)$- and $(m, =)$-decision reducts. Let us consider a decision table presented in Table II. All three objects have exactly the same values on feature subset B but different values on attribute a. According to definition of $(m, =)$-decision reduct attribute a cannot be removed. $(m, =)$-decision reduct requires that generalized majority decision sets are exactly the same for merged equivalence classes. In case of $(m, \subseteq)$-decision reduct this attribute can be removed as it requires only set inclusion.

Table II
ATTRIBUTE REMOVAL IN $(m, \subseteq)$- AND $(m, =)$-*decision reducts*.

Id	B	a	$m_d([x]_{B \cup \{a\}})$
x_1		1	{0,1}
x_2	...	2	{0,1,2}
x_3		3	{0,2}

V. GENERALIZED APPROXIMATE MAJORITY DECISION

The generalized majority decision function points only to a single most frequent decision for a given equivalence class. Its drawback is that it does not keep any information about minority decision classes and there is no way to control the threshold how much information about the minority decision classes should be maintained. We propose to determine this threshold based on decision class distribution within equivalence classes and a level of approximation given by the user.

Let us introduce a definition of *generalized approximate majority decision* function. In comparison to Equation 1 we include not only the most frequent decision values but also those which are almost as frequent as the majority decision values. We express this statement by introducing a threshold parameter ε that controls the ratio between allowed decision values and the majority decision.

Definition 8. *For a given decision table $\mathbb{A} = (U, A \cup \{d\})$ and approximation threshold $\varepsilon \in [0, 1)$, we say that a **generalized approximate majority decision function** is a function $m_d^\varepsilon : 2^U \to 2^{V_d}$ defined as follows:*

$$m_d^\varepsilon(E) = \{k : |X_k \cap E| \geq (1 - \varepsilon) \max_j |X_j \cap E|\} \tag{22}$$

Proposition 5. *For $\varepsilon = 0$ and $\varepsilon \to 1^-$ we have:*

$$\bigwedge_E m_d^0(E) = m_d(E) \qquad \text{and} \qquad \bigwedge_E \lim_{\varepsilon \to 1^-} m_d^\varepsilon(E) = \partial_d(E)$$

Let us propose an alternative way of the attribute reduction based on the new concept of $(m^\varepsilon, \cap)$-decision reduct and generalized decision set intersections. Let us further notice that this approach is equivalent to the previous one for $\varepsilon = 0$.

Definition 9. *Let $\mathbb{A} = (U, A \cup \{d\})$ be given. We say that $B \subseteq A$ is an $(m^\varepsilon, \cap)$-decision superreduct if and only if:*

$$\bigwedge_{x \in U} \bigcap_{y \in [x]_B} m_d^\varepsilon([y]_A) \neq \emptyset \tag{23}$$

We say that B is an $(m^\varepsilon, \cap)$-decision reduct if and only if it is an $(m^\varepsilon, \cap)$-decision superreduct and none of it proper subsets satisfy the above condition.

Proposition 6. *Let $\mathbb{A} = (U, A \cup \{d\})$ be given. The following holds for every subset $B \subseteq A$:*

$$\bigwedge_{x \in U} \bigcap_{y \in [x]_B} m_d^\varepsilon([y]_A) \subseteq m_d^\varepsilon([x]_B) \tag{24}$$

Proposition 7. *If $B \subseteq A$ in an $(m^\varepsilon, \cap)$-decision superreduct then the following inequality holds:*

$$M(B) \geq (1 - \varepsilon) M(A) \tag{25}$$

Let us notice that in case of Proposition 2 we have the equivalence between of $B \subseteq A$ being an $(m, \subseteq)$-decision reduct and given the equality $M(B) = M(A)$, whereas in case of Proposition 7 and $(m^{\varepsilon}, \cap)$-decision reduct we have only the implication.

In the traditional decision rule induction approach one can use contents of a decision table and the set of attributes B to induce decision rules. Similarly, decision rules calculated with respect to the generalized majority decisions take such a form that the ascendants are based on values of B over objects $x \in U$ but consequents point at disjunctions of possible decisions belonging to $m_d^{\varepsilon}(E)$. For example, let us consider objects $\{x_8, x_9, x_{10}, x_{11}, x_{12}\}$ from Table I. Based on these objects we can produce the following decision rule:

$$a_1 = 1 \wedge a_2 = 1 \wedge a_3 = 2 \wedge a_4 = 1 \Rightarrow d = 1 \vee d = 2$$

It is important to note that such a rule could be rewritten as two separate rules, both having the same support and confidence:

$$a_1 = 1 \wedge a_2 = 2 \wedge a_3 = 2 \wedge a_4 = 1 \Rightarrow d = 1$$
$$a_1 = 1 \wedge a_2 = 2 \wedge a_3 = 2 \wedge a_4 = 1 \Rightarrow d = 2$$

In fact all decision values belonging to generalized approximate majority decision set are treated as equally distributed within a particular equivalence class. Moreover, all decision rules based on this equivalence class are considered to have equal quality support and confidence.

VI. Heuristic Search for Approximate Reducts

The above results could be utilized in a procedure for attribute reduction. First we calculate generalized majority decision for the whole attribute set A. Next we can remove the attribute a whenever the Equality 23 is satisfied. This means that if the intersection of all generalized decision of smaller equivalence classes induced by $B \cup \{a\}$ is nonempty we can remove attribute a without a significant loss of information. Algorithm 2 presents the procedure that utilizes *Reduce* function described in Algorithm 3. In Table III and Table IV we give brief examples of attribute removal tries, the unsuccessful and successful, respectively.

Table III
Impossible removal of attribute a_2

Id	a_1	a_3	a_4	$m_d^{\varepsilon}([x]_B)$	
$\{x_1, ..., x_4\}$	1	2	2	{0,1}	✓
$\{x_5, x_6\}$	3	1	2	{0,1}	✓
$\{x_7\}$	2	1	2	{1}	✓
$\{x_8 ... x_{12}\}$	1	2	1	{1,2}	✓
$\{x_{13} \cup x_{14}\}$	2	1	1	$\{1\} \cap \{0\} = \emptyset$	✗

Let us focus on decision rules that could be generated before and after removal of attribute a_1 based on objects $\{x_5, x_6, x_7\}$ Before removal of attribute a_1 that rules are as follow:

$$a_1 = 3 \wedge a_2 = 3 \wedge a_3 = 1 \wedge a_4 = 2 \Rightarrow d = 0 \vee d = 1$$
$$a_1 = 2 \wedge a_2 = 3 \wedge a_3 = 1 \wedge a_4 = 2 \Rightarrow d = 1$$

Algorithm 2 Generalized Majority Decision Reduct (GMDR2)

Input: $\mathbb{A} = (U, A \cup \{d\})$, $\varepsilon \in [0, 1)$, $\sigma : \{1, ..., n\} \to \{1, ..., n\}$, $n = |A|$,
$\mathbb{A}_{temp}$ - temporary decision table for storing equivalence classes)
Output: $B \subseteq A$

```
Calculate Generalized Majority Decision m_d^ε(E_A) for all objects in 𝔸
B ← A
E_B ← CreateEquivalenceClasses(A)
for i = 1 → n do
    E_C ← Reduce(E_B, B, a_σ(i))
    if E_C ≠ E_B then
        E_B = E_C
        B ← B \ {a_σ(i)}
    end if
end for
return B
```

Algorithm 3 Attribute reduction function

Input: Collection of equivalence classes $E_B \in U/B$,
Attribute subset $B \subseteq A$, Candidate attribute $a \in B$ for removal
Output: Collection of equivalence classes $E_C \in U/C$ where $C \subseteq B$ if attribute $a \in B$ was removed or $E_B \in U/C$ otherwise

```
function REDUCE(E_B, B, {a})
    C ← B \ {a}
    E_C ← ∅
    for all EquivalenceClasses e ∈ E_B do
        DEC ← ∅
        v_B ← GetInstance(e)
        v_C ← Remove(v_B, {a})
        e_tmp ← Find(E_C, v_C)
        if e_tmp ≠ NULL then
            DEC ← GetDec(e_tmp) ∩ GetDec(e)
            if |DEC| > 0 then
                SetDec(e_tmp, DEC)
            else
                return E_B
            end if
        else
            AddEquivalenceClass(E_C, e_tmp)
        end if
    end for
    return E_C
end function
```

After removal of attribute a_1:

$$a_2 = 3 \wedge a_3 = 1 \wedge a_4 = 2 \Rightarrow d = 1$$

From this perspective we can understand the described procedure as creating a lowe approximation of decision rules generated on the whole attribute set.

Table IV
SUCCESSFUL REMOVAL OF ATTRIBUTE a_1

Id	a_2	a_3	a_4	$m_d^{\varepsilon}([x]_B)$	
$\{x_1, ...x_4\}$	1	2	2	{0,1}	✓
$\{x_5, x_6\} \cup \{x_7\}$	3	1	2	$\{0,1\} \cap \{1\} = \{1\}$	✓
$\{x_8 ... x_{12}\}$	2	2	1	$\{1,2\}$	✓
$\{x_{13}\}$	1	1	1	$\{1\}$	✓
$\{x_{14}\}$	2	1	1	$\{0\}$	✓

VII. APPROXIMATE DECISION REDUCTS WITH EXCEPTIONS

Let us propose an another way of expressing levels of allowed inconsistency in the decision model. We introduce a new approximation threshold ϕ that relates to the maximal ratio of objects that can be misclassified. By analogy, we introduce a new definition of $(m^{\phi}, \cap)$-decision reduct which utilizes the generalized majority decision function.

Definition 10. *Let $\phi \in [0,1)$ and $\mathbb{A} = (U, A \cup \{d\})$ be given. Subset $B \subseteq A$ is an $(m^{\phi}, \cap)$-decision superreduct if and only if there exists subset $X \subseteq U$ satisfying inequality $|X| \geq (1-\phi)|U|$, such that the following condition holds:*

$$\bigvee_{x \in X} \bigcap_{y \in [x]_B \cap X} m_d([y]_A) \neq \emptyset \tag{26}$$

We say that B is an $(m^{\phi}, \cap)$-decision reduct if and only if it is an $(m^{\phi}, \cap)$-decision superreduct and none of its proper subsets satisfy the above conditions.

The idea behind $(m^{\phi}, \cap)$-decision reducts $B \subseteq A$ is to cover sufficiently the data set with decision rules constructed using attributes in B and provide a comparable level of information to decision rules constructed using all features from A. If $B \subseteq A$ is an $(m^{\phi}, \cap)$-decision superreduct, then there exists subset $X \subseteq U$ satisfying inequality $|X| \geq (1-\phi)|U|$ and condition (26) such that X is definable by A, which means that X is the set-theoretic sum of some equivalence classes of U/A. This explains in what sense subsets $X \subseteq U$ can be treated as set-theoretic sum of supports of premises of rules induced by A whose information is kept at comparable level after shortening them to more general decision rules induced by B. Thus, from now on, we will implicitly assume that all subsets $X \subseteq U$ discussed in context of $(m^{\phi}, \cap)$-decision reducts take a form of sums of some equivalence classes of U/A.

Proposition 8. *If $B \subseteq A$ is an $(m^{\phi}, \cap)$-decision superreduct, then the following inequality holds*

$$M(B) \geq M(A) - \phi \tag{27}$$

Let us notice that we could combine definitions of $(m^{\phi}, \cap)$-decision reduct with $(m^{\varepsilon}, \cap)$-decision reduct and formulate a combined definition of $(m^{\varepsilon,\phi}, \cap)$-decision reduct. For simplicity, we will further assume that $\varepsilon = 0$.

In Section IV we discussed an example showing that the attribute a_2 cannot be removed. This was due to the fact that objects x_{13} and x_{14} have an empty intersection of their generalized majority decision sets. In other words, there was only one object that was blocking the algorithm from reducing the number of features. If we removed either object x_{13} or object x_{14} from the discussed decision table the reduction would be possible but we would also lose some information. Exception rules allow us to reduce the number of features and at the same time to save additional information about lost objects. Let us focus on Tables V and VI where we again consider the same example but this time we can remove attribute a_2. We create the following exception rule (supported only by one object ($\{x_{14}\}$) but with confidence = 1.0)

$$a_1 = 2 \wedge a_2 = 1 \wedge a_3 = 1 \wedge a_4 = 1 \Rightarrow d = 0$$

In the next step we try to remove another attribute, again if some objects do not allow attribute removal we can remove them and save in form of exception rules, remembering that no more than $\phi \cdot |U|$ objects can be removed in total. This procedure is presented as Algorithm 4.

During the classification phase, the exception rules are always searched in the first place. If a proper exception rule is

Algorithm 4 Attribute reduction function with exceptions

Input: Collection of equivalence classes $E_B \in U/B$, Approximation degree $\phi \in [0,1)$; Attribute subset $B \subseteq A$, Candidate attribute $a \in B$ for removal

Output: Updated exception rules set R_{ex}, Collection of equivalence classes $E_C \in U/C$ where $C \subseteq B$ if attribute $a \in B$ was removed or $E_B \in U/B$ otherwise

```
1: function REDUCE2(E_B, B, {a}, φ, R)
2:     C ← B \ {a}
3:     E_C ← ∅
4:     w_C ← GetWeight(E_B)
5:     for all EquivalenceClasses e ∈ E_B do
6:         DEC ← ∅
7:         v_B ← GetInstance(e)
8:         v_C ← Remove(v_B, {a})
9:         e_tmp ← Find(E_C, v_C)
10:        if e_tmp ≠ NULL then
11:            DEC ← GetDec(e_tmp) ∩ GetDec(e)
12:            if |DEC| > 0 then
13:                SetDec(e_tmp, DEC)
14:            else
15:                w_C ← w_C − |e|
16:                if w_C ≤ (1 − φ)|U| then
17:                    return E_B
18:                end if
19:                StoreExceptionRule(R, e_tmp)
20:            end if
21:        else
22:            AddEquivalenceClass(E_C, e)
23:        end if
24:    end for
25:    SetWeight(E_C, w_C)
26:    return E_C
27: end function
```

Table V
SUCCESSFUL REMOVAL OF ATTRIBUTE a_2

Id	a_1	a_3	a_4	$m_d^{\varepsilon}([x]_B)$	
$\{x_1, ..., x_4\}$	1	2	2	{0,1}	✓
$\{x_5, x_6\}$	3	1	2	{0,1}	✓
$\{x_7\}$	2	1	2	{1}	✓
$\{x_8 ... x_{12}\}$	1	2	1	{1,2}	✓
$\{x_{13} \cup x_{14}\}$	2	1	1	$\{1\} \cap \{0\} = \emptyset$	✓**(-1)**

Table VI
STATE AFTER REMOVAL OF ATTRIBUTE a_2 – OBJECT x_{14} WAS REMOVED

Id	a_1	a_3	a_4	$m_d^{\varepsilon}([x]_B)$
$\{x_1, ..., x_4\}$	1	2	2	{0,1}
$\{x_5, x_6\}$	3	1	2	{0,1}
$\{x_7\}$	2	1	2	{1}
$\{x_8 ... x_{12}\}$	1	2	1	{1,2}
$\{x_{13}\}$	2	1	1	{1}

found, we do not search any other rules related to this reduct, i.e., any shorter more general rule that had been created in the subsequent stages of the learning process.

VIII. BIREDUCTS AND GENERALIZED MAJORITY DECISION

The first formulation of decision bireducts occurred in [9], [6], where their Boolean characterization was described and a simple ordering algorithm aimed at their heuristic extraction from data was proposed by an analogy to classical decision reducts. The motivation to introduce decision bireducts was to allow for explicit analysis whether classifiers that use different selected subsets of attributes do not repeat classification mistakes on the same objects in the training data set. Experiments reported in [9] show that diversification of subsets of attributes with this respect may be important in practice.

Definition 11. *Let $\mathbb{A} = (U, A \cup \{d\})$ and subsets $B \subseteq A$, $\mathcal{X} \subseteq U$ be given. We say that B determines d within $\mathcal{X}$, further denoted as $B \Rrightarrow_{\mathcal{X}} d$, if and only if B discerns all pairs $u_i, u_j \in \mathcal{X}$ such that $d(u_i) \neq d(u_j)$. Further, we say that the pair $(\mathcal{X}, B)$ is a decision bireduct if and only if the following holds:*

1) *There is $B \Rrightarrow_{\mathcal{X}} d$,*
2) *There is no proper subset $B' \subsetneq B$ such that $B' \Rrightarrow_{\mathcal{X}} d$,*
3) *There is no proper superset $\mathcal{X}' \supsetneq \mathcal{X}$ such that $B \Rrightarrow_{\mathcal{X}'} d$.*

The original ordering algorithm to calculate a decision bireduct for a given data set is presented in Algorithm 5. It is recalled here as a reference point to a modified version of the algorithm (presented as Algorithm 6) that works with equivalence classes instead of single objects and with generalized majority decision intersections constraint instead of the original discernibility condition.

A major difference of the ordering algorithm for decision bireducts extraction, comparing to the ordering algorithms obtaining decision reducts (reduction on attributes only, e.g., Algorithm 2), can be observed in the main loop iterating on both the attributes and the objects at the same time. An appropriate action is taken, i.e., either removal of unnecessary attribute or addition of an object to the result, on base of a given condition for the generalized majority decisions intersections.

It is worth to notice that the reduct and bireduct versions of the generalized majority decision algorithms are not equivalent, i.e., there exist data sets for which certain results can be obtained only by one of the methods.

As an example, let us consider the following fragment of a decision table. All presented objects have the same values on feature subset B and the specified values on attributes a and

Algorithm 5 Original Ordering Decision Bireduct Algorithm

Input: a decision table $\mathbb{A} = (U, A \cup \{d\})$, σ – a permutation of a set $\{1, \ldots, |U| + |A|\}$
Output: a decision bireduct for $\mathbb{A}$

$\mathcal{X}_0 \leftarrow \emptyset, B_0 \leftarrow A$
for $i = 1 \rightarrow |U| + |A|$ **do**
  $B_i \leftarrow B_{i-1}$
  $\mathcal{X}_i \leftarrow \mathcal{X}_{i-1}$
  if $\sigma(i) \leq |U|$ **then**
    if $B_i \Rrightarrow_{\mathcal{X}_i \cup \{u_{\sigma(i)}\}} d$ **then**
      $\mathcal{X}_i \leftarrow \mathcal{X}_i \cup \{u_{\sigma(i)}\}$
    end if
  else
    if $B_i \setminus \{a_{\sigma(i)-|U|}\} \Rrightarrow_{\mathcal{X}_i} d$ **then**
      $B_i \leftarrow B_i \setminus \{a_{\sigma(i)-|U|}\}$
    end if
  end if
end for
return $(\mathcal{X}_{|U|+|A|}, B_{|U|+|A|})$

Algorithm 6 Ordering Generalized Majority Decision Bireduct Algorithm

Input: a decision table $\mathbb{A}^{m_d} = (U^{m_d}, A \cup \{m_d\})$ obtained from $\mathbb{A} = (U, A \cup \{d\})$, where equivalence classes induced by A are considered objects in $\mathbb{A}^{m_d}$, i.e., $U^{m_d} = U/A$ and where m_d is used to compute their decision values, σ – a permutation of a set $\{1, \ldots, |U^{m_d}| + |A|\}$
Output: a decision bireduct for $\mathbb{A}^{m_d}$

$\mathcal{X}_0 \leftarrow \emptyset, B_0 \leftarrow A$
for $i = 1 \rightarrow |U^{m_d}| + |A|$ **do**
  $B_i \leftarrow B_{i-1}$
  $\mathcal{X}_i \leftarrow \mathcal{X}_{i-1}$
  if $\sigma(i) \leq |U^{m_d}|$ **then**
    if $Test^{m_d}(\mathcal{X}_i \cup \{u^{m_d}_{\sigma(i)}\}, B_i)$ **then**
      $\mathcal{X}_i \leftarrow \mathcal{X}_i \cup \{u^{m_d}_{\sigma(i)}\}$
    end if
  else
    if $Test^{m_d}(\mathcal{X}_i, B_i \setminus \{a_{\sigma(i)-|U^{m_d}|}\})$ **then**
      $B_i \leftarrow B_i \setminus \{a_{\sigma(i)-|U^{m_d}|}\}$
    end if
  end if
end for
return $(\mathcal{X}_{|U^{m_d}|+|A|}, B_{|U^{m_d}|+|A|})$

Algorithm 7 Test function used by Algorithm 6.

```
1: function Test^{m_d}(X ⊆ U^{m_d}, B ⊆ A)
2:     X_Test ← ∅
3:     for all EquivalenceClasses e ∈ X do
4:         v ← Remove(GetInstance(e), A \ B)
5:         e_tmp ← Find(X_Test, v)
6:         if e_tmp ≠ NULL then
7:             DEC ← GetDec(e_tmp) ∩ GetDec(e)
8:             if |DEC| > 0 then
9:                 SetDec(e_tmp, DEC)
10:            else
11:                return false
12:            end if
13:        else
14:            AddEquivalenceClass(X_Test, e)
15:        end if
16:    end for
17:    return true
18: end function
```

b.

B	a	b	$m_d([x]_{B\cup\{a,b\}})$
$\cdots$	0	0	$\{0,1,2\}$
	0	1	$\{0,1,3\}$
	1	0	$\{0,2,3\}$
	1	1	$\{1,2,3\}$

Let as consider the generalized majority decision reduct computation flow. If we try to remove only one column a or b, respectively, we would get the following:

B	b	$m_d([x]_{B\cup\{b\}})$
$\cdots$	0	$\{0,2\}$
	1	$\{1,3\}$

B	a	$m_d([x]_{B\cup\{a\}})$
$\cdots$	0	$\{0,1\}$
	1	$\{2,3\}$

We can observe than in either case the second column cannot be removed because $\{0,2\}\cap\{1,3\}$ as well as $\{0,1\}\cap\{2,3\}$ are equal to $\emptyset$. Let us now consider a fragment of the input permutation $a,1,2,3,b,4$ to the generalized majority decision bireduct algorithm. We can remove the attribute a, add equivalence classes $1,2,3$ and remove the attribute b because the intersection of the appropriate decisions are equal to $\{0\}$:

B	$m_d([x]_B)$
$\cdots$	$\{0,1,2\}$
	$\{0,1,3\}$
	$\{0,2,3\}$
$\cdots$	$\{1,2,3\}$

The subsequent addition of the equivalence class 4 is not possible due to the fact that intersection of the decision $\{1,2,3\}$ with all other already added to the decision bireduct is equal to $\emptyset$. Therefore, we would finish with a decision bireduct consisting of the attribute subset B and the subset containing only the first three equivalence classes.

IX. Experiments

We conducted our experiments on a collection of benchmark data sets available from the University of California at Irvine (UCI) Repository [15]. We present our results based on two data sets (*DNA-splices* and *ZOO*) but similar results can be observed on other benchmark data. We compared different types of approximate decision reducts calculation: the approximate decision reducts (ADR) calculated according to Algorithm 1 and three variations on the generalized approximate majority decision reducts (GMDA) with exceptions rules (EXEP) and with exceptions pointing to unknown decision (GAPS) and without any exception rules (NONE) but with the same reduction criteria. We have also tested each algorithm based on M and R measures. In case of GMDA procedure, the indications M and R correspond to the way how generalized approximate majority decision set was constructed. In case of the R measure we considered the relative frequency of a decision class within each equivalence class prior to a decision frequency in the testing set.

The data sets that were not given with *a priori* split into training and testing data sets, have been tested using 5-fold cross validation. Every experiment was conducted at least 20 times. We followed the well known ensemble technique to construct a classifier consisting of 10 reducts, each yielding a single weak classifier with a set of decision rules (including exceptions) calculated according to a particular method. We calculated 200 decision reducts and selected reducts with the smallest number of features. In case of exception rules we also included exceptions in calculation of the average reduct size but we used weighted average, where number of recognized objects is understood as a weight of the reduct.

Like in any ensemble-based classifier, some combination technique of classifier outputs must be applied. We used voting based on decision rule confidence, but other rule measures can be also utilized – see our previous research [16] where we described and tested several methods for rule identification and voting mechanisms.

X. Conclusions

We have presented several new methods for calculating approximate decision reducts. All of presented methods are based on the generalized majority decision function and the concept of analyzing intersections between generalized majority decision sets. We consider new methods to be more flexible in expressing the user preferences to control the distribution of decision classes inside each equivalence class. The presented methods are also less computationally expensive in comparison to the standard methods for approximate decision reducts, as they allow, in most situations, a quicker test for attribute removal. We have also presented the idea of constructing a decision rule-based classifier that use a special type of rules referred as exceptions. These rules are produced during the decision model construction and allow constructing models with a reduced dimensionality but with information about original training data preserved. We have experimentally validated our propositions and algorithms. Results show that

Figure 1. Average accuracy on *DNA-splices* data set

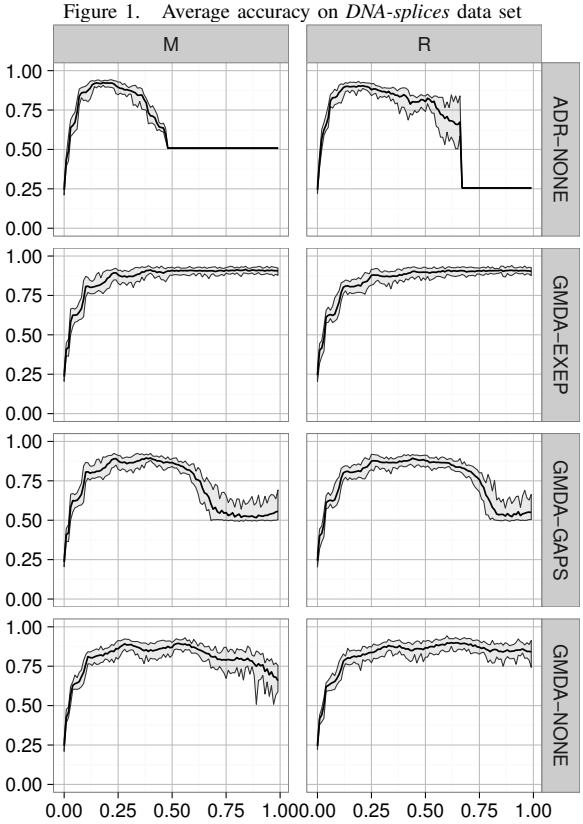


Figure 2. Average length of decision rules – *DNA-splices* data set

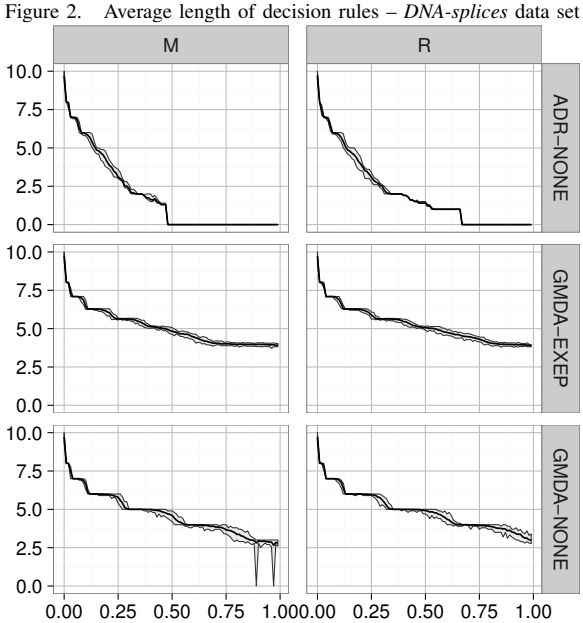


Figure 3. Average accuracy on *ZOO* data set

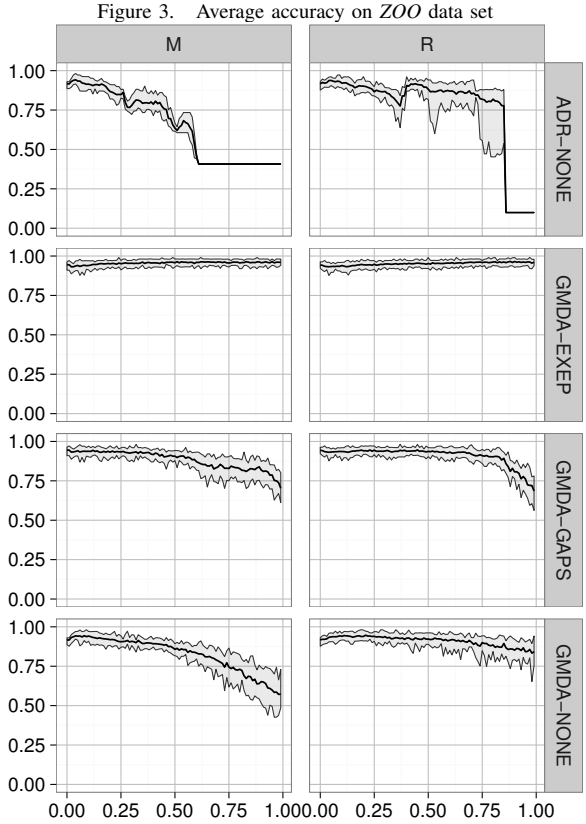
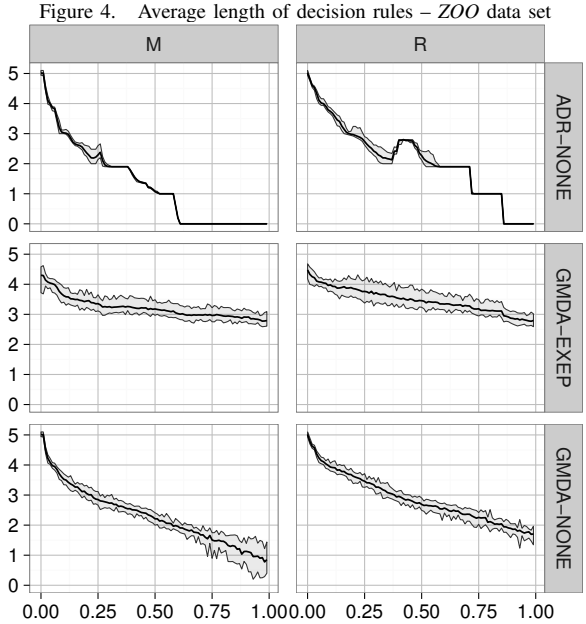


Figure 4. Average length of decision rules – *ZOO* data set



the new way of constructing approximation decision reducts have very comparable accuracy but is less computationally complex. Last but not least, we compared our method with bireduct framework and proposed new method for bireduct calculation based on the generalized majority decision set intersections. We plan to analyze the last method more deeply in the nearest future.

REFERENCES

[1] H. Liu and H. Motoda, Eds., *Computational Methods of Feature Selection*. Chapman & Hall/CRC, 2008.

[2] Z. Pawlak and A. Skowron, "Rudiments of rough sets," *Information sciences*, vol. 177, no. 1, pp. 3–27, 2007.

[3] D. Ślęzak, "Rough Sets and Functional Dependencies in Data: Foundations of Association Reducts," *LNCS Transactions on Computational Science*, vol. V, pp. 182–205, 2009.
[4] S. Widz and D. Ślęzak, "Approximation Degrees in Decision Reduct-based MRI Segmentation," in *FBIT*. IEEE, 2007, pp. 431–436.
[5] A. Janusz and S. Stawicki, "Applications of Approximate Reducts to the Feature Selection Problem," in *RSKT*. Springer, 2011, pp. 45–50.
[6] S. Stawicki and S. Widz, "Decision bireducts and approximate decision reducts: Comparison of two approaches to attribute subset ensemble construction," in *Proc. of FedCSIS'2012 Conf.* IEEE, 2012, pp. 331–338.
[7] D. Ślęzak, "Normalized Decision Functions and Measures for Inconsistent Decision Tables Analysis," *Fundamenta Informaticae*, vol. 44, no. 3, pp. 291–319, 2000.
[8] D. Ślęzak and W. Ziarko, "The Investigation of the Bayesian Rough Set Model," *International Journal of Approximate Reasoning*, vol. 40, no. 1-2, pp. 81–91, 2005.
[9] D. Ślęzak and A. Janusz, "Ensembles of Bireducts: Towards Robust Classification and Simple Representation," in *Proc. of FGIT 2011*, ser. LNCS, vol. 7105, 2011, pp. 64–77.
[10] S. Widz and D. Ślęzak, "Rough Set Based Decision Support – Models Easy to Interpret," in *Selected Methods and Applications of Rough Sets in Management and Engineering*, ser. Advanced Information and Knowledge Processing, G. Peters, P. Lingras, D. Ślęzak, and Y. Yao, Eds. Springer, 2012, pp. 95–112.
[11] M. J. Moshkov, M. Piliszczuk, and B. Zielosko, *Partial covers, reducts and decision rules in rough sets: theory and applications*. Springer, 2009, vol. 145.
[12] W. Ziarko, "Probabilistic approach to rough sets," *International Journal of Approximate Reasoning*, vol. 49, no. 2, pp. 272–284, 2008.
[13] G. Peters, P. Lingras, D. Ślęzak, and Y. Yao, *Rough sets: Selected methods and applications in management and engineering*. Springer Science & Business Media, 2012.
[14] J. Wroblewski, "Ensembles of classifiers based on approximate reducts," *Fundam. Inform.*, vol. 47, no. 3-4, pp. 351–360, 2001. [Online]. Available: http://content.iospress.com/articles/fundamenta-informaticae/fi47-3-4-14
[15] M. Lichman, "UCI machine learning repository," 2013. [Online]. Available: http://archive.ics.uci.edu/ml
[16] D. Ślęzak and S. Widz, "Is It Important Which Rough-Set-Based Classifier Extraction and Voting Criteria Are Applied Together?" in *Proc. of Int. Conf. on Rough Sets and Current Trends in Computing (RSCTC)*, ser. LNAI, vol. 6086. Springer, 2010, pp. 187–196.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 175–180
DOI: 10.15439/2016F462
ACSIS, Vol. 8. ISSN 2300-5963

Clustering Documents on Case Vectors Represented by Predicate-Argument Structures – Applied for Eliciting Technological Problems from Patents

Hitomi Yanaka, Yukio Ohsawa
Department of Systems Innovation, Faculty of Engineering,
University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, Japan
Email: h2.yanaka@gmail.com, ohsawa@sys.t.u-tokyo.ac.jp

***Abstract*—Patent analysis is useful to understand the trends of technological problems and develop strategies for technologies. Here patent classification is a method to support the analysis. The purpose of this study is to propose a method for patent classification, with the use of hierarchical clustering based on the structural similarity of problems to be solved. The structural similarity can be calculated with case vectors based on predicate-argument structures of the contents of the patents. The interview survey indicated that this classification plays an essential role in analogical problem solving, by allowing visualization of similar technological problems.**

I. Introduction

IN COMPANIES, it is important to understand the trends of technological problems by analyzing patents. By this, they learn to improve strategies for the development of technologies. Classification symbols have been used for classifying patent documents in patent analysis. Patent examiners use them to search similar patent documents written in different technical words or languages. The classification symbols are updated manually by experts. For improving the efficiency, a method to classify patents automatically based on the semantic content is required.

A patent map, a model of patent visualization, is also a method of patent classification. Here patent information is collected for a specific purpose of use and depicted in a visual form of presentation such as a chart, matrix, graph, or table. Fig. 1 shows an example of a patent matrix map [1]. As can be seen from Fig. 1, a patent map is mainly produced to grasp its technology trends by gathering related patent information of target technology fields. However, it cannot be identified at a glance which patent documents are contained in each cluster. Also, the target technology fields have been specified manually by experts.

Analogy is the process toward understanding and solving the problem from the relation between the knowledge base (sometimes called the source) and the target problem. Finding the relationship between the base and the target can be useful for patent analysis. According to the structure mapping theory [2], there are two kinds of similarities: superficial similarity and structural similarity. Superficial similarity is characterized by elements contained in the target concept and the base concept. Structural similarity is characterized by the primary or higher-order relationships between elements included in them. MAC / FAC model has been proposed as a model to search common elements between the base and the target based on the superficial similarity, and to assess the validity of the reasoning by evaluating the structural similarity [3].

In this study, to support the creation of technical development strategy, we proposed a method for clustering to find a relationship between patent documents and classifying them, based on the structural similarity of texts expressing technological problems to be solved by invention. In addition, we proposed a method of visualization of patent documents to grasp technological problems of patent documents in each cluster at a glance.

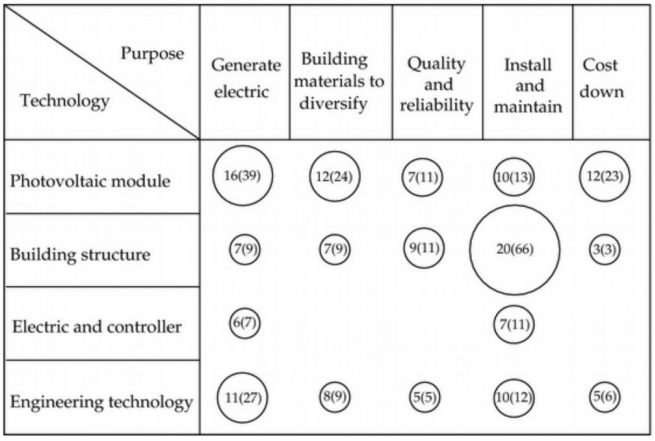


Fig. 1 example of a patent map: the size of each bubble correlates the number of registered patent publication (patent publication in parentheses.)[1]

This work was supported by CREST, Japan Science and Technology Agency. This paper has been accomplished under collaborative research project with Toppan Forms, Tokyo, Japan.

II. Related Work

A. Extraction of Technological Problems by Clustering

A previous study proposed a method of extracting a topic by automatic classification of newspaper articles [4]. This method composes a document vector by using the frequency of each noun in the document as features and extracts the topic by performing hierarchical clustering. However, when the similarity between two documents is calculated by the frequency of each noun in the documents, a document including different words with the same structure can be regarded to be dissimilar. Therefore, using the frequency of each noun yields superficial similarity, not structural similarity. However, there is a possibility that a new topic of technological problems can be found by gathering technological problems with the same structure. Therefore, we considered both the superficial similarity and the structural similarity and represented the feature amount of patent document. This enables us to make clusters based on MAC / FAC Model, a human analogy model.

B. Conversion to a First-order Predicate Logic Formation

For supporting analogy, it is necessary to consider both structural and superficial similarities, not only for construction of sentence vectors but also for visualization. The previous study reports the usefulness of converting sentence to predicate-argument structures for the recognition of implicational relation and the analysis of dialogues [5]. This study applies the conversion result from a conversion system called Boxer [6, 7] to applied tasks of semantic analysis using analogy. However, the conversion system to predicate-argument structures has not been used for supporting idea creation. Furthermore, the conversion system from Japanese sentence has not been confirmed yet. Japanese grammar is different from English grammar and it is necessary to propose a method to convert into logical form particular in Japanese language.

Therefore, we proposed a method to support analogy with the use of predicate-argument structures. We expressed sentences of technological problems by combining important predicate-argument structures. Comparing technological problems written by predicate-argument structures with each other, patent researchers can find the structural similarity between the problems. With the use of predicate-argument structures, it enables the researchers to easily use analogy and to understand how each patent document approaches technological problems.

III. METHOD

The proposed method consists of six steps. Below let us describe details of each step.

Step.1 Summarization of Technological Problem

We extract important sentences about technological problems from patent documents by summarizing the content filled in the blank "problems to be solved by the invention." Here we use a basic model for summarizing text by selecting the important sentences [8]. The important sentences are selected based on the scores until the length limit is reached. The length of the summary is fixed as around 100 words to recognize at a glance. The score of each sentence, $score(s)$, is calculated by the amount of information included in the sentence s as shown in Eq. (1).

$$score(s) = \sum_i log\, freq(w_i, s) pos(w_i, s) / \|s\| \quad (1)$$

In a Bag-of-Words model, the probability of a word w in a sentence s can be measured by its frequency as $freq(w,s)/\|s\|$, where $freq(w,s)$ indicates the frequency of w in s and $\|s\|$ indicates the total number of words in s. w_i indicates the i th word in s. $pos(w, s)$ is the weight of the position of w in s. We take a geometric sequence as a calculate function of $pos(w, s)$ based on the position hypothesis that earlier appearances of a word are more informative. A geometric sequence is defined as $f(w, s, i)$ in Eq.(2), based on the assumption that the degree of every appearance of a word is the sum of the degree of all the following appearances of it.

$$f(w, s, i) = f(w, s, i+1) + f(w, s, i+2) + \ldots + f(w, s, n) \quad (2)$$

The content of "problems to be solved by the invention" represents the technological problem, starting from what previous methods could not solve, followed by details, concluding with the purpose of the invention, i.e. the target technological problems. Important words of technological problems tend to appear at the beginning and repeatedly appear in the content. Therefore, the scoring method is adequate to summarize the content.

Step.2 Extraction of Predicate-argument Structures

Next, the summary of a technological problem is expressed as the combination of predicate-argument structures. We use dependency parsing of the summary and extract predicate-argument structures. As a simple format of the technological problem, predicate-argument structures are composed of nouns, case-marking particles, and verbs. MeCab[9] is used for morphological analysis and CaboCha[10] is used for dependency parsing. Both of them are appropriate for Japanese language.

In this research, the research object is confined to predicate-argument structures with adnominal case particles "-ga" (means subjective case), "-wo" (means accusative case), and "-ni" (means objective case). These three factors construct the framework of a sentence, especially if the sentences expressing an aimed function of a subject interacting with things to be given as objects. That is, nouns are regarded as indications of subjects and objects in the sentence. If consecutive nouns including prefix or suffix are contained in the sentence, they are regarded as a whole word. Numbers, pronouns, and syncategorematic words are excluded from the extent of research object.

Categorematic verbs are extracted as predicates. When the verbal auxiliary, "-nai" (means adding negative) followed behind the verb, we mention it behind the verb. The noun related to the nominal verb, "suru" (means "do") are extracted together as one verb.

From the result of dependency parsing, each verb written in original form is defined as a predicate, and relevant nouns are defined as values of its variables. The verbs, which appear in over 90% of the datasets, are meaningless verbs and are excluded from technical problems.

Step.3 Representation of Case Vector

To consider both the superficial similarity and the structural similarity, we also focus on a predicate-argument structure, which consists of nouns, verbs, and their relations. The predicate-argument structure represents case form, which is defined by a form of generative grammar such as subjective and objective, based on the semantic relationship of noun phrases to verbs. As one sentence contains at least one predicate-argument structure, the structure of the sentence can be represented by case vectors.

In Japanese language, a semantic relationship between a verb and a noun is confirmed by a case-marking particle. Furthermore, a verb tends to have a relationship with certain nouns to express its meaning. For instance, the verb "boil" tends to have a relationship with the nouns such as "human being" as the subject and has the relationship with the nouns such as "meals" as the object. Therefore, we represent a verb vector based on the verb's semantic role by using frequency of the nouns, which have a relationship with the verb such as subjects or objects.

Additionally, a case relationship represents the influence of a verb to/from a related noun. That is, the case relationship is useful for a semantic vector representation of a verb. We represent a verb vector by using the appearance of the nouns, which are in the same case relationship. The same case relationship means that they have the same case-marking particle. We define a case vector as a verb vector of each case-marking particle. For example, verb "Iku" (meaning "goes") in sentence "Prime Minister Abe ga Russia ni Iku." in Japanese, meaning "Prime Minister Abe goes to Russia," is put into:

Iku: (subject: Prime Minister Abe, Object: Russia).

This will be then put into a case vector where numerical values are filled as elements corresponding to "subject: Prime Minister Abe", "Object: Russia".

The case vector is constructed by a Bag-of-Words model. This model is the simplest vector representation for a sentence or a document. Each dimension of the case vector is correlated with a noun. Let us take a subjective case vector as an example. The value of the dimension is calculated as 1 if there is a subjective noun for a verb in the document and otherwise calculated as 0. Furthermore, the vector for the *i*-th verb is weighted by the *j*-th noun, which is related to the verb in the sentence. When the verb *i* has a predicate-argument structure with *m* nouns in the corpus, the weight of the *j*-th noun is calculated in Eq. (3).

$$w_i = (w_{i1},\ w_{i2}, \dots, 2\ w_{ij}, \dots,\ w_{im}) \tag{3}$$

In this study, to identify patterns in the relationships between nouns and concepts of verbs, a method of Latent Semantic Indexing (LSI) is applied to nouns/verbs matrix. LSI is an indexing and retrieval method to identify patterns in the relationships between the terms and concepts contained in an unstructured collection of text [11]. Additionally, the computational complexity decreases and the accuracy of a clustering algorithm can be improved with the use of LSI. This study is different from the previous study in applying LSI to nouns/verbs matrix.

If *n* is the number of verbs and *m* is the number of unique nouns in the corpus, *A* is the matrix about verbs *V* (*V*=1, …, *n*) which are related to nouns *N* (*N*= 1 , …, *m*) in the sentence *s* as shown in Eq. (4).

$$A = \begin{array}{c} \\ N_1 \\ N_2 \\ N_m \end{array} \begin{array}{c} \begin{array}{ccc} V_1 & V_2 & V_n \end{array} \\ \begin{pmatrix} 0 & 2 & 0 \\ 2 & 0 & 0 \\ 0 & 0 & 4 \end{pmatrix} \end{array} \tag{4}$$

This matrix *A* is decomposed by Singular Value Decomposition (SVD). When the rank of *A* is *r*, *T* is an *m* by *r* noun-concept vector matrix. *S* is an *r* by *r* singular values matrix. *D* is an *n* by *r* concept-verb vector matrix. Then, the matrix *A* is compressed to *k* dimensional space by restructuring the matrix *A'* using top *k* amounts of large and primary factor shown in Eq. (5). Dimension *k* is defined from preliminary experiment (In this study, k=200).

$$A' = T_k S_k {D_k}^T \tag{5}$$

Step.4 Calculate Distance Between Patents

Next, hierarchical clustering by using average linkage is selected as a method of cluster analysis of technological problems of patents. The relationship of each cluster is visualized by hierarchical clustering. As the predicate-argument structure is the framework of the document, the distance between the documents can be treated as a total of the distance between the predicate-argument structures in the document. Therefore, we calculate a distance between two patent documents as the sum of distances between the case vectors in the documents. By this calculation, the distance between two documents can be determined according to their predicate-argument structures. For simplicity, in this study, we defined three kinds of case vectors: subjective case vector, accusative case vector, and objective case vector. Nouns, which prefer to relate to verbs are different on the type of case and the distance between case vectors should be calculated in the same type of case each other. Therefore, we calculate a distance of the same type of case vectors and add them together as shown in Eq. (6).

$$sdist = sdist_{sub} + sdist_{acc} + sdist_{obj} \quad (6)$$

$sdist$ means the distance between two documents, $sdist_{sub}$ means the distance between these documents in the subjective case.

The distance $cdist\ (V_1, V_2)$ between two case vectors, V_1 and V_2 is calculated by Euclidean distance divided by the

$$cdist\ (V_1, V_2) = \frac{1}{cdist_{MAX}} \sqrt{\sum_{k=1}^{n} (V_{1k} - V_{2k})^2} \quad (7)$$

maximum of all distances, as shown in Eq. (7).

V_{1k} and V_{2k} are values of V_1 and V_2 at a dimension k. Each distance takes a value within the range of $0 \leqq cdist \leqq 1$ so that the distance of each type of case should be treated as equivalent.

When the document p contains k subjective nouns and document q contains l subjective nouns, the distance $sdist_{sub}\ (p, q)$ between these documents in the subjective case is defined as Eq. (8).

$$sdist_{sub}\ (p, q) = \frac{1}{kl} \sum_{i \in k} \sum_{j \in l} cdist(V_i, V_j) \quad (8)$$

As shown in Eq. (8), the sum of distances of the subjective cases is divided by the number of the cases to consider the average of the distances as the distance of the case.

If one document contains one direct objective noun and the other does not, $sdist_{acc}$ cannot be calculated in Eq. (8). In this case, these two documents are completely different in accusative case and $dist_{acc}$ is regarded as the maximum $sdist_{acc}$, 1. Similarly, if two documents contain no indirect objective nouns, $dist_{obj}$ cannot be calculated in Eq. (8). In this case, these two documents are the same in objective case and $dist_{obj}$ is regarded as the minimum $sdist_{obj}$, 0.

To illustrate the above, there are two example sentences of calculating distances.

Example 1: Subject A / Verb B / Direct Object C

Example 2: Subject N / Verb O / Indirect Object P, Subject Q / Verb R / Direct Object S

There are two pairs of the subjective case (A-N and A-Q) and one pair of the accusative case (C-S). Therefore, $sdist_{sub}(s_1, s_2)$ is expressed to be $\frac{1}{2}\ (cdist(V_A, V_N) + cdist(V_A, V_Q))$ and $sdist_{acc}(s_1, s_2)$ is expressed to be $cdist(V_C, V_S)$. In addition, these two documents are completely different in objective case and $sdist_{obj}(s_1, s_2)$ is 1. In this example, the formula to calculate the distance $sdist(s_1, s_2)$ between Example 1 (s_1) and Example 2 (s_2) is shown in Eq. (9).

$$sdist(s_1, s_2) = \frac{1}{2}\ (cdist(V_A, V_N) + cdist(V_A, V_Q)) + cdist(V_C, V_S) + 1 \quad (9)$$

Step.5 Hierarchical Clustering of Patents

The hierarchical clustering method in the previous study[4] has a problem of fluctuating extracted topics because these topics subjected to the number of clusters and the number of clusters is decided in the hand. We here employed Upper Tail method [12, 13], clustering with an automatic decision of the number of clusters. This method is based on the stopping rule of Eq. (10).

$$\alpha_{j+1} > \alpha_{ave} + ks_\alpha \quad (10)$$

When the number of documents is n, the most appropriate number of clusters is j, and the number of clusters is n-j (j=0, 1,..., n-1), α_j is the minimum distance between two documents belonging to different clusters. α_{ave} is an average of a distribution of α_j. s_α is a square root of the unbiased estimate of variance. k is determined by the number of elements per cluster. We determine k=2, considering that the number of documents per cluster is estimated from 10 to 50. The previous study [13] shows normalizing α to be distributed by chi-square as shown in Eq. (11) achieved better accuracy and we perform the same procedures.

$$\alpha' = \Phi^{-1}(F_p(\alpha / s_\alpha \cdot p) \quad (11)$$

Φ^{-1} is the inverse function of the normal distribution. F_p is distribution function of chi-square when a degree of freedom is p (In this study, p=1).

Step.6 Visualization of Similar Technological Problems

In the end, we propose a method of visualization of technological problems. The format of output data is JSON, which is a lightweight data-interchange format. The embedded structure of the data is composed of the number of the cluster, technological problem solved by each patent document, and the patent number in sequence. In visualization, we use D3.js, which is a JavaScript library for visualizing data with JSON and HTML.

IV. Results and Discussion

A. Datasets

In this study, we used unexamined Japanese patent applications, which were published from 2013 to 2015 and contain the word "condiment" in the content of "problems to be solved by the invention" as datasets to make clusters and visualize. The number of documents was 348. The dictionary data for constructing the case vector was the content of "problems to be solved by the invention" of seventy thousand patent documents relating to "Foods" (theme code: 4H).

Fig. 2 Part of clusterization of patent documents in three conditions

First condition used the distance determined by the case vector, second condition used the distance by the sentence vector with the verbs as feature amount and third condition used the distance by the sentence vector with the nouns as feature amount.

B. Evaluation of Summarization of Technological Problem

324 summaries were created from the datasets in this study. The reason why some summaries could not be created is that the contents of "problems to be solved" of such patent documents were too short to summarize. In order to evaluate the validity of the summarization, randomly selected 100 summaries extracted by the method were compared to summaries determined by discussion among patent experts for the same documents. The rate of concordance in this comparison was 81% and it shows the method is adequate to summarize patent documents.

C. Evaluation of Cluster Analysis and Visualization

To evaluate the utility of the proposed method, we performed cluster analysis of patent documents, using three kinds of distances and compared the results with each other. The first distance was determined by the proposed method, the case vector. The second distance was determined by the sentence vector using frequency of verbs as the feature. The third distance was determined by the sentence vector using frequency of nouns as the feature. We used TF/IDF to weight the features in the second and the third distances. TF means the frequency of words in each sentence and IDF means logarithm of inverse number of the frequency of words used in total sentences of the summary of the technological problem. Fig. 2 shows part of the result of the cluster analysis in three conditions, corresponding to the three kinds of distances above. The number of clusters under each condition was determined as 44 (the first condition), 20 (the second condition), and 40 (the third condition).

We interviewed two patent experts engaged in the Intellectual Property Department of the food company for more than three years. We showed them the result of the cluster analysis in three conditions and asked two questions below.

Question 1. What kind of difference do you find by comparing the three types of clustering results? What similarities do you find in the technical problems in the same cluster?

Question 2. Can technological problems of patent document be grasped at the moment? Is there any use to develop technology strategy or intellectual property strategy?

The answers to Question 1 are shown below:

- At the first glance, each cluster in the first condition seems to lack in coherence. However, for example, the cluster No.9 seems to be aggregated by the same concept of verbs. The verbs such as "maintain", "take in" and "fill up" represent the concept of keeping and the verbs such as "sprinkle", "attach" and "add" represent the concept of adding.
- Each cluster at the second condition seems to be aggregated by the same verb roughly.
- As the cluster No.18 is aggregated by the same noun "extract" in the third condition, the same goes for others at the third condition.

These answers indicate that case vectors have a possibility to represent the concept of the verbs, which forms the

structural similarity of the sentences. On the other hand, as a result of the second and third condition, the sentence vector using frequency of words as the feature represents the superficial similarity based on the words. The third condition also indicates that the meaning of the sentence is more influenced by the meaning of nouns than that of verbs.

The answers to Question 2 are shown below:

- When technological problems are expressed by the predicate-argument structures, modifiers are removed. Therefore, removing terminology of confidential information with this method is useful for sharing confidential data with others. On the other hand, some technological problems are hard to read. For example, “advantage convenience+-ga+loses” is grammatically incorrect. The correct sentence is “advantage convenience is lost”.
- The words which indicate not only the material of condiment but also the method for processing were extracted as technological problems.
- Some technological problems were extracted with accuracy, while others with unnecessary words.

The first answer indicates that the proposed method of visualization is useful in not only patent documents but also confidential documents. The answer also indicates that sentences of technological problems including passive verbs are a little difficult to read. This is because an auxiliary verb that represents the passive voice was not considered in this study and excluded. The method of expressing a sentence with a passive verb as predicate-argument structures will be an issue to be addressed in the future.

The second answer indicates that technological problems extracted by the proposed method are useful to find a point of view in patent analysis, whereas existing patent classification is mainly divided by the material of a condiment.

The third answer suggests that the interviewee could find the structural similarity of some technological problems and the usefulness of this method. However, as the structural similarity is not based on the commonality of words, the interviewee had difficulty to find the structural similarity of others. Therefore, it is necessary to get an evaluation of this method to more interviewees. Another reason for this result is the structural similarity of two documents is not always directly represented by predicate-argument structures including in the documents. Therefore, it is necessary to consider the extraction method of the abstract predicate-argument structure in the documents.

V. Conclusion

In this study, we proposed a novel method of text clustering by using the distance of case vectors derived from predicate-argument structures of patent documents. A method for calculating the distance between documents based on predicate-argument structures has not been approached. This study suggested the possibility to capture not only superficial similarity based on nouns in sentences but also structural similarity based on the meaning of verbs in sentences. Furthermore, the interview survey indicated that the proposed visualization method of technical problems is useful to overview the problem, which patent researcher should search for. Therefore, this visualization method is thought to be effective to support to search patent documents relating to researcher's own technological problems.

In this research, we calculated the distance between patents as an average distance between case vectors included in the documents. However, as this method cannot consider the causal structures of documents, the similarity of documents could have been inaccurate. Therefore, an issue in the future is to evaluate the method for deriving similarity considering causal structures in documents.

References

[1] Chiu, Y.J., Ying, T., A Novel Method for Technology Forecasting and Developing R&D Strategy of Building Integrated Photovoltaic Technology Industry, Mathematical Problems in Engineering, 2012, 2012, pp.1-24, http://dx.doi.org/10.1155/2012/273530.

[2] Falkenheiner, B., Forbus, K., Gentner, D., The Structure Mapping Engine: Algorithm and Examples, Artificial Intelligence, 41, 1989, pp.1-63, http://dx.doi.org/10.1016/0004-3702(89)90077-5.

[3] Forbus, K., Gentner, D., and Law, K., MAC/FAC: A model of similarity-based retrieval, Cognitive Science, 19, 1994, pp.141- 205, http://dx.doi.org/10.1016/0364-0213(95)90016-0.

[4] Hashimoto, T., Murakami, K., Inui, K., Uchiumi, K., Ishikawa, M., Topic Extraction and Social Problem Detection Based on Document Clustering, SocioTechnica, Vol.5, 2008, pp.216-226, http://dx.doi.org/10.3392/sociotechnica.5.216

[5] Bos, J., Markert, K.. Recognising Textual Entailment with Logical Inference, Proceedings of the 2005 Conference on Human Language Technology and Empirical Methods in Natural Language Processing, Vol. 2012-NL-206, 2005, pp. 628–635, http://dx.doi.org/10.3115/1220575.1220654.

[6] Bos, J., Wide-Coverage Semantic Analysis with Boxer, Proceedings of the 2008 Conference on Semantics in Text Processing, 2008, pp. 277–286, http://dx.doi.org/10.3115/1626481.1626503.

[7] Bos, J., Clark, S., Steedman, M., Curran, J., Hockenmaier, J.. Wide-Coverage Semantic Representations from a CCG Parser. Proceedings of the 20th international conference on Computational Linguistics, 2004, pp.1240-1246, http://dx.doi.org/10.3115/1220355.1220535.

[8] Ouyang, Y., Li, W., Lu, Q., Zhang, R., A Study on Position Information in Document Summarization, Proceedings of the 23rd International Conference on Conputational Linguistics, Beijing, China, 23-27 August.2010, pp.919-927.

[9] MeCab: Yet another part-of-speech and morphological analyzer. http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html Accessed: 2016-05-06.

[10] Cabocha: Yet another Japanese dependency structure analyzer. http://taku910.github.io/cabocha/ Accessed: 2016-05-06.

[11] Dumais, S., Latent Semantic Analysis, Annual Review of Information Science and Technology, Vol.38, Issue.1, 2004, pp.188-230, http://dx.doi.org/10.1002/aris.1440380105.

[12] Mojena, R., Hierarchical grouping methods and stopping rules: an evaluation, The Computer Journal, Vol.20, 1977, pp.359-363, http://dx.doi.org/10.1093/comjnl/20.4.359.

[13] Shizu, A., Matsuda, S., Comparison of the Cluster Number Automatic Determination Method in a Cluster Analysis, Academia. Information sciences and engineering : journal of the Nanzan Academic Society, Vol.11, 2011, pp.17-34.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 181–186
DOI: 10.15439/2016F195
ACSIS, Vol. 8. ISSN 2300-5963

Evaluating Model of Traffic Accident Rate on Urban Data

Jianshi Wang
Department of System Innovation
Graduate School of Engineering
University of Tokyo
Email: jianshiwang0329@gmail.com

Yukio Ohsawa
Department of System Innovation
Graduate School of Engineering
University of Tokyo
Email:ohsawa@sys.t.u-tokyo.ac.jp

***Abstract*—Public safety, especially the daily traffic accident is concerned by the public. Previous studies have already discussed accident reasons associated with accidents statistically. There is a method called Innovators Marketplace on Data Jackets created by Professor Ohsawa. This method is used to externalize the value of data via stakeholders' requirement communication. This paper applied the solution from an IMDJ workshop to research this topic creatively. This novel solution suggested to do analysis on the combination of urban data and traffic accident rate to find the impact factors to the traffic accident rate in the urban system. This paper used factor analysis, structure equation modeling and data mining to construct a theoretical frame for traffic accident rate analysis for urban data. Different accident indexes, such as total number of accident, fatality rate, injury rate, and casualty rate are combined to construct a traffic accident risk evaluation model. This paper chosen the urban data as the solution from IMDJ workshop, such as population structure information, vehicle information, road characters, public traffic system information, and the other kinds of data to explore factor meaning, and to identify relationships between different factors. It segmented these urban data based on their categories, and determined accident risk for each section. By doing analysis on not only the original data but also the changing rate of these data each year, the result analytical results showed that traffic accident rate on urban data could be described by the combination of population structure, road characters, public traffic system and public facilities. These four sections affects traffic accident rate significantly during the development of urban; however, the vehicle factor does not have influence on traffic accident rate. And it proves the solution from IMDJ workshop is not only novel but also practical strongly. Making some solution from IMDJ into reality, we will find another new way to affect the world.**

I. INTRODUCTION

In Beijing, the fatal accident rate was 39.56 per million populations in 2014, it decreased roughly 2.8% in comparison with 2013, even though the population of 2014 has increased 1.7%. Different losses caused by traffic accident in urban are considerable. Many previous researches on traffic safety tried to find methods for preventing traffic accidents. Among these researches, the analytical methods are used frequently, including Clustering Analysis (CA), Statement Statistic (SS), Regression Analysis (RA), evaluation of the correlation between independent variable sets and a dependent variable; but these models could not cover the causal relationship between these variables or combine different traffic accident indexes to identify accident risk. Even though all of the final purpose of them attempts to avoid traffic accident, they mostly used discriminate analysis or regression model in analyzing accident characteristics and identifying behaviors of drivers, weather condition or road sections prone to accidents, then constructed an accident prediction model by parametric approaches (i.e. on distribution model). However, these methods only analyzed accidents on drivers with different attributes and relevance, weather data or different road sections; they cannot define the causality between factors clearly enough. Some current studies have already utilized data mining, including non-parametric approaches where traditional models were not employed, to identify key accident related impact factors. However, data mining can just identify traffic accident risk indexes (i.e. number of traffic accident, of death, of injuries, or of casualties). But it is necessary to identify the latent correlations among the different factors which is suggested by solution from IMDJ, and find the correlations among the traffic accident rate and urban data-population structure, vehicles, road characters or some other cause-and-effect relationships. In this study, we focused on urban data, applied structural equation modeling to find impact factors associated with traffic accident. Important traffic accident factors were identified, population structure, vehicle information, road characters and other cause-and-effect relationship associated with traffic accident risk were discussed, and prediction model which describe the latent relationship between urban city and accident was constructed. This paper discussed six factor sets by constructing an accident risk causal framework based on urban data and the component factor sets of each feature and influence on traffic accident.

II. LITERATURE REVIEW

This paper builds a theoretical framework for traffic accident risk by the analysis on the latent relationship between different aspects of urban data and traffic accident to identify causal relationships. In the past, Etienn(2006) used logit regression analysis to discuss reasons that led to different degree of injury condition in the traffic accident. Gender, age, and protective gear were related to the condition of injury. By using bivariate and simple analysis, Wong and Chung(2007) applied a rough set approach for accident chains to analyzing the influence of traffic accident for different factors. They found that a single factor accident chain had poor quality and

978-83-60810-90-3/$25.00

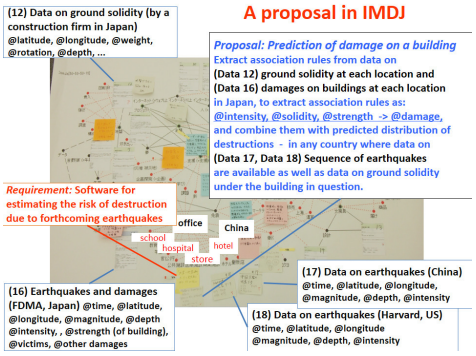


Fig. 1. A proposal in IMDJ

a multi-factor accident chain should be used when analyzing accidents. John(2007) has used mixed logit model to construct a prediction model for serious traffic accident. Variables related to daily traffic flow, such as average daily traffic flow, percentage of big vehicles, and number of access roads, and variables related to road features, such as curvature and road friction were significant variables impacting road accident risk. Niloofar(2015) utilized highly disaggregate spatial units for an analysis of area specific variables associated with urban traffic accident. These researches suggested that factors affecting traffic accident risk include drivers, road structure of city, public traffic system, which may affect each other. On the other hand, this paper applies the solution from IMDJ workshop to do factor analysis and structure equation modeling to analyze these features as well as the cause-and-effect relationships between urban factors and traffic accident rate. The main goal of this paper is necessary in order to have a clear enough structure of urban and traffic accident for improving safety which will be done with urban planning in the future.

III. MOTIVATION OF RESEARCH

The reason of doing this research comes from one solutions of IMDJ workshop. The IMDJ(Innovators Marketplace on Data Jackets) has been metioned about many times aboved. And this research is the application of a solution from IMDJ workshop, so I think it is necessary to introduce this amazing method. The purpose of IMDJ is making a social environment where analysits and decision makers in active businesses and science could be provided with data they need. During the workshop, each members could give requirements and provide solutions by checking or combining DataJackets on the Keygraph. And finally members can buy/sell their solutions in a reasonable condition. The Fig.1 shows a proposal in IMDJ workshop. DataJackets are small pieces of information containing the abstracts of data that exist but cannot be disclosed. And Keygraph created by Professor Ohsawa is the high-effective visualization method to show the relationship between different Datajackets by analyzing contents in Data jackets. Before an IMDJ workshop, the topic of the workshop should be determined by the organizer. During an IMDJ workshop which topic is ”Improving the quality of urban life”, a requirement about public safety has been asked. That requirement asked for a method to check the safety condition of objective people in an area. And another member supplied a solution of checking the safety index of the area by using combination of two Datajackets. One is the structure data of the city, and the other is the accident rate of the city. The research angle of this solution is totally different from the previous researches We selected the urban data just like the solution suggested to do this research. The society has paid lots of cost on traffic accidents significantly. And it is time to do the research to reveal the complex relationship between urban data and public saftery. The solution from IMDJ workshop gives a very novel and meaningful angle to do research on this problem.

IV. RESEARCH EXTENT AND METHODOLOGY

This study focused on urban data in Beijing City. Beijing is still in construction after Beijing Olympics 2008, and it is becoming bigger and advanced every year. Public facility and mass rapid public transit lines have been under construction in Beijing City. And also with some policy changed, the other social structures have been keeping changing every year, such as the different structure of population, the GDP of Beijing, the sales of different products and etc. These different structures have affected all of aspects in Beijing City. In respect of methodology, this paper used structure equation modeling with population, driver, vehicle, road characters, public transportation system, other urban data as independent variables, and accident rate as the dependent variable. All these data related to this urban were compiled into six categories. In addition to applying structure equation modeling, this paper used structure equation modeling method to explain the relationship between factors. To choose whether observation variables in each aspect are appropriate before model construction, each variable was subjected to factor analysis. It is necessary to exclude the unrelated factors in order to reduce the interference terms. After determining the variables, this paper analyzed the quantized the effect between these variable.

A. Data content

Urban data: We covered 30 factors, and clustered these factors into 6 sections according their categories. These 30 factors involving population, population in different age, number of truck, number of bus, gasoline consumption each year, number of hospital, number of police stations, road area, length of bus lane, green area in urban and some other aspects of urban. We assume, these data actually have the relationship among themselves and affect the traffic situation, it also includes the traffic accident naturally. Accident data: Accident rate was expressed by the sum of Fatality rate, Injury rate and Casualty rate multiplied by different weight.

$$Fatality\ rate = number\ of\ deaths/number\ of\ accident \quad (1)$$

$$Injury\ rate = number\ of\ injuries/number\ of\ accident \quad (2)$$

$$Casualty\ rate = number\ of\ casualties/number\ of\ accident \quad (3)$$

$$Accident\ rate = (1) * 3 + (2) * 2 + (3) * 1 \qquad (4)$$

Here, the above accident rate means the personal safety loss of the traffic accident.Theses different weights mean the degree of damage after accident based on the influence of traffic accident. The Fatal consequence is the most serious one, so I give the weight 3 to it; to the injury consequence, I give the weight 2 to evaluate it; to the casualty, because it prefers to refer to a range of affected people, I just use 1 to depict its weight. And the number of accident refers to the number of traffic accidents in that year. These traffic fatality, injury and casualty data were collected for the period 2010 to 2014 from the following sources: publications in the same period such as Beijing Statistics yearbook, official statistic site likes National Bureau of Statistics of China and official city traffic and accident reports.

B. Variables

We chose the data by which can be used to describe the urban objective situation in order to find the relationship between structure of urban, involving the structure of urban planning and social structure in urban, and accident rate. Then identified causality with the accident rate on the result. With the development of the urban and technology, the structure of urban is keeping changing every year. Doing analysis on the cause of the traffic accident to find the main reason by using statistical method or clustering method is common method. But we could not ignore the relationship between urban data and traffic accident rate. The changes in urban can affect every small aspects in our life, absolutely involving the accident. For example: alcohol drinking sales in urban may reflect the change of drunk driving rate in urban, and drunk driving is one of main reason of traffic accident, this chain shows the alcohol drinks sales has a tacit relationship with accident. And there must be other causes of the change in traffic accident from this kind of chain based on the urban system.

C. Model Framework

To build a traffic accident risk model on urban construction, different urban data were compiled. In this paper, we collected and compiled different data in an equal structure. This structure means that these urban data are on the same important position at the beginning of analysis. In this structure, the relationship does not only exist between the traffic accident rate and different urban factors, but also among these factors. It is more like a net of relationship. The model of analysis should be much more comprehensive, only in this way the result has persuasiveness, and intuitive. Fig.2 shows the framework in this research.

D. Analysis process

1) step1: Preprocessing the urban data: Using the raw data to do analysis is easily to ignore the influence of small change automatically during the analysis. It is necessary to preprocess the urban data. The purpose is to enlarge the influence of these factors and increase the efficiency of analysis.Also the final visualization will be clear enough to observe. Transforming the raw data into different ratio value, such as gender ratio, potential support ratio, vehicle type ratio, road light per meter, length ratio between road and total public traffic system, the ratio of green area, living space, and etc. Fig.3 shows the preprocess of urban data.

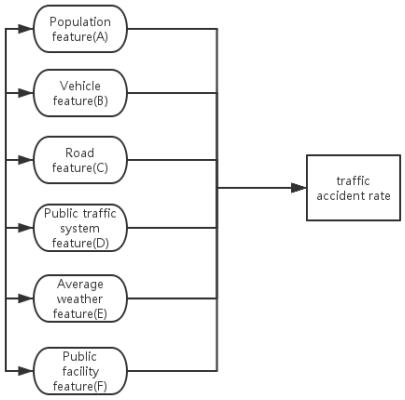


Fig. 2. model framework

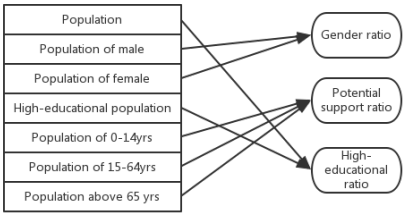


Fig. 3. Example of preprocessing

2) step2: Clustering latent factor sets: In order to distinguish the effects of different aspects in urban, before the analysis, it is necessary to define the preprocessed data into different sets. After the main components were analyzed, variables have been clustered in six features-population, vehicle,road characters, public traffic system, average weather data, public facility data. In this way, the result can give the structure and relationship between accident and variables, also among variables themselves. The TABLE I shows the result of clustering.

$$Population\ feature \longrightarrow A\ factor\ sets(A)$$
$$Vehicle\ feature \longrightarrow B\ factor\ sets(B)$$
$$Road\ feature \longrightarrow C\ factor\ sets(C)$$
$$Public\ traffic\ system\ feature \longrightarrow D\ actor\ sets(D)$$
$$Weather\ average\ feature \longrightarrow E\ factor\ sets(E)$$
$$Public\ facility\ feature \longrightarrow F\ factor\ sets(F)$$
$$\alpha.Urban\ data\{A, B, C, D, E, F\} \Rightarrow Accident\ Rate$$
$$\beta.\Delta Urban\ data\{A, B, C, D, E, F\} \Rightarrow \Delta Accident\ Rate$$

Population feature(A)	Vehicle Feature(B)
A1.All population A2.Gender ratio A3.High educational ratio A4.Low educational ratio A5.Potential support ratio A6.Driver ratio	B1.Number of vehicle B2.Type ratio(Big/small) B3.Driver/number of vehicle B4.Average gasoline consumption
Road characters(C)	Public traffic system(D)
C1.Length of road C2.Area of road C3.Road light ratio C4.Number of bridge C5.Wide road/narrow road	D1.Length of bus line D2.Length of subway line D3.Subway station/urban area D4.Total passengers per year D5.Bus passenger/subway passenger
Average weather feature(E)	Public facility(F)
E1.Good weather/Bad weather E2.Strong wind day E3.Days of low visibility	F1.Area of parks/areas F2.Green area F3.Hospitals F4.Schools F5.Police station

TABLE I

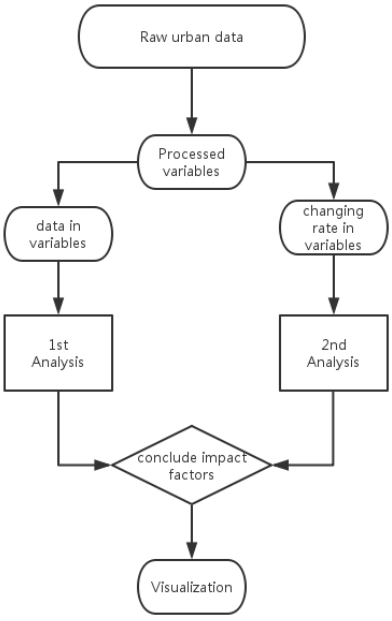


Fig. 4. Impact factors from result of analysis

3) step3: Calculating the changing rate. The whole process of analysis has been designed to do two times in order to confirm the effect of impact factors. The first used data in TABLE I directly, and the second used changing rate of factors in each year. The reason of second analysis on changing rate is to confirm the relationship between factors and the accident rate. I selected the factors which have strong relationship with accident rate by checking the P-value of analysis results. And two results can show overlap part of significant impact factors clearly, also it can confirm the influence of impact factors and quantize the influence of these impact factors.

4) step4: After getting results of these two analysis, factors which have no relationships with traffic accident are deleted. We used data of these impact factors to calculate the influence by numeric on urban traffic accident. Making a graph to visualize the relationship and degree of influence between these impact factors and traffic accident rate is necessary. With this graph, impact factors and their influence between each other can be observed clearly. The numbers marked on lines mean the degree of correlation between them. The positive number means the positive correlation, and the negative number means the negative correlation. The dotted line means the relationship is bidirectional relationship, and the solid line means the direction of relationship between sections is just one way. The Fig.4 aboved shows the whole process of analysis.

V. RESULT

The TABLE II showed results from analysis. Two columns showed impact factors and their quantized influence calculated by the assumed model.

After the first analysis, A5 factor(potential support ratio), D2 Factor(length of subway line), D3 Factor(subway station/urban area), F3 Factor(number of school) and F4 Factor(number of school) has been selected, they are from three different sections. The first result showed the potential support ratio, length of subway line, subway station/area, and number of school are the impact factors, among them the length of subway line showed the strongest relationship (0.938) with traffic accident rate. It gives some reasonable assumptions: the Beijing attracts more working population (due to the A5(potential support ratio)), and with the construction of subway line system, it is much more convenient, more and more people choose the public traffic system, and the traffic accident rate has been controlled. And with the number of hospital, the number of fatality in traffic accident has been affected. The number of school means, firstly the increasing working population takes more students, secondly the number of school has the relationship with public traffic system. The second analysis has used the changing rate in the same factors. The first result could be used to combine with it, then the combination of results could solve the problem comprehensively. After the second analysis, A2 factor(gender ratio), A3 Factor(high educational ratio), C1 Factor(length of road), D2 Factor(length of subway line) and F4 Factor(num of school) has been selected, from four different sections. The result of second analysis showed that gender ratio, high educational ratio, length of road, length of subway line and number of school are the impact factors. It showed the number of school has the strongest relationship with traffic accident rate.And it is apparently to find that the influence of F4 factor(number of school) has been improved. With considering the factor-High educational ratio, it showed some reasonable possibility: with the increasing of educational level, more people come to recognize the importance of traffic safety.And they know how to protect their safety,also it means in the traffic accident, the student accident has affected it strongly. With choosing the impact factors from these two analysis, we can visualize

First analysis	Second analysis
A5.Potential support ratio(0.55)	A2.Gender ratio(0.056)
D2.Length of subway line(0.938)	A3.High educational ratio(0.129)
D3.Subway station/city area(0.036)	C1.Length of road(0.028)
F3.Hospital(0.558)	D2.Length of subway line(0.004)
F4.Number of school(0.07)	F4.Number of school(0.867)

TABLE II

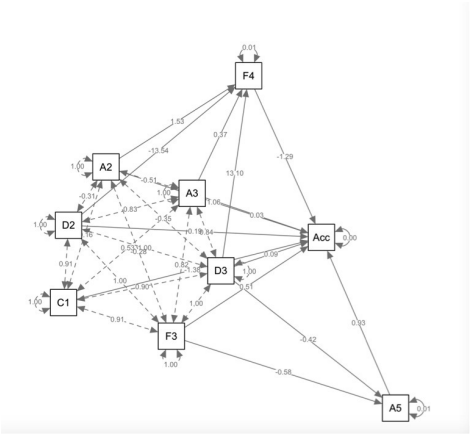


Fig. 5. Visualization of Impact Factors

the structure and the influence about them. Fig.5 showed the visualization of the result. The selected nine factors from these two analysis made this net diagram. As the interpretation of the diagram before, every impact factors showed the degree of influence, direction of the relationship and the combination of the relationship. For example, F3 (number of hospital) has affected the A5 (potential support ratio) by negative effect of 0.58, and the direction is just from F3 to A5. C1(Length of road) has the relationship with F3, they affected each other by the positive effect of 0.91. Generally, the traffic accident rate is affect by all of impact factors on this diagram. The A5(potential support ratio) has positive effect of 0.93 on traffic accident rate. The F4 (number of school) has negative effect of 1.29 on the traffic accident rate. The D2 (Length of subway line) has positive effect of 0.19 on the traffic accident. By gathering these coefficents, using these different factors as variables, we can define a model to describe the relationship between each factors and accident rate and the relationship among these different factors.

VI. DISCUSSION

"The more vehicles, the more accident.", it is a general consideration. But in this research, it showed there is no relationship with number of vehicles. Neither in the first analysis nor the second analysis, there is no result showing the accident rate has the relationship with vehicle factor. From the raw data, it showed even though the number of vehicle is increasing, the traffic accident rate has decreased, and by the result, it showed the vehicle data has no influence on the traffic accident data. On the other hand there is a strong relationship among the other urban data,such as-population structure, public traffic accident and traffic accident rate. Even though the traffic accident mainly about the vehicle caused accident on the road, the results here showed the other factors affected this rate obviously from the angle of urban structure data. And the other possibility exists, the changing data of vehicle variables has effects on the traffic accident rate, but the influence is too small, and the influence has been neutralized by other changed urban data. It still needs more and more data to make the model much more comprehensively, and make the correlation correctly.

VII. CONCLUSION

This paper constructed an assessment model for traffic accident risk which is novel in that we defined the relationship between urban data and traffic accident data. The four main sections-population structure, road characters, public traffic system and public facility had the significant impact on traffic accident rate, while the vehicle section was insignificant. A higher gender ratio (men/women) can reduce traffic accident risk, while a higher potential support increased the traffic accident risk. The longer length of road and public traffic lane can decrease the traffic accident rate.The results show that public traffic system and facilities are important factors to controlling the traffic accident rate. Traffic accident is mostly caused by people themselves, researchers can do many analysis on policy or reason analysis in order to reduce the traffic accident rate. However, from the new angle of IMDJ workshop, one should pay more attention to the significance from urban structure designing. This paper successfully made one traffic accident risk assessment model.Four sections-population structure, road characters, public traffic system and public facility are adequate for assessing traffic accident risk.It may require the deeper research in future, but this study was still restricted by its access to data. Although 30 variables were included,it was not enough. If the experiment add other kind of urban data, the model would be modified on causality and deduction. This paper was focus on the influence of urban data to traffic accident risk, it is suggestive to use this methodology to apply to other kind of public safe problem.

VIII. FUTURE WORK

Based on the conclusion of this paper, it showed that the relationship between urban data and accident could be quantized in number and described in a mathematical model. If there are other urban data, the model could be more perfect. So finding other urban data to construct the model more comprehensively. Trying to find what extent does this model match real sitiuation, and make it more comprehensive.
The requirement from that IMDJ workshop needs to check the safety condition of the objective people. So the model of traffic accident rate is not enough, because the public safety also covers the crime rate. From this solution of IMDJ, we prove the proposal which shows the traffic accident rate is related to the urban social data, so the crime rate of objective area could also be defined in that kind of mathematical model. And with combining these models, a synthesized prediction

model which can describe danger index of area based on urban data could be made. It could be used in checking the danger status of objective area. Also it could give reference to urban construction in the future. Because this interesting research angle was from IMDJ workshop, and this useful result was from the applicaiton of solution in IMDJ. I think the more results from IMDJ would make more exciting and useful things. The IMDJ workshop let members from different fields to externalize use value of data by combining their familiar data and other datas noveltly. It could help us to dig the deep value of the hidden datasets. Also it could help us to share the experience from professional people in different fields. I could not wait to hold a workshop with topic of public safety to find more new points in solving these kinds of problems.

ACKNOWLEDGMENT

This research was supported by Japan Science and Technology Agency (JST) and Core Research for Evolutionary Science and Technology (CREST).

REFERENCES

[1] Al-Ghamdi, A. S., *Analysis of traffic accidents at urban intersections in Riyadh* Accident Analysis and Prevention, 35(5), pp. 717-724, 2003, http://dx.doi.org/10.1016/S0001-4575(02)00050-7

[2] K. Ivan, I. Haidu, J. Benedek, and S. M. Ciobanu*Identification of traffic accident risk-prone areas under low-light conditions* Nat. Hazards Earth Syst. Sci., 15, 2059-2068, 2015. http://dx.doi.org/10.5194/nhessd-3-1453-2015

[3] Ariana Vorko-Jovic, Josipa Kern, Zrinka Biloglav, *Risk factors in urban road traffic accidents* Journal of Safety Research, 37(1), pp. 93-98, 2006. http://dx.doi.org/10.1136/ip.2010.029215.429

[4] John C. Milton, Venky N. Shankar, Fred L. Mannering, *Highway accident severities and the mixed logit model: An exploratory empirical analysi* Analysis and Prevention, Volume: 40, Issue: 1, January, pp. 260-266, 2008. http://dx.doi.org/10.1016/j.aap.2007.06.006

[5] Kim, K., Nitz, L., Richardson, J., Li, L., *Personal and behavioral predictors of automobile crash and injury severity* Accident Analysis and Prevention, 27(4), pp.469-481,1995. http://dx.doi.org/10.1016/0001-4575(95)00001-G

[6] Y. Ohsawa, H.Kido,T.Hayashi,C.Liu,*Innovators Marketplace on Data Jackets for Externalizing the Value of Data via Stakeholders Requirement Communication* Procedia Computer Science,pp.709-716,2013. http://dx.doi.org/10.1007/978-3-319-13545-8_6

[7] Kuhnert, P. M., Do, K. A., McClure, R., *Combining non-parametric models with logistic regression: an application to motor vehicle injury data* Statistics and Data Analysis, 34(3), pp. 371-386,2000. http://dx.doi.org/10.1016/S0167-9473(99)00099-7

[8] ARCHER,J., VOGEL,K.,*The Traffic Safety Problem in urban areas* Royal Institute of Technology Publication,2000. http://dx.doi.org/10.1016/j.aap.2016.03.017

[9] Kathleen,L. Wolf and Nicholas Bratton,*Urban Trees and Traffic Safety: Considering U.S. Roadside Policy and Crash Data* International Society of Arboriculture,pp.170-179,2006. http://dx.doi.org/10.1061/(ASCE)0733-947X(1990)116:1(90)

[10] K. Ivan, I. Haidu, J. Benedek, and S.M.Ciobanu,*Identification of traffic accident risk-prone areas under low-light conditions* Nat. Hazards Earth Syst. Sci.,pp. 20592068, 2015. http://dx.doi.org/10.5194/nhess-15-2059-2015

[11] Zajac, S. S., Ivan, J. N., *Factors influencing injury severity of motor vehicle-crossing pedestrian crashes in rural Connecticut* Accident Analysis and Prevention, 35(3), pp. 369-379,2003. http://dx.doi.org/10.1016/S0001-4575(02)00013-1

[12] Milton, John C., Shankar, Venky N., Mannering, Fred L., *Highway accident severities and the mixed logit model: An exploratory empirical analysis* Accident Analysis and Prevention, Volume: 40, Issue: 1, pp. 260-266, 2008. http://dx.doi.org/10.1016/j.aap.2007.06.006

[13] Cass DT, Ross F, Lam L., *School Bus Related Deaths And Injuries In New South Wales* Med J Austr;166(2), pp.07-108,1997. PMID: 8709875

[14] Darrell.S, Dana.H, *Gender, structural disadvantage, and urban crime:do macro-social variables also explan female offending rates* Criminology,volume38,number 2, pp.403-438,2000. http://dx.doi.org/10.1111/j.1745-9125.2000.tb00895.x

[15] Judith R.Blau, Peter M.Blau, *The cost of inequality: metropolitan structure and violent crime* American sociological review1982, Vol47,pp:114-129,1982. http://dx.doi.org/10.2307/2095046

[16] Adam Krasuski, *A framework for Dynamic Analytical Risk Management at the emergency scene. From tribal to top down in the risk management maturity model* Proceedings of the 2014 Federated Conference on Computer Science and Information Systems, ACSIS,Vol.2,pp.323–330.,2014. http://dx.doi.org/10.15439/2014F371

[17] Yau, K. K. W., *Risk factors affecting the severity of single vehicle traffic accidents in Hong Kong* Accident analysis and prevention,36(3), pp.333-340,2004. http://dx.doi.org/10.1016/S0001-4575(03)00012-5

[18] ODonnell, C. J., Connor, D.H., *Predicting the severity of motor vehicle accident injuries using models of ordered multiple choic* Accident Analysis and Prevention, 28(6), pp. 739-753.,1996. http://dx.doi.org/10.1016/S0001-4575(96)00050-4

FedCSIS'16 Plenary Panel on the Legacy of Professor Zdzisław Pawlak

Background

This is the year of the 90th anniversary of the birth and the 10th anniversary of the death of Professor Zdzisław Pawlak – a Polish mathematician and computer scientist known for his contribution to many branches of theory and applications, a full member of Polish Academy of Sciences, a recipient of Order of Polonia Restituta and many prestigious awards in Poland and abroad. The scientific work of Professor Pawlak is probably most recognized from the perspective of rough set theory that he founded in the early 80's. However, it is important to emphasize also some of his other achievements, e.g., designing the first Polish computer (GAM-1, 1950), introducing a new approach to random number generation (1953, the first international publication of a scientist from Poland in the area of computer science), introducing a positional numeral system with base -2, introducing a generalized class of reverse Polish notation languages, proposing a new formal model of digital machine, creating the first mathematical model of DNA (1965), or proposing a new, very-well received mathematical model of conflict analysis.

Topics and Format

The goal of this panel is to summarize the scientific journey of Professor Zdzisław Pawlak and discuss how to make further progress by following his great intuitions, emphasis on clarity of ideas, as well as willingness to work on interdisciplinary projects. Some of areas of Professor Pawlak's activities will be discussed by his friends and co-workers:

- **Stan Matwin –** Introduction: Professor Pawlak as a Scientist, a Teacher and a Mentor
- **Victor W. Marek –** Working with Zdzisław Pawlak: Excursions in Computer Science and Mathematics
- **Ewa Orłowska –** The Evolution of Rough Sets
- **James F. Peters, Sheela Ramanna –** Gentle Art of Zdzisław Pawlak's Paintings
- **Alicja Wakulicz-Deja, Małgorzata Przybyła-Kasperek –** Pawlak's Conflict Model: Directions of Development
- **Andrzej Skowron** – From Information Systems to Interactive Information Systems

FedCSIS'16 Plenary Panel will form a thematic block with one of FedCSIS'16 / AAIA'16 regular sessions focused on rough sets and approximate reasoning. Moreover, the panel's contents will be coordinated with the analogous event held at IJCRS'16 conference co-organized by International Rough Set Society (IRSS) in Santiago de Chile, October 7-11, 2016.

Panel Moderator

Dominik Ślęzak, University of Warsaw & Infobright, Poland

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 189–190
DOI: 10.15439/2016F002
ACSIS, Vol. 8. ISSN 2300-5963

Working with Zdzisław Pawlak – Personal Reminiscences

Victor Marek
Department of Computer Science
University of Kentucky
Lexington, KY 40506, USA

***Abstract*—This paper accompanies panel contribution of the author to the session devoted to personal reminiscences of Professor Zdzisław Pawlak, a computer scientist and engineer. In particular we discuss some aspects of the work of Pawlak and researchers in his circle of collaborators in 1960ies, and especially, 1970ies. Given the lack of archival materials, the author bases this writing on personal recollections which may, at places, be imprecise.**

***Index Terms*—Zdzisław Pawlak, Information Storage and Retrieval Systems, Rough Sets**

I. How Did It Start

GIVEN that this text refers to personal recollections, I need to introduce myself to the reader. I finished High School in 1960, and started studies in Mathematics at Warsaw University. Soon, I met socially Andrzej Ehrenfeucht and he suggested attending a seminar that a group of scientists conducted at the Mathematical Institute of the Academy of Sciences. It was a series of meetings that can be characterized by a completely informal atmosphere and truly interdisciplinary subject of talks. Among the leaders of the groups were Robert Bartoszyński, Andrzej Ehrenfeucht, Zdzisław Pawlak and other future leaders of Computer Science in Warsaw. The atmosphere was very informal and the hierarchical relationship so often visible in the university education and research was, simply, absent. I was the youngest person attending the seminar. This informal atmosphere was very different from the seminars lead at Warsaw University by Professors Andrzej Mostowski (the head logician in Warsaw) and by Professor Helena Rasiowa. The main topics of these informal seminars were foundations and applications of computers. Of course, there were very few actual computers (but there were some) and the topics were an interesting mixture of entirely theoretical topics (for instance Turing Machines) and Hao Wang's experiments with proving theorems of "Principia Mathematicae" on IBM machines. There was strong presence of individuals interested in biomedical information (or, to be precise, what today is called such). This resulted from the research of Zdzisław Pawlak who, at the time, was interested in DNA as built by formal grammar. Eventually, Pawlak wrote a book (in Polish! [1]) devoted to this research. From my own perspective, the important aspect of the seminar was opening of a perspective – computers and their foundational issues.

In 1964 I completed my M.Sc. in Mathematics studies and joined Professor Mostowski group. This was Foundations of Mathematics group but also comprised of algebraists. Soon, Andrzej Ehrenfeucht emigrated to United States eventually settling in Boulder, CO (where he still teaches). Mostowski and his research group got involved in research on a new (and very exciting) topic called "forcing". This was, at the time completely mysterious, technique for independence proofs in Set Theory. In my case, I worked in this area for a number of years. But in 1970 I went for a post-doc appointment in Utrecht, Holland, in a group of Professor Dirk van Dalen. Their interests were different and were motivated by different aspects of Foundations. A strong influence on my thinking was exercised by Henk Barendregt and his research on λ-calculus. Even more importantly, as I was returning from Holland, I visited Arhus University in Denmark (where my colleague Dr. Janusz Onyszkiewicz was a post-doc) and noticed that logicians there were mainly interested in Computer Science considerations. This made me thinking that maybe it is time to look again at computers.

II. Changed Perspective

The year 1970 and the revolt of workers resulted in significant changes in Poland. From the point of view of Computer Science studies there were significant changes, too. Warsaw University reorganized its science programs. Mathematicians and physicists parted ways. The Mathematics and Physics faculty divided and Mathematics faculty became Mathematics, Mechanics and Computer Science. Computing Center of the Academy of Sciences evolved (first informally, then formally) into the Institute of Computer Science. Mathematicians in the Academy created a venue for Mathematical research called Banach Center. This place welcomed not only "pure" mathematicians but also computer scientists. Yet another important change was creation of "Technical Physics and Applied Mathematics" program at Warsaw Technical University. This program attracted students that were interested in applications, and became a premier program in Computer Science. The alumni of this program included Witold Lipski, Tomasz Imieliński, Mirosław Truszczyński and several other outstanding researchers. At the same time Professors Pawlak and Rasiowa created a publication venue; a journal *Fundamenta Informaticae*. This journal, associated with Polish Mathematical Society became, eventually one of premier places for the publication of Computer Science research. The very name of that journal (with the word *Fundamenta*) alluded to the great

978-83-60810-90-3/$25.00

traditions of foundational research, as done in Poland after WWI.

III. Information Storage and Retrieval Systems

So, after I came back from post doctoral stint in Holland, and seeing that many logicians started to do research in Computer Science, I was certainly open to look at the problems that were grounded in applications of computers. Given the academic system in Poland where one had to write *Habilitation* (a degree established in Germanic and other countries of Europe) took time, sometimes around 1973 I was ready to expand my interests into Computer Science. Then came a series of phone calls for Professor Pawlak (i.e. Zdzisław, recall the 1960ies seminar). After some time I understood the idea. It was a model of databases. Recall that since the work of E.F. Codd on the relational model of database, computer scientists investigated logical formalisms that would eliminate need for the fluency in data structures (such as double linked lists) to operate databases. Eventually several proposals for logic-based languages were proposed and SQL and many of its variations were adopted as a logic-based language for what is known as relational databases. But the Pawlak's proposal [2] differed from the Codd's proposal in several important details. The single table idea was closely related to so-called universal instance. In fact significant effort of a large group of researchers of relational model was spent on decomposing tables, since universal instances were too big to be effectively processed. This research and the enormous effort spent by the database community is at least partly forgotten now, as, with the increased processing capacity of computers, often one denormalizes databases. Other differences of Pawlak's proposal versus SQL-based databases was that SQL admitted null values (thus is, really, based on multivalued logic). But the real difference (and the one that was quietly incorporated into SQL-based systems) was the fact that Information Storage and Retrieval Systems (ISRs), from the very beginning admitted duplicate objects. In other words, different objects could have the same descriptions. This property of ISRs implied existence of (potentially nontrivial) equivalence relation on the set of objects, $\equiv$, namely: $o_1 \equiv o_2$ iff description of o_1 is the same as description of o_2.

The studies of ISRs were conducted by many researchers in Poland and other countries, and the reason for it was that the logicians were ready to investigate ISRs. Let me mention that the group of researchers in Warsaw alone consisted of over 10 individuals. Led mostly by Witold Lipski it included people in the Technical University, Warsaw University and the Academy of Sciences: Tomasz Imieliński, Paweł Traczyk, Michał Jeagermann, Cecylia Rauszer, Andrzej Jankowski, myself, and many others. Lipski and Imieliński tied ISR to the relational model and very soon we were aware of similarities and of differences with Relational Model.

The motivation of ISR came, not surprisingly, from the work of Pawlak with the physicians, more generally, biomedicine researchers. Once one thinks about it, it is clear that computerized medical records of patients may be quite similar or even (after elimination of personal identifiable information) identical. Since the idea in the background was to "mine" the information and find dependencies present in such data, the presence of duplicates was, actually, an interesting information.

IV. Rough Sets, Handling Different Description Languages

The ISRs were based on a first-order language; the objects possessed descriptions. But one observation (motivated by potential biomedical applications) was that there may be more than one language associated with a set of objects. To give one example let us look at a collection of patients. We may have a language associated with symptoms exhibited by the patients but there may be another language to describe the patients, one based on objective data such as various biochemical data (levels of enzymes, presence or absence of specific genes, etc.). The question asked by Pawlak, and one that lies behind the concept of rough sets [3] is this: do we need to specify the description language for the set of objects, or could we, instead, abstract from the specific language and consider equivalence relation "having the same description"? Once this fundamental question was asked, the concept of rough sets, and the associated "interior" and "closure" operations became very natural. Moreover, such approach opened, immediately, the need to look at numerical parameters associated with rough sets (for instance: measures of roughness such as various ratios (for instance of the boundary to the closure), and various characteristics of the equivalence relation (such as the discrepancy - the ratios of sizes of "large" equivalence classes to the small ones). The rest is history; rough sets were introduced, studied, and – most importantly – applied. Like many other fundamental concepts, they were specialized, generalized, investigated for relationship with other concepts (from logic, topology, universal algebra, and combinatorics). Contributions of many researchers, from many countries, and with many interests witness to importance of Pawlak's intuitions.

References

[1] Z. Pawlak, *Grammar and Mathematics (In Polish)*. PZWS, 1965.
[2] V. W. Marek and Z. Pawlak, "Information Storage and Retrieval Systems: Mathematical Foundations," *Theor. Comput. Sci.*, vol. 1, no. 4, pp. 331–354, 1976.
[3] Z. Pawlak, *Rough Sets – Theoretical Aspects of Reasoning about Data*. Kluwer, 1991.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 191–197
DOI: 10.15439/2016F003
ACSIS, Vol. 8. ISSN 2300-5963

Pawlak's Conflict Model: Directions of Development

Alicja Wakulicz-Deja, Małgorzata Przybyła-Kasperek
University of Silesia
Institute of Computer Science
Będzińska 39, 41-200 Sosnowiec, Poland
Email: {alicja.wakulicz-deja, malgorzata.przybyla-kasperek}@us.edu.pl

***Abstract*—The article provides an overview of different approaches to the methods of conflict analysis that are inspired by the model of Zdzisław Pawlak. In the first part of the paper, Pawlak's original model is described. In the second part, the model proposed by Skowron and Deja is discussed. In the third part, the model proposed by the authors is presented.**

I. Introduction

IN THE paper issues related to the problem of conflict analysis are considered. In 1984 Zdzisław Pawlak proposed a simple and intuitive model of conflict analysis. This model allows the relations between the units involved in the conflict to be defined and enables the conflict situation to be visualized and coalitions to be identified. In the model, the concept of agents was introduced and it is a multi-agent system, although it has nothing in common with the definition of the multi-agent systems that are known from the literature [9], [13], [37]. The first articles about this model were written by Pawlak and Skowron [18], [19], [20], [21], [22], [23], [24], [25], [38]. However, the intensive development of solutions to these problems occurred after Pawlak's death and was, in some sense, the realization of his intentions. The extensions of the model that were proposed allow for the analysis of conflicts in different situations as well as for numerous applications in real life situations. Some of these approaches stem from attempts to find solutions to real situations and the effects of different solutions to conflict. In this study, both the original conflict model of Pawlak as well as a review of the directions of its development are presented.

II. Pawlak's conflict model

Conflicts are inscribed in human nature and they have been with us since the dawn of history. Conflict analysis plays an important role in government, politics, business, lawsuits, disputes, negotiations, military operations, labor-management and others. In 1984, in the paper [18], a model to describe the complicated structure of conflict in a simple way was proposed by Zdzisław Pawlak. This issue was further developed by Pawlak in the papers [19], [20], [21], [22], [23], [24]. During the period in which Pawlak was interested in the issue of conflict, some other models of conflict situations were discussed in [1], [8], [10], [12], [15], [36]. Pawlak predicted that rough and fuzzy sets are the perfect candidates for modeling conflict situations in the presence of uncertainty, but at that time, not very much had been done in this area.

In the proposed model, the parties involved in the conflict are called agents. Depending on the type of conflict situation, that is considered, agents can be individuals, groups of individuals, institutions, companies, countries, etc. Each agent expresses his opinion about some discussed issue by assigning one of three values: -1 means, that an agent is against, 0 neutral toward the issue 1 means favorable. Knowledge about conflict situation is written in the form of table, in which rows are labeled by agents, and columns are labeled by discussed issues. The entries of the table are the values that were uniquely assigned to each agent and an issue. Each entry represents opinion of agent about issue. This type of table is an example of an information system $S = (U, A)$, that definition can be found in the literature [16], [17]. In the case of the Pawlak's conflict model elements of the universe U are agents, A is a set of issues, and the set of values of $a \in A$ is equal $V^a = \{-1, 0, 1\}$. The value $a(x)$, where $x \in U, a \in A$ is opinion of agent x about issue a. For each $a \in A$ function $\phi_a : U \times U \to \{-1, 0, 1\}$ is defined:

$$\phi_a(x,y) = \begin{cases} 1 & \text{if } a(x)a(y) = 1 \text{ or } x = y, \\ 0 & \text{if } a(x)a(y) = 0 \text{ and } x \neq y, \\ -1 & \text{if } a(x)a(y) = -1. \end{cases}$$

Then over $U \times U$ three relations are defined: R_a^+ alliance, R_a^0 neutrality, R_a^- conflict, that express the relations between agents:

$$\begin{array}{ll} R_a^+(x,y) & \text{if and only if } \phi_a(x,y) = 1, \\ R_a^0(x,y) & \text{if and only if } \phi_a(x,y) = 0, \\ R_a^-(x,y) & \text{if and only if } \phi_a(x,y) = -1. \end{array}$$

Directly from the definition indicates that the alliance relation is

- reflexive: $\forall_{x \in U} R_a^+(x,x)$,
- symmetric: $\forall_{x,y \in U} \big(R_a^+(x,y) \Rightarrow R_a^+(y,x)\big)$,
- transitive: $\forall_{x,y,z \in U} \big(R_a^+(x,y) \wedge R_a^+(y,z) \Rightarrow R_a^+(x,z)\big)$.

Relation R_a^+ is an equivalence relation. Each equivalence class of alliance relation R_a^+ is called coalition on a.

In order to evaluate views between agents x and y with respect to the set of issues $B \subseteq A$ a function of distance

between agents $\rho_B^* : U \times U \to [0,1]$ is defined

$$\rho_B^*(x,y) = \frac{\sum_{a \in B} \phi_a^*(x,y)}{card\{B\}},$$

where

$$\phi_a^*(x,y) = \frac{1 - \phi_a(x,y)}{2} =$$

$$= \begin{cases} 0 & \text{if } a(x)a(y) = 1 \text{ or } x = y, \\ 0.5 & \text{if } a(x)a(y) = 0 \text{ and } x \neq y, \\ 1 & \text{if } a(x)a(y) = -1. \end{cases}$$

In the definition of the function of distance between agents it is assumed that distance between agents that are in conflict is greater than distance between agents which are neutral. The function of distance between agents for the set of all issues $B = A$ is written in short as ρ.

Now we can in a more general way than before, without reference to specific issues, define the relations between agents. A pair $x, y \in U$ is said to be:

- allied $R^+(x,y)$, if $\rho(x,y) < 0.5$,
- in conflict $R^-(x,y)$, if $\rho(x,y) > 0.5$,
- neutral $R^0(x,y)$, if $\rho(x,y) = 0.5$.

Set $X \subseteq U$ is a coalition if for every $x, y \in X$, $R^+(x,y)$ and $x \neq y$. Coalitions defined in this way does not have to be pairwise disjoint, which was shown in Example 2.1.

Each agent has the strength, that is expressed in the form of non-negative real number $\mu : U \to [0, \infty)$. The strength $\mu(x)$ of agent x can represent economic or military power of a given agent. Each agent distributes his forces against his enemies according to a chosen strategy and knowledge of the situation. A strategy is defined as a function $\lambda : U \times U \to [0, \infty)$ that assigns a non-negative real number to each pair of agents x, y. This number expresses how much strength the agent x directed against the agent y. It is reasonable to assume that for every x, y:

$$\text{if } \rho(x,y) \leq 0.5 \text{ then } \lambda(x,y) = 0,$$

$$\sum_{y \in E_x} \lambda(x,y) \leq \mu(x),$$

where E_x is the set of all enemies of x, i.e., $E_x = \{y \in U : \rho(x,y) > 0.5\}$. The first condition indicates that the strength directed by x against agents who are allied with x or neutral to x is zero. The second condition ensures that the sum of all the forces led by the x against their enemies may not exceed the strength of agent x. Of course, the choice of strategy is essential in the case, which agents will win in a conflict situation, and which will lose.

Pawlak has defined a particularly important strategy, in which each agent has enough strength to destroy all of his enemies. A strategy of intimidation is a strategy λ that fulfills the conditions:

$$\forall_{x \in U} \quad \sum_{y \in E_x} \lambda(x,y) = \mu(x),$$
$$\forall_{x,y \in U} \quad \lambda(x,y) = \lambda(y,x).$$

The strategy of intimidation is unfavorable for all of the agents involved in the conflict, since its realization will cause that all agents will destroy each other.

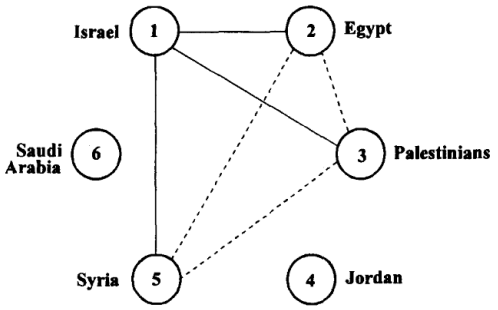


Fig. 1. A graphical representation of the Middle East conflict, issue a

Example 2.1: We consider an example named the Middle East conflict that is from the paper [22]. In the example there are six agents $U = \{1,2,3,4,5,6\}$

1 - Israel,
2 - Egypt,
3 - Palestinians,
4 - Jordan,
5 - Syria,
6 - Saudi Arabia

and five issues $A = \{a, b, c, d, e\}$

a - autonomous Palestinian state on the West Bank and Gaza,
b - Israeli military outpost along the Jordan River,
c - Israeli retains East Jerusalem,
d - Israeli military outposts on the Golan Heights,
e - Arab countries grant citizenship to Palestinians who choose to remain within their borders.

The relationship of each agent to a specific issue is presented in Table I.

As can be seen Egypt, Palestinian and Syria are allied on issue a, Jordan and Saudi Arabia are neutral to this issue whereas, Israel and Egypt, Israel and Palestinian, and Israel and Syria are in conflict about this issue. This can be easily illustrated by a graph. Figure 1 shows a graphical representation of the conflict situation. Agents are represented by circles in the figure. When a pair of agents is in conflict about the issue a, the circles representing the agents are linked. When agents are allied no this issue, the circles representing the agents are connected by dotted line.

The value of the function of the distance between agents is calculated for each pair of agents. These values are given in Table II. Now a graphical representation of a conflict situation, that takes into account all issues being considered by agents, is presented in Figure 2. As before when a pair of agents is in conflict, the circles representing the agents are linked. When agents are allied, the circles representing the agents are connected by dotted line. Neutral agents are not present in this conflict situation. In order to find coalitions, all cliques should be identified in the graph. So the subset of vertices such that every two vertices are connected by dotted line is determined. There are two coalitions in the Middle East conflict $\{1,6\}$ and $\{2,3,4,5\}$.

TABLE I
INFORMATION SYSTEM FOR THE MIDDLE EAST CONFLICT

U	a	b	c	d	e
1	-1	+1	+1	+1	+1
2	+1	0	-1	-1	-1
3	+1	-1	-1	-1	0
4	0	-1	-1	0	-1
5	+1	-1	-1	-1	-1
6	0	+1	-1	0	+1

TABLE II
VALUES OF THE DISTANCE FUNCTION BETWEEN AGENTS

	1	2	3	4	5	6
1						
2	0.9					
3	0.9	0.2				
4	0.8	0.3	0.3			
5	1.0	0.1	0.1	0.2		
6	0.4	0.5	0.5	0.4	0.6	

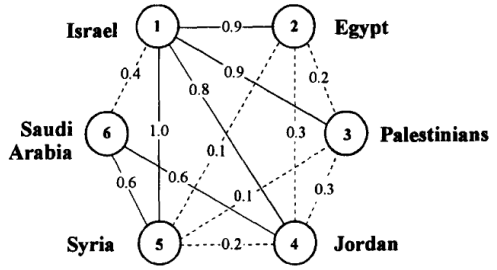


Fig. 2. A graphical representation of the Middle East conflict, distance between agents

The model of conflict analysis that was proposed by Pawlak allows to make advanced analysis of relations between agents and provides guidance that help to decide about strategy. Presented mathematical model is a simple way to illustrate the basic properties of conflicts.

The concepts described above were pursued by many authors [2], [3], [4], [5], [6], [7], [11], [14], [32], [33], [34], [35], [38], [39], [44], [45], [46], [47], [48], [49], [50], [51]. Some of these proposals are briefly described below.

III. CONFLICT MODEL PROPOSED BY SKOWRON AND DEJA

Around 1996, Skowron and Deja proposed conflict model [2], which is an enhancement of the Pawlak model. This concept is described in the papers [3], [4], [5], [6], [38], [7]. Motivation to propose a new model was that, as the authors noted, in the Pawlak model a set of the attribute's values may be too limited in many real situations. Moreover, the assumption that the issues the agents vote represent the issues each agent takes care of causing that the model can be applied in very few real situations. In addition, the reason for the conflict cannot be determined because views on the issues to vote are consequences of the decision taken, based on the local issues, the current state and some background knowledge that are the real cause of the conflict. The main aim of a new model is to define and analyse causes of conflict. It was assumed that the conflict between agents is the consequence of the limited resources which are available to agents in a situation. If the number of resources is insufficient to attain agents' goals it often comes into the conflicts. In the proposed model, the fact was taken into consideration that any agent is describing the situation in its own way. The manner of understanding the same world by each agent can be completely different. A reflection of this assumption in the mathematical model is that for each agent a separate information system is assigned. It was assumed that the sets of attributes of different agents are pairwise disjoint.

Let Ag be a set of agents involved in conflict. Knowledge about the views of agent $ag \in Ag$ on conflict situation is represented in an information system $S_{ag} = (U_{ag}, A_{ag})$, where U_{ag} - is a set of local states and A_{ag} - is a set of attributes. It was assumed that the sets of attributes of different agents are pairwise disjoint

$$\forall_{ag,ag' \in Ag;\, ag \neq ag'} \quad A_{ag} \cap A_{ag'} = \emptyset.$$

For each information system a target function $e_{ag} : U_{ag} \rightarrow [0, 1]$, that determines a subjective evaluation score to each state, is defined. The target function is used to determine the subset of local states, which are accepted by the agent with a fixed threshold value. A set of target states is $T_{ag} = \{s \in U_{ag} : e_{ag}(s) > \mu_{ag}\}$, where μ_{ag} is a level of acceptance, chosen subjectively by the agent ag. The set of target states in this model are presented in the form of Boolean formula.

Conflict is described by a situation and constraints. A situation S is any element of the Cartesian product $S_{Ag} = \prod_{ag \in Ag} U^*_{ag}$, where U^*_{ag} is a set of all possible local states of the agent ag. Constraints come from the bound on the number of resources and describe some dependencies among local states of agents. Constraints restrict the set of possible situations to admissible situations and are described by Boolean formulas.

The situations are evaluated. The score assigned to each situation can reflect the agents preferences (subjective states

evaluation) or the expert judgement - who takes into account the global good. In the second evaluation method, a quality function $q : S_{Ag} \rightarrow [0, 1]$ is defined. The set of situations satisfying a given level of quality t is defined by

$$Score_{Ag}(t) = \{S \in S_{Ag} : q(S) \geq t\}.$$

Boolean formula that describes the set $Score_{Ag}(t)$ is determined. Assessment of the situation expressed by preferences of the agents is defined by the global preference function $p : S_{Ag} \rightarrow R$, $p(S) = F\big(e_{ag_1}(s_1), \ldots, e_{ag_n}(s_n)\big)$, where $S = (s_1, \ldots, s_n)$ and $n = card\{Ag\}$. Aggregation function F may be chosen in many different ways. Boolean formula that describes the set of all preferred situations is also determined.

The multi-agent system, with local states for each agent defined and the global situations satisfying constraints, will be called the system with constraints and is denoted by M_{Ag}. In such systems, conflict can be defined on several levels. Three types of conflicts are distinguished

- local conflict - that arises from the low level of subjective evaluation of the current state,
- global conflict (based on an expert evaluation) - indicates the existence of a situation which is not preferable for the global good,
- global conflict (based on agents preferences) - indicates that the current situation is not preferred by most of the agents.

The concepts of conflict, that are defined above, are the basis for investigation of the most fundamental problem - the possibility to achieve the consensus. In this model, the solution of conflict is searched on various levels: local and global, subjective and objective. In all cases, this is accomplished by determining the conjunction of Boolean formulas that describe appropriate conditions.

IV. Conflict model proposed by Wakulicz-Deja and Przybyła-Kasperek

Around 2009, Wakulicz-Deja and Przybyła-Kasperek proposed conflict model, which is also an extension of the Pawlak model. This concept is described in the papers [26], [27], [28], [29], [30], [31], [40], [41], [42], [43]. Motivation to propose a new model was that, the authors wanted to use conflict model for making decisions based on dispersed knowledge that is stored in many local decision tables. These tables are given in advance and collected by different units for example in different medical centers. It is not assumed that the sets of conditional attributes or the universe of different local decision tables are disjoint or equal. In the situation that is described above the Pawlak model can not be directly applied, because we are dealing with a set of decision tables - not just one information system. The Skowron – Deja model can not be applied, because in order to resolve conflicts the Boolean reasoning is used there and the algorithm has exponential pessimistic execution time. The identification of constraints in the considered dispersed situation is not possible. In addition, the assumption that the sets of conditional attributes of all decision tables must be disjoint is not fulfilled.

The main assumptions that were adopted in the proposed model are that knowledge is stored in several decision tables. There are a set of resource agents, one agent has access to one decision table. The resource agents, that are similar, in some specified sense, are combined into a group. In the process of groups creating elements of conflict analysis and negotiation are used. System has a hierarchical structure. For each group of agents a superordinate agent is defined - a synthesis agent. The synthesis agent has access to knowledge that is the result of the process of inference that is carried out by the resource agents that belong to its subordinate group. Based on local decisions taken by synthesis agents, global decisions are generated using certain fusion methods and methods of conflict analysis.

Different approaches to creating a system's structure were proposed: from a very simple solution to a more complex method of creating groups of agents. Below, a brief overview of the proposed approaches was presented.

The first approach was proposed in the papers [40], [41], [42]. In these articles definitions of resource agents and synthesis agents are given, and the hierarchical structure of the system was established. We call ag in Ag a resource agent if it has access to the resources that are represented by a decision table $D_{ag} := (U_{ag}, A_{ag}, d_{ag})$, where U_{ag} is „the universe"; A_{ag} is a set of conditional attributes, and V^a_{ag} is a set of attribute a values that contain the special signs * and ?. The equation $a(x) = *$ for some $x \in U_{ag}$ means that for an object x, the value of attribute a has no influence on the value of the decision attribute, while the equation $a(x) =?$ means that the value of attribute a for object x is unknown; d_{ag} is referred to as a decision attribute and V^d_{ag} is called the value set of d_{ag}. The only condition that must be satisfied by decision tables of agents is the occurrence of the same decision attributes in all of the decision tables of the agents. In this first approach it was assumed that resource agents taking decisions on the basis of common conditional attributes form a group of agents called a cluster. For each cluster that contains at least two resource agents, a superordinate agent is defined, which is called a synthesis agent as. By a dispersed system we mean

$$WSD_{Ag} = \langle Ag, \{D_{ag} : ag \in Ag\}, As, \delta \rangle,$$

where Ag is a finite set of resource agents; $\{D_{ag} : ag \in Ag\}$ is a set of decision tables of resource agents; As is a finite set of synthesis agents, $\delta : As \rightarrow 2^{Ag}$ is a injective function which each synthesis agent assign a cluster. A significant problem that must be solved when making decisions based on dispersed knowledge has identified - inconsistencies of knowledge may occur within the clusters. This problem stems from the fact that there are no assumptions about the relation between the sets of the conditional attributes of different resource agents in the system. We understand an inconsistency of knowledge to be a situation in which conflicting decisions are made on the basis of two different decision tables that have common conditional attributes and for the same values for the common attributes using logical implications. The process of generating the common knowledge (an aggregated decision table) for each cluster was proposed that was named as the process for the elimina-

tion of inconsistencies in the knowledge. This method consists in constructing new objects that are based on the relevant objects from the decision tables of the resource agents that belong to one cluster. The aggregated objects are created by combining only those relevant objects for which the values of the decision attribute and common conditional attributes are equal. Based on the aggregated decision tables the values of decisions with the highest support of synthesis agents are selected. Global decisions are taken using the DBSCAN algorithm.

The second approach was proposed in the papers [26], [27], [43]. The same methods as in the first approach are used, however, additionally the decision tables of the resource agents are subjected to a certain transformation. In the paper [43], the method of editing and condensing were used on the decision tables of the resource agents. In the paper [26], based on the decision tables of the resource agents decision rules, using rough set theory, are generated. Then, in a single cluster, these rules are aggregated. In the article [27], from each decision tables of the resource agents unnecessary attributes are removed.

The third approach was proposed in the papers [28], [30]. The main modification, compared to the first approach, was conducted within the process of creating a system's structure. In the previous approaches a dispersed system has a static structure (created once, for all test objects), and this time a dynamic structure is used (created separately for each test object). The aim of this approach is to identify homogeneous groups of resource agents. The agents who agree on the classification into the decision classes for a test object will be combined in a group. The modified definitions of relations of friendship and conflict as well as the method for determining the intensity of the conflict, which were introduced by Pawlak, are used. Relations between agents are defined by their views for the classification of the test object to the decision classes. In the first step of the process of clusters creating for each resource agent ag_i a vector of probabilities that reflects the classification of the test object is generated. Then, based on this vector, the vector of ranks $[r_{i,1}(x), \ldots, r_{i,c}(x)]$, where $c = card\{V^d\}$ is generated. We define the function $\phi^x_{v_j}$ for the test object x and each value of the decision attribute $v_j \in V^d$; $\phi^x_{v_j} : Ag \times Ag \to \{0, 1\}$

$$\phi^x_{v_j}(ag_i, ag_k) = \begin{cases} 0 & \text{if } r_{i,j}(x) = r_{k,j}(x) \\ 1 & \text{if } r_{i,j}(x) \neq r_{k,j}(x) \end{cases}$$

where $ag_i, ag_k \in Ag$.

We also define the intensity of conflict between agents using a function of the distance between agents. We define the distance between agents ρ^x for the test object x: $\rho^x : Ag \times Ag \to [0, 1]$

$$\rho^x(ag_i, ag_k) = \frac{\sum_{v_j \in V^d} \phi^x_{v_j}(ag_i, ag_k)}{card\{V^d\}},$$

where $ag_i, ag_k \in Ag$.

We say that agents $ag_i, ag_k \in Ag$ are in a friendship relation due to the object x, which is written $R^+(ag_i, ag_k)$, if and only if $\rho^x(ag_i, ag_k) < 0.5$. Agents $ag_i, ag_k \in Ag$ are in a conflict relation due to the object x, which is written $R^-(ag_i, ag_k)$, if and only if $\rho^x(ag_i, ag_k) \geq 0.5$.

Then, the groups of agents who are in agreement about the classification to the decision classes of the test object are defined. Two different approaches to combining agents in the friendship relations into one cluster were considered.

In the paper [30] the approach is proposed in which disjoint clusters of resource agents remaining in the friendship relations are created. The process of clusters creating in this approach is very similar to the hierarchical agglomerate clustering method and proceeds as follows. Initially, each resource agent is treated as a separate cluster. These two steps are performed until the stop condition, which is given in the first step, is met.

1) One pair of different clusters is selected (in the very first step a pair of different resource agents) for which the distance reaches a minimum value. If the selected value of the distance is less than 0.5, then agents from the selected pair of clusters are combined into one new cluster. Otherwise, the clustering process is terminated.
2) After defining a new cluster, the value of the distance between the clusters are recalculated. The following method for recalculating the value of the distance is used. Let $\rho^x : 2^{Ag} \times 2^{Ag} \to [0, 1]$, let D_i be a cluster formed from the merger of two clusters $D_i = D_{i,1} \cup D_{i,2}$ and let it be given a cluster D_j then

$$\rho^x(D_i, D_j) = \begin{cases} \frac{\rho^x(D_{i,1}, D_j) + \rho^x(D_{i,2}, D_j)}{2}, & \text{if } \rho^x(D_{i,1}, D_j) < 0.5 \text{ and } \rho^x(D_{i,2}, D_j) < 0.5 \\ \max\{\rho^x(D_{i,1}, D_j), \rho^x(D_{i,2}, D_j)\}, & \text{if } \rho^x(D_{i,1}, D_j) \geq 0.5 \text{ or } \rho^x(D_{i,2}, D_j) \geq 0.5 \end{cases}$$

In the paper [28] the approach is proposed in which not disjoint clusters are created. The cluster due to the classification of object x is the maximum, due to the inclusion relation, subset of resource agents $X \subseteq Ag$ such that $\forall_{ag_i, ag_k \in X} \; R^+(ag_i, ag_k)$. Thus, the cluster is the maximum, due to inclusion relation, set of resource agents that remain in the friendship relation due to the object x.

In both approaches, the same definition of dispersed system is used. By a dispersed decision-making system with dynamically generated clusters, we mean

$$WSD^{dyn}_{Ag} = \big\langle Ag, \{D_{ag} : ag \in Ag\}, \{As_x : x \text{ is a classified object}\}, \{\delta_x : x \text{ is a classified object}\} \big\rangle$$

where Ag is a finite set of resource agents; $\{D_{ag} : ag \in Ag\}$ is a set of decision tables of the resource agents; As_x is a finite set of synthesis agents defined for the clusters that are dynamically generated for test object x, $\delta_x : As_x \to 2^{Ag}$ is an injective function that each synthesis agent assigns to the cluster that is generated due to classification of object x.

Also in both approaches the method of elimination of inconsistencies in the knowledge and the DBSCAN algorithm, that were proposed in the first approach, are used.

The fourth approach was proposed in the paper [29]. In this approach a dynamic structure is also used, but the process

of clusters creating is more extensive and as a consequence the clusters are more complex and better reconstructed and illustrate the views of the agents on the classification. The main differences between this approach and the previous approach are as follows. Now, three types of relations between agents are defined: friendship, neutrality and conflict (previously, only two types were used). The clustering process consists of two stages (previously, only one stage process was used). In the first step initial groups are created, it contains agents in friendship relation. In the second stage, a negotiation stage, agents which are in neutrality relation are attached to the existing groups. In order to define the intensity of conflict between agents two function are used: the distance function between agents (was used in the previous approach) and a generalized distance function. The process of clusters creating is as follows. At first the distance between agents ρ^x is defined. Definitions of the relations between agents are modeled on the definitions that were given by Pawlak. Let p be a real number that belongs to the interval $[0, 0.5)$. We say that agents $ag_i, ag_k \in Ag$ are in a friendship relation due to the object x, which is written $R^+(ag_i, ag_k)$, if and only if $\rho^x(ag_i, ag_k) < 0.5 - p$. Agents $ag_i, ag_k \in Ag$ are in a conflict relation due to the object x, which is written $R^-(ag_i, ag_k)$, if and only if $\rho^x(ag_i, ag_k) > 0.5 + p$. Agents $ag_i, ag_k \in Ag$ are in a neutrality relation due to the object x, which is written $R^0(ag_i, ag_k)$, if and only if $0.5 - p \leq \rho^x(ag_i, ag_k) \leq 0.5 + p$. The first step in the process of creating clusters is to define the initial cluster. The initial cluster due to the classification of object x is the maximum, due to the inclusion relation, subset of resource agents $X \subseteq Ag$ such that $\forall_{ag_i, ag_k \in X} \ R^+(ag_i, ag_k)$. After the first stage of clustering we obtain a set of initial clusters and a set of agents which are not included in any cluster. In this second group of agents there are agents which remained undecided. So those which are in neutrality relation with agents belonging to some initial clusters. In the second step, the negotiation stage, this agents play a key role. As it is known the goal of negotiation process is to reach a compromise by accepting some concessions by the parties involved in a conflict situation. In the negotiation process, the intensity of the conflict is determined by using the generalized distance function. This definition assumes that during the negotiation, agents put the greatest emphasis on compatibility of ranks assigned to the decisions with the highest ranks. That is the decisions that are most significant for the agent. Compatibility of ranks assigned to less meaningful decision is omitted. We define the function ϕ_G^x for test object x; $\phi_G^x : Ag \times Ag \to [0, \infty)$

$$\phi_G^x(ag_i, ag_j) = \frac{\sum_{v_l \in Sign_{i,j}} |r_{i,l}(x) - r_{j,l}(x)|}{card\{Sign_{i,j}\}}$$

where $ag_i, ag_j \in Ag$ and $Sign_{i,j} \subseteq V^d$ is the set of significant decision values for the pair of agents ag_i, ag_j. We also define the generalized distance between agents ρ_G^x for the test object x; $\rho_G^x : 2^{Ag} \times 2^{Ag} \to [0, \infty)$

$$\rho_G^x(X, Y) = \begin{cases} 0 & \text{if} \quad card\{X \cup Y\} \leq 1 \\ \dfrac{\sum\limits_{ag, ag' \in X \cup Y} \phi_G^x(ag, ag')}{card\{X \cup Y\} \cdot (card\{X \cup Y\} - 1)} & \text{else} \end{cases}$$

where $X, Y \subseteq Ag$. As can be easily seen, the value of the generalized distance function for two sets of agents X and Y is equal to the average value of the function ϕ_G^x for each pair of agents ag, ag' that belong to the set $X \cup Y$. For each agent which is not attached to any cluster we calculate the value of generalized distance function for this agent and every initial cluster. Then the agent is included to all initial clusters, for which the generalized distance does not exceed a certain threshold, which is set by the system's user. Also agents without coalition, for which the value does not exceed the threshold, are combined into a new cluster. We do not connect agents who are in conflict relation in one cluster. After completion of the second stage of the process of clustering we get the final form of clusters. Then the method of elimination of inconsistencies in the knowledge and the DBSCAN algorithm, that were proposed in the first approach, are used.

The fifth approach was proposed in the paper [31]. In this paper, a method of creating clusters is the same as in the previous approach but when the global decisions are taken the agents' weights are additionally calculated. Different methods of calculating the strength of a cluster was proposed:

- with respect to the number of component agents,
- with respect to the decisiveness of agents,
- with respect to the number of component agents and the decisiveness of agents,
- with respect to the decisiveness-based cluster strength.

Further work is carried out on the development of a dispersed system. The authors propose another approach in order to achieve a better efficiency of the system and always find inspiration in the work of Pawlak. The relations as well as the method for determining the intensity of a conflict between agents that were proposed by Pawlak are always the basis of the entire system.

V. Summary

In this article, the conflict analysis model that was proposed by Pawlak has been described. The aim of the study was to show the main extensions of this model that have been proposed in the literature. Pawlak's model is simple, intuitive and very useful in the analysis of the complex nature of conflicts. The model has many applications and is inspiration for developing new approaches.

References

[1] Coombs, C.H.,Avruin, G.S.: The Structure of Conflicts. Lawrence Erlbaum, London, 1988.

[2] Deja, R.: Conflict analysis. in: S. Tsumoto, S. Kobayashi, T. Yokomori, H. Tanaka, A. Nakamura, (Eds.), Proceedings of the Fourth International Workshop on Rough Sets, Fuzzy Sets and Machine Discovery, The University of Tokyo, 6-8 November, (1996) 118–124.

[3] Deja, R.: Conflict Model with Negotiation. Bulletin of the Polish Academy of Sciences, Technical Sciences, 44 (4), (1996) 475–498.
[4] Deja, R.: Conflict analysis. Proceedings of the 7th European Congress on Intelligent Techniques & Soft Computing, Aachen, Germany, September 13-16, 1999.
[5] Deja, R.: Conflict Analysis. Rough Set Methods and Applications; New Developments. In: L. Polkowski, et al. (eds.), Studies in Fuzziness and Soft Computing, Physica-Verlag, (2000) 491–520.
[6] Deja, R.: Application of rough set theory in conflict analysis. Instytut Podstaw Informatyki Polskiej Akademii Nauk, Dissertation, supervisor: A. Skowron 2000.
[7] Deja, R.: Conflict analysis, Int. J. Intell. Syst., 17, 2, (2002) 235–253, http://dx.doi.org/10.1002/int.10019
[8] Dorclan, P.: Interaction under conditions of crisis: Application of graph theory to internation relations. Peace Research Society International Papers 11 (1969) 89–109.
[9] Franklin, S., Graesser, A.: Is it an Agent, or just a Program - A Taxonomy for Autonomous Agents, Proceedings of the Third International Workshop on Agent Theories, Architectures, and Languages, Berlin: Springer-Verlag, (1996) 21–35.
[10] Hart, H.: Structures of influence and cooperation-conflict. International Interactions 1 (1974) 141–162.
[11] Kanczewski, A.: On regular conflicts. Bull. Polish Acad. Sci. Math. 33 (11/12) (1985) 685–692.
[12] Maeda, Y., Senoo, K., Tanaka, H.: Interval density function in conflict analysis. in: N. Zhong, A. Skowron, S. Ohsuga (Eds.), New Directions in Rough Sets, Data Mining and Granular-Soft Computing, Springer, New York, (1999) 382–389.
[13] Maes, P.: Artificial Life Meets Entertainment: Life like Autonomous Agents, Communications of the ACM (38) (1995) 108–114.
[14] Nabialek, I.: Convex sets of balenced strategies in conflict situations. Bull. Polish Acad. Sci. Math. 36 (1988) 425–428.
[15] Nakamura, A.: Conflict Logic with Degrees. in: S.K. Pal, A. Skowron (Eds.), Rough Fuzzy Hybridization–A New Trend in Decison-Making, Springer, New York, (1999) 136–150.
[16] Pawlak, Z.: Information Systems. Theoretical foundation. WNT, Warszawa (1983) (in polish)
[17] Pawlak, Z.: Rough Sets: Theoretical aspects of reasoning about data, Kluwer Academic Publishers, Boston, 1991.
[18] Pawlak, Z.: On conflicts. Int. J. of Man-Machine Studies 21, (1984) 127–134.
[19] Pawlak, Z.: About conflicts (in Polish), Polish Scientic Publishers, Warsaw, (1987) 1–72.
[20] Pawlak, Z.: Anatomy of conflict. Bulletin of the European Association for Theoretical Computer Science, 50, (1993) 234–247.
[21] Pawlak, Z.: On some issues connected with conflict analysis. Institute of Computer Science Reports, 37/93, Warsaw University of Technology 1993.
[22] Pawlak, Z.: An Inquiry Anatomy of Conflicts. Journal of Information Sciences 109, (1998) 65–78.
[23] Pawlak, Z.: Some remarks on conflict analysis. European Journal of Operational Research 166, (2005) 649–654, http://dx.doi.org/10.1016/j.ejor.2003.09.038
[24] Pawlak, Z.: Conflicts and Negotations. In: Wang, G.-Y., Peters, J.F., Skowron, A., Yao, Y. (eds.) RSKT 2006. LNCS (LNAI), vol. 4062, Springer, Heidelberg (2006) 12–27, http://dx.doi.org/10.1007/11795131_2
[25] Pawlak, Z., Skowron, A.: Rough Sets and Conflict Analysis. E-Service Intelligence: Methodologies, Technologies and Applications, (2007) 35–74, http://dx.doi.org/10.1007/978-3-540-37017-8_2
[26] Przybyła-Kasperek M., Wakulicz-Deja A.: Application of decision rules, generated on the basis of local knowledge bases, in the process of global decision-making, Intelligent Decision Technologies Smart Innovation, Systems and Technologies, Vol. 1, Part 2, (2012) 375-388.
[27] Przybyła-Kasperek M., Wakulicz-Deja A.: Application of reduction of the set of conditional attributes in the process of global decision-making, Fundamenta Informaticae 122 (4), (2013) 327-355, http://dx.doi.org/10.3233/FI-2013-793
[28] Przybyła-Kasperek M., Wakulicz-Deja A.: Global decision-making system with dynamically generated clusters, Information Sciences Volume 270, (2014) 172–191, http://dx.doi.org/10.1016/j.ins.2014.02.076
[29] Przybyła-Kasperek M., Wakulicz-Deja A.: A dispersed decision-making system - The use of negotiations during the dynamic generation of a system's structure, Information Sciences, Volume 288, (2014) 194–219, http://dx.doi.org/10.1016/j.ins.2014.07.032
[30] Przybyła-Kasperek M., Wakulicz-Deja A.: Global decision-making in multi-agent decision-making system with dynamically generated disjoint clusters, Applied Soft Computing, 40, (2016) 603–615, http://dx.doi.org/10.1016/j.asoc.2015.12.016
[31] Przybyła-Kasperek M., Wakulicz-Deja A.: The strength of coalition in a dispersed decision support system with negotiations, European Journal of Operational Research, 252, (2016) 947–968, http://dx.doi.org/10.1016/j.ejor.2016.02.008
[32] Ramanna, S., Peters, J.F., Skowron, A.: Generalized Conflict and Resolution Model with Approximation Spaces. Rough Sets and Current Trends in Computing, 5th International Conference, RSCTC 2006, Kobe, Japan, November 6-8, 2006, Proceedings, (2006) 274–283, http://dx.doi.org/10.1007/11908029_30
[33] Ramanna, S., Peters, J.F., Skowron, A.: Approaches to Conflict Dynamics Based on Rough Sets. Fundam. Inform. 75(1-4) (2007) 453–468.
[34] Ramanna, S., Skowron, A.: Requiremnents Interaction an Conflicts A Rough Set Approach. Proceedings of the IEEE Symposium on Foundations of Computational Intelligence, FOCI 2007, part of the IEEE Symposium Series on Computational Intelligence 2007, Honolulu, Hawaii, USA, 1-5 April 2007, (2007) 308–313, http://dx.doi.org/10.1109/FOCI.2007.372185
[35] Ramanna, S., Skowron, A., Peters, J.F.: Approximation Space-Based Socio-Technical Conflict Model. Rough Sets and Knowledge Technology, Second International Conference, RSKT 2007, Toronto, Canada, May 14-16, 2007, Proceedings (2007) 476–483, http://dx.doi.org/10.1007/978-3-540-72458-2_59
[36] Roberts, F.: Discrete Mathematical Models with Applications to Social, Biological and Environmental Problems, Prentice-Hall, Englewood Cliffs, NJ, 1976.
[37] Russell, S., Norvig, P.: Artificial Intelligence: A Modern Approach, Prentice Hall, Englewood Cliffs, New Jersey, 1995.
[38] Skowron, A., Deja, R.: On Some Conflict Models and Conflict Resolutions. Romanian Journal of Information Science and Technology 3(1-2), (2002) 69–82.
[39] Skowron, A., Ramanna, S., Peters, J.F.: Conflict Analysis and Information Systems: A Rough Set Approach. Rough Sets and Knowledge Technology, First International Conference, RSKT 2006, Chongqing, China, July 24-26, 2006, Proceedings, (2006) 233–240, http://dx.doi.org/10.1007/11795131_34
[40] Wakulicz-Deja A., Przybyła-Kasperek M.: Hierarchical Multi-Agent System, Recent Advances in Intelligent Information Systems, Academic Publishing House EXIT, (2009) 615-628.
[41] Wakulicz-Deja A., Przybyła-Kasperek M.: Global decisions Taking on the Basis of Multi-Agent System with a Hierarchical Structure and Density-Based Algorithm, CS&P, Uniwersytet Warszawski, (2009) 616-627.
[42] Wakulicz-Deja A., Przybyła-Kasperek M.: Multi-Agent Decision Taking System, Fundamenta Informaticae 101(1-2), (2010) 125-141, http://dx.doi.org/10.3233/FI-2010-280
[43] Wakulicz-Deja A., Przybyła-Kasperek M.: Application of the method of editing and condensing in the process of global decision-making, Fundamenta Informaticae 106 (1), (2011) 93-117, http://dx.doi.org/10.3233/FI-2011-378
[44] Wąsowski, J.: Existence of the balanced Strategy in theory of conflicts. Bull. Polish Acad. Sci. Math. 35 (1987) 535–537.
[45] Wiweger, A.: On the notation of a conflict, Bull. Polish Acad. Sci. Math. 34 (5/6) (1986) 381–391.
[46] Xuat, N.V.: Security in the theory of conflicts. Bull. Polish Acad. Sci. Math. 32 (1984) 539–541.
[47] Żakowski, W.: On new characterization of regular configurations in theory of conflict situations, Demonstration Mathematica 17 (1984) 211–218.
[48] Żakowski, W.: Investigation of balanced situation in theory of conflicts, Bull. Polish Acad. Sci. Math. 32 (7/8) (1984) 379–382.
[49] Żakowski, W.: The balanced state in a total-conflict situation, Bull. Polish Acad. Sci. Math. 33 (7/8) (1985) 379–382.
[50] Żakowski, W.: On some properties of sets of configurations in theory of conflicts, Bull. Polish Acad. Sci. Math. 34 (1986) 123–126.
[51] Żakowski, W.: The balanced strategy in an oriented total-conflict, Bull. Polish Acad. Sci. Math. 35 (1987) 525–530.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 199–202
DOI: 10.15439/2016F004
ACSIS, Vol. 8. ISSN 2300-5963

Maximal Nucleus Clusters in Pawlak Paintings. Nerves as approximating tools in Visual Arts

James Peters
Dept Electrical & Computer Engineering
University of Manitoba
Winnipeg, Manitoba R3T 5V6, CANADA
Email: james.peters3@umanitoba.ca

Sheela Ramanna
Dept. of Applied Computer Science
University of Winnipeg
Winnipeg, Manitoba R3B 2E9, CANADA
Email: s.ramanna@uwinnipeg.ca

***Abstract*—This paper is an application of Edelsbrunner-Harer (EH) nerves as approximating tools in discovering interesting perceptual clusters in Pawlak's painting of landscapes, thus giving us an insight into the style of the artist. A variation of EH nerves (collections of Voronoï regions called nucleus clusters) are used in this paper. The Rényi entropy is used to measure the information level of Voronoï regions. It is shown that the information levels (i.e., Rényi entropy) of maximal nucleus clusters in tesselled paintings are the highest compared with surrounding regions, thereby highlighting regions in the paintings with the greatest detail by the artist.**

I. Introduction

IN THIS paper, we are seeking to discover interesting clusters using the concept of nerves useful in visual arts. Visual art forms include paintings, ceramics, photography amongst others. Both spatial and descriptive forms of representation will be considered [1]. First, we start by considering a representative space for the visual information in paintings by Z. Pawlak. *Spatial representation* can be considered in two ways: i) one where a space is given and its characteristics are studied via geometry and topology ii) a space is approximated using some form of tool [2]. *Descriptive representation* starts with probe functions that map features of objects to numbers in $\mathbb{R}^n$ [1]. Probe values provide a description of an object. The problem of finding interesting clusters in an object space X is mapped to the problem of finding interesting clusters in a feature space $\Phi(X)$. The nearness of feature space clusters is studied in the context of proximity spaces.

Various forms of geometric nerves are usually collections known as simplicial complexes in a normed linear space (for details see, e.g. , [3], [1, §1.13]). A nerve $N(C)$ in a finite collection of sets C is a simplicial complex with vertices of sets in C and with simplices corresponding to all non-empty intersections among these sets.

In a descriptive representation, the simplicial complexes are a result of nerve constructions of observations (objects) in the feature space. To construct the simplicial complexes, we tessellate Pawlak Paintings with Voronoi diagram overlays. Then we compute nerves of sets of collections derived from these Voronoi regions [4]. A variation of Edelsbrunner-Harer nerves which are collections of Voronoï regions (called nucleus clusters) are used in this paper. Rényi entropy is used to measure the information level of Voronoï tessellation cells [5]. The focus here is on maximal nucleus clusters (MNCs) that are strongly proximal Edelsbrunner-Harer nerves. A proximity space setting for MNCs makes it possible to investigate the strong closeness of subsets in MNCs as well as the spatial and descriptive closeness of MNCs themselves.

Voronoi tessellation has great utility and has many applications such as geometric modelling in physics, astrophysics, chemistry and biology [6] and in the study of digital images [1], [7], [8]. The form of clustering introduced in this article has proved to be important in the analysis of brain tissue [9]. The contribution of this paper is an application of Edelsbrunner-Harer nerves as approximating tools in discovering interesting perceptual clusters in Pawlak's painting of landscapes, thus giving us an insight into the style of the artist.

II. Definitions Plus MNC Construction

Every Voronoï region of a site s is a convex polygon containing all points that are nearer s than another site in a Voronoï tessellation of a surface. Voronoï regions are *strongly near*, provided the regions have points in common. In Fig. 1, Voronoï region N in the tessellation, is the nucleus of a mesh cluster containing all of those polygons adjacent to N. This form of clustering leads to the introduction of what are known as nucleus-clusters.

A Voronoï *mesh nucleus* is any Voronoï region that is the center of a collection of Voronoï regions adjacent to the nucleus. A *maximal nucleus cluster* is a collection of a maximal number of Voronoï regions that are *strongly near* the mesh nucleus. Maximal nucleus clusters (MNCs) serve as indicators of high object concentration in a tessellated image.

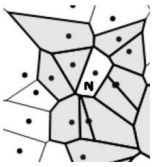


Fig. 1. Voronoï nucleus

Definition 1 (Voronoï region $V(s)$)**.** Let E be the Euclidean plane, $S \subset E$ (set of mesh generating points), $s \in S$.

$$V(s) = \{x \in E : \|x - s\| \leq \|x - q\|, \text{for all } q \in S\}.$$

Nonempty sets A, B in a topological space X equipped with the relation $\overset{\ast\ast}{\delta}$, are *strongly near* (i.e., strong proximity) (denoted $A \overset{\ast\ast}{\delta} B$), provided the sets have at least one point in common.

Definition 2. Nucleus Cluster (see, *e.g.*, Fig. 1). Let X be a collection of Voronoï regions containing N, endowed with the strong proximity $\overset{\ll}{\delta}$, $A \subset X, \mathrm{cl}A = \left\{x \in X : x \overset{\ll}{\delta} A\right\}$ (closure of A), and

$$\mathfrak{C}\, N = \left\{A \in X : \mathrm{cl}\ A \overset{\ll}{\delta} N\right\} \text{ (NC)} \qquad \blacksquare.$$

Let $A, B \subset X$ and let $\Phi(x)$ be a feature vector for $x \in X$, a nonempty set of non-abstract points such as picture points. $A\ \delta_\Phi\ B$ reads A is *descriptively near* B, provided $\Phi(x) = \Phi(y)$ for at least one pair of points, $x \in A, y \in B$ where $\Phi(A) = \{\Phi(x) \in \mathbb{R}^n : x \in A\}$ which are a set of feature vectors

The descriptive strong proximity $\overset{\ll}{\delta_\Phi}$ is the descriptive counterpart of $\overset{\ll}{\delta}$ defined in the feature space $\Phi(A)$. Let regions A, B be described by a feature vector of the form $(x, y, area, diameter)$. Then $A\ \overset{\ll}{\delta_\Phi}\ B$, provided A and B have matching descriptions. Formal proofs of the connection between relations and proximities are given in [1], [4].

Definition 3. Maximum Nuclear Cluster [4]. A nucleus cluster with nucleus N is *maximal*, provided N has the highest number of adjacent polygons in a tessellated surface denoted by $\max\mathfrak{C}N$. Similarly, a descriptive nucleus cluster is maximal, provided N has the highest number of polygons in a tessellated surface descriptively near N, (denoted by $\max\mathfrak{C}_\Phi N$). $\blacksquare$

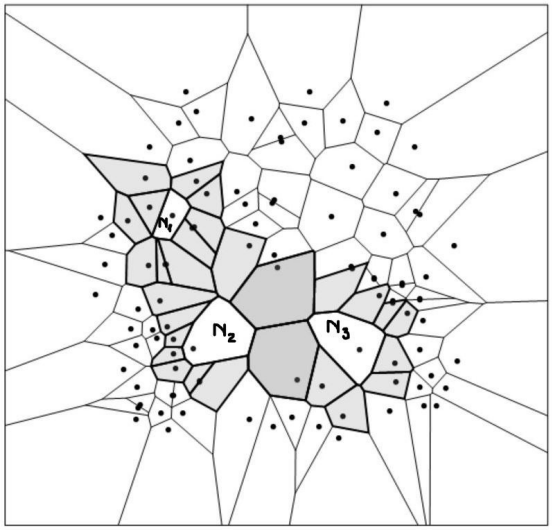


Fig. 2. $\mathfrak{C}\ N_1\ \overset{\ll}{\delta}\ \mathfrak{C}\ N_2$ and $\mathfrak{C}\ N_1\ \overset{\ll}{\delta_\Phi}\ \mathfrak{C}\ N_2$

Example 1. Let X be the collection of Voronoï regions in a tessellation of a subset of the Euclidean plane shown in Fig. 2 with nuclei $N_1, N_2, N_3 \in X$. In addition, let 2^X be the family of all subsets of Voronoï regions in X containing maximal nucleus clusters $\mathfrak{C}N_1, \mathfrak{C}N_2, \mathfrak{C}N_3 \in 2^X$ in the tessellation. Then, for example, $\mathrm{int}\mathfrak{C}N_2 \cap \mathrm{int}\mathfrak{C}N_3 \neq \emptyset$ where int is the interior of a set, since $\mathfrak{C}N_2, \mathfrak{C}N_3$ share Voronoï regions. Hence, $\mathfrak{C}N_2\ \overset{\ll}{\delta}\ \mathfrak{C}N_3 \neq \emptyset$ (for proof, see [4]). Similarly, $\mathfrak{C}N_1\ \overset{\ll}{\delta}\ \mathfrak{C}N_2$.

$\blacksquare$

Algorithm 1: Construct Maximal Nucleus Cluster

Input : Digital images img.
Output: MNCs on image img.

$img \longmapsto TitledImg$/*(Voronoï tessellation)*/;
Choose a Voronoï region in $TitledImg$: *;
$ngon \longleftarrow\!\shortmid TiledImg$;
$NoOfSides \longleftarrow\!\shortmid ngon$;
/* Count no. of sides in $ngon$ & remove it from $TitledImg$. */;
$TiledImg := TiledImg \setminus ngon$;
$ContinueSearch := True$;
while $(TiledImg \neq \emptyset$ *and* $ContinueSearch)$ **do**
  $ngonNew \longleftarrow\!\shortmid TiledImg$;
  $TiledImg := TiledImg \setminus ngonNew$;
  $NewNoOfSides \longleftarrow\!\shortmid ngonNew$;
  if $(NewNoOfSides > NoOfSides)$ **then**
    $ngon := ngonNew$;
  else
    /* Otherwise ignore $ngonNew$: */
  if $(TiledImg = \emptyset)$ **then**
    $ContinueSearch := False$;
    $max\mathfrak{C}N := ngon$;
    /* MNC found; Discontinue search */;

Let $\mathscr{F}$ be a finite collection of sets. An *Edelsbrunner-Harer nerve* (denoted by Nrv $\mathscr{F}$) consists of all nonempty subcollections of $\mathscr{F}$ that have a non-void common intersection, *i.e.*,

$$\mathrm{Nrv}\ \mathscr{F} = \left\{X \in \mathscr{F} : \bigcap X \neq \emptyset\right\}.$$

Lemma 1. [4] *Let $\mathscr{F}_{MNC}$ be a collection of polygons in a Voronoï MNC endowed with the strong proximity $\overset{\ll}{\delta}$. The structure $Nrv\mathscr{F}_{MNC}$ is an Edelsbrunner-Harer nerve.*

Theorem 1. [10, §III.2, p. 59] *Let $\mathscr{F}$ be a finite collection of closed, convex sets in Euclidean space. Then the nerve of $\mathscr{F}$ and the union of the sets in $\mathscr{F}$ have the same homotopy type.*

Theorem 2. [4] *Let the nucleus cluster $\mathfrak{C}N$ be a finite collection of closed, convex sets in a Voronoï mesh V in the Euclidean plane. The nerve $Nrv\mathscr{F}_{MNC}$ in $\mathfrak{C}N$ and the union of the sets in $\mathfrak{C}N$ have the same homotopy type.*

Theorem 3. [4] *Let X be a finite collection of MNC Edelsbrunner-Harer nerves $Nrv\mathscr{F}_{MNC}$ in a Voronoï mesh with nuclei N in the Euclidean plane and let X be equipped with the relator $\left\{\overset{\ll}{\delta}, \overset{\ll}{\delta_\Phi}\right\}$ with strongly close mesh nerves. Each nucleus N has a description $\Phi(N)$ = number of sides of N. Then $\bigcap_\Phi Nrv\mathscr{F}_{MNC} \neq \emptyset$.*

III. Experiments and Discussion

Let $p(x_1), \ldots, p(x_i), \ldots, p(x_n)$ be the probabilities of a sequence of events $x_1, \ldots, x_i, \ldots, x_n$ and let $\beta \geq 1$. Then the

Rényi entropy [11] $H_\beta(X)$ of a set of event X is defined by

$$H_\beta(X) = \frac{1}{1-\beta}\ln\sum_{i=1}^{n} p^\beta(x_i) \text{ (Rényi entropy).}$$

Rényi's entropy is based on the work by R.V.L. Hartley [12] and H. Nyquist [13] on the transmission of information. The information of order β contained in the observation of the event x_i with respect to the random variable X is defined by $H(X)$. Here, $H(X)$ is used to measure the information levels of maximal nucleus clusters in tesselled paintings by Z. Pawlak (places reflecting the greatest detail by the artist).

Fig. 3. 1954 Waterscape by Zdzisław Pawlak

Fig. 4. MNC in 1954 Waterscape by Zdzisław Pawlak

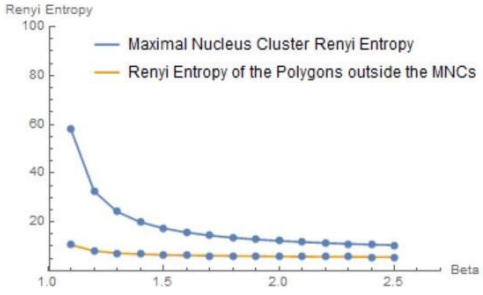


Fig. 5. MNC Rényi's entropy in 1954 Waterscape

The three sample paintings by Z. Pawlak, span 45 years, starting in 1954 and ending in 1999. In Z. Pawlak's paintings, places where the artist rendered with the greatest detail (splashes of colour, slanting brush strokes, clustering of paint

Fig. 6. 1999 Landscape by Zdzisław Pawlak

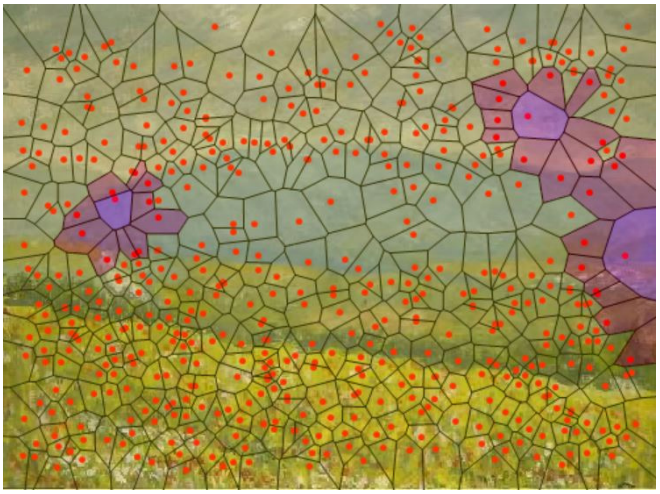


Fig. 7. MNC in 1999 Landscape by Zdzisław Pawlak

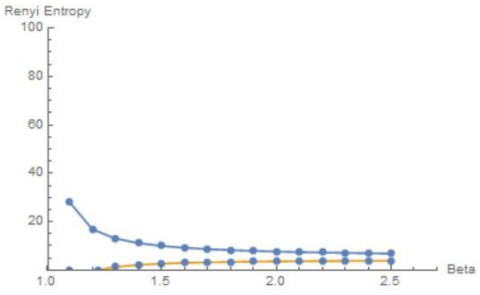


Fig. 8. MNC Rényi's entropy in 1999 Landscape

Fig. 9. 1999 Waterscape by Zdzisław Pawlak

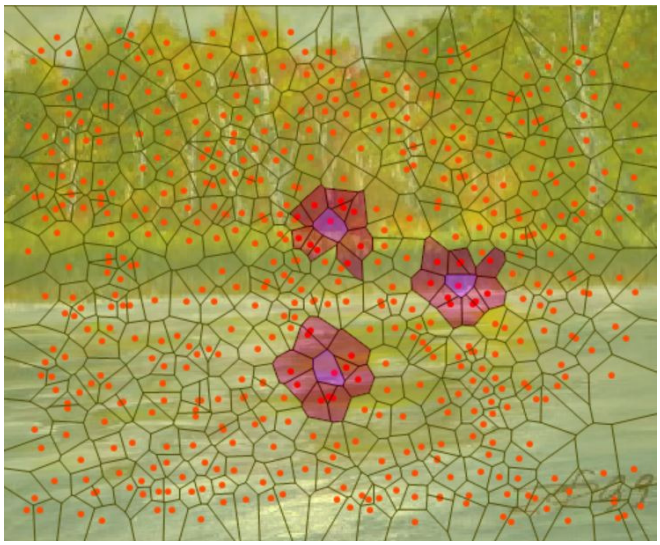


Fig. 10. MNC in 1999 Waterscape by Zdzisław Pawlak

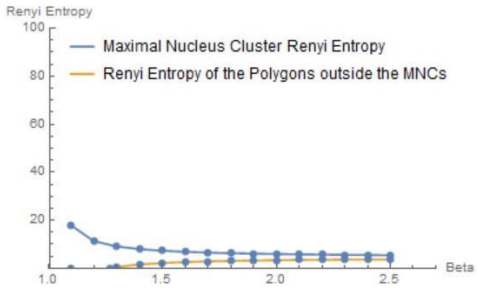


Fig. 11. MNC Rényi's entropy in 1999 Waterscape

patches) occurs where the Rényi's entropy is highest. In the examples, the entropy of the MNCs is consistently higher than the non-MNC areas in the paintings. Each of these paintings reflects a style very similar to the impressionist painting by Oscar-Claude Monet (1840-1926). The basic approach was to express one's perceptions of nature in which he created a record of the French countryside. Similarly, Pawlak's perceptions of Polish countryside are represented by dabs of paint to suggest things like buildings (see the red-roofed building in Fig. 3), long and short strokes of colour (see the trees in Fig. 9 and bits of white for distant roads and villages in Fig. 6). The MNCs in Z. Pawlak's paintings occur in those places in the paintings where the artist has expended his greatest efforts on the detailed woodland and waterscape structures he observed.

In conclusion, this paper presents a Voronoï diagram-based clustering which partitions a descriptive space represented by paintings into regions around a set of seed points. The clustering approach leads to the introduction of *maximal nucleus clusters* that are collections of a maximal number of Voronoï regions that are *strongly near* the mesh nucleus. It can also be observed that there can be several MNCs depending on the number of selected seed points. It is interesting to note that the Rényi's entropy shown in Figures 5,8,11 indicate the entropy of the MNC is higher than those of the surrounding polygons.

REFERENCES

[1] J. Peters, *Computational Proximity. Excursions in the Topology of Digital Images.* Berlin: Springer, 2016, Intelligent Systems Reference Library, 102.

[2] J. Gratus and T. Porter, "Spatial representation: Discrete vs. continuous computational models a spatial view of information," *Theoretical Computer Science*, vol. 365, no. 3, p. 206.

[3] H. Edelsbrunner, "Modeling with simplical complexes," ser. Proceedings of the Canadian Conference on Computational Geometry, Canada, 1994, pp. 36–44.

[4] J. Peters and E. İnan, "Strongly proximal edelsbrunner-harer nerves," *Proceedings of the Jangjeon Mathematical Society*, vol. 19, no. 3, pp. 563–582, 2016.

[5] E. A-iyeh and J. Peters, "Rényi entropy in measuring information levels in Voronï tessellation cells with application in digital image analysis," *Theory and Applications of Math. & Comp. Sci.*, vol. 6, no. 1, pp. 77–95, 2016.

[6] Q. Du and M. Gunzburger, "Advances in studies and applications of centroidal voronoï tessellations," *Numer. Math. Theory Methods Appl.*, vol. 3, no. 2, pp. 119–142, 2010.

[7] R. Hettiarachchi and J. Peters, "Multi-manifold LLE learning in Pattern-Recognition," *Pattern Recognition*, vol. 48, pp. 2947–2960, 2015.

[8] J. Wang, "Edge-weighted centroidal voronoï tessellation based algorithms for image segmentation," Ph.D. dissertation, Department of Scientific Computing, 2011.

[9] J. Peters, A. Tozzi, and S. Ramanna, "Brain tissue tessellation shows absence of canonical microcircuits," *Neuroscience Letters*, vol. 626, pp. 99–105, 2016.

[10] H. Edelsbrunner and J. Harer, *Computational Topology. An Introduction*. Providence, RI: Amer. Math. Soc., 2010, xii+241 pp. ISBN: 978-0-8218-4925-5, MR2572029.

[11] A. Rényi, "On measures of entropy and information," in *Proceedings of the 4th Berkeley Symposium on Math., Statist. and Probability*. University of California Press, Berkeley, Calif., 2011, pp. 547–547, vol. 1, Math. Sci. Net. Review MR0132570.

[12] R. Hartley, "Transmission of information," *Bell Systems Technical Journal*, p. 535, 1928.

[13] H. Nyquist, "Certain factors affecting telegraph speed," *Bell Systems Technical Journal*, p. 324, 1924.

AAIA'16 Data Mining Competition: Predicting Dangerous Seismic Events in Active Coal Mines

AAIA'16 Data Mining Challenge is the third data mining competition associated with International Symposium on Advances in Artificial Intelligence and Applications (AAIA'16, https://fedcsis.org/2016/aaia) which is a part of FedCSIS conference series. This time, the task is related to the problem of predicting periods of increased seismic activity which may cause life-threatening accidents in underground coal mines. Prizes worth over 3,000 USD will be awarded to the most successful teams. The contest is sponsored by Research and Development Centre EMAG (http://ibemag.pl) with support from Polish Information Processing Society (http://www.pti.org.pl/).

Introduction

Providing safety of miners working underground is the fundamental requirement for the coal mining industry in Poland. Coal mining companies are obligated by the law to introduce many safety measures to secure proper working conditions of their underground personnel. However, expert knowledge-based safety monitoring systems sometimes fail to foresee dangerous seismic events which have disastrous consequences. In this data mining competition we would like to address this challenging problem. In particular, we want to ask participants to come up with reliable methods for predicting periods of increased seismic activity perceived in longwalls of a coal mine.

More details regarding the task and a description of the competition data can be found in Task Description section.

Special session at AAIA'16

A special session devoted to the competition will be held at the conference. We will invite authors of selected reports to extend them for publication in the conference proceedings (after reviews by Organizing Committee members) and presentation at the conference. The publications will be treated as short papers and will be indexed by IEEE Digital Library and Web of Science. The invited teams will be chosen based on their final rank, innovativeness of their approach and quality of the submitted report.

Awards

Authors of the top ranked solutions (based on the final evaluation scores) will be awarded with prizes funded by our sponsors:

- First Prize: 1000 USD + one free FedCSIS'16 conference registration,
- Second Prize: 500 USD + one free FedCSIS'16 conference registration,
- Third Prize: one free FedCSIS'16 conference registration.

The award ceremony will take place during the FedCSIS'16 conference (Sep 11-14, 2016, Gdańsk, Poland).

Contest Organizing Committee

Andrzej Janusz, University of Warsaw
Marek Sikora, Institute of Innovative Technologies EMAG
Łukasz Wróbel, Institute of Innovative Technologies EMAG
Sebastian Stawicki, University of Warsaw
Marek Grzegorowski, University of Warsaw
Sinh Hoa Nguyen, University of Warsaw
Dominik Ślęzak, University of Warsaw & Infobright Inc.

References

[1] A. Janusz, M. Sikora, Ł. Wróbel, S. Stawicki, M. Grzegorowski, P. Wojtas, D. Ślęzak: "Mining Data from Coal Mines: IJCRS'15 Data Challenge", in *Proceedings of RSFDGrC 2015,* LNAI 9437, Springer, 2015

[2] M. Kozielski, A. Skowron, Ł. Wróbel, M. Sikora: "Regression Rule Learning for Methane Forecasting in Coal Mines", *Beyond Databases, Architectures, and Structures, CCIS,* Vol. 521, Springer International Publishing, pp. 495-504, 2015

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 205–211
DOI: 10.15439/2016F560
ACSIS, Vol. 8. ISSN 2300-5963

Predicting Dangerous Seismic Events: AAIA'16 Data Mining Challenge

Andrzej Janusz*, Dominik Ślęzak*‡
*Institute of Informatics, University of Warsaw
ul. Banacha 2, 02-097 Warsaw, Poland
{janusza,slezak}@mimuw.edu.pl
‡Infobright Inc.
ul. Krzywickiego 34, lok. 219, 02-078 Warsaw, Poland

Marek Sikora†§, Łukasz Wróbel†§
†Institute of Informatics, Silesian University of Technology
ul. Akademicka 2A, 44-100 Gliwice, Poland
marek.sikora@polsl.pl
§Institute of Innovative Technologies EMAG
Leopolda 31, 40-189 Katowice, Poland

***Abstract*—This paper summarizes AAIA'16 Data Mining Challenge: Predicting Dangerous Seismic Events in Active Coal Mines which was held between October 5, 2015 and March 4, 2016 at the Knowledge Pit platform. It describes the scope and background of this competition and explains our research objectives which motivated the specific design of the competition rules. The paper also briefly overviews the results of this challenge, showing the way in which those results can help in solving practical problems related to the safety of miners working underground. In particular, our analysis focuses on applications of prediction models in order to facilitate the assessment of seismic hazards, in a situation when the exploration of a given working site has just started and there is very little historical data available.**

Keywords*—*data mining competition; multivariate time series data; attribute engineering; cold start problem; hazards assessment;

I. Introduction

THE COAL MINING is one of the most important industries which according to a report by *IBISWorld* employs worldwide over $3.5M$ people [1]. The exploration of coal often requires working in hazardous conditions. Miners in an underground coal mine can face many threats, such as, e.g. methane explosions, rock-burst or seismic tremors. To provide protection for people working underground, systems for active monitoring of the coal extraction processes are typically used. One of their fundamental applications is to screen the seismic activity in order to minimize the risk of severe mining incidents. To facilitate this task, data exploration and decision support tools can be employed, e.g. for predicting seismic activity in the nearest future.

From a data processing point of view, a decision support system which could aid in active monitoring of the coal mining process requires efficient methods for handling continuous streams of data [2]. Such methods have to be able to handle large volumes of data from multiple sensors. They also need to be robust with regard to missing or corrupted data. Moreover, a good decision support system should be easy to comprehend by the experts and end-users who need to have access not only to its outcomes, but also to arguments or causes that were taken into account. A few practical studies have been already conducted with this respect, relying on rule-based models for predicting the methane level [3]. However, the literature on this important subject is still very scarce.

One of very few research initiatives in that field is DISESOR – a Polish national R&D project aimed at creation of an integrated decision support system for monitoring of the mining process and early detection of viable threats to people and equipment working underground [4]. The system developed in the frame of DISESOR project integrates data from different monitoring tools. It contains an expert system module that can utilize specialized domain knowledge and an analytical module which can be applied to make a diagnosis of the mining processes. When combined, these modules are capable of reliable prediction of natural hazards, such as those related to increased seismic activity. The idea to popularize this topic among the data science community by organizing open data mining challenges originated within this project.

The competition described in this paper is the second one in the series. The first one – IJCRS'15 Data Challenge – was focused on the problem of active monitoring and prevention of dangerous methane outbreaks [5]. The task was to design an efficient classifier for multivariate time series data that is generated by various sensors placed in corridors of underground coal mines. The main difficulty in that task was related to the problem of, so called, concept drift [6] and the necessity of constructing robust representation of the available data [7]. This competition was hosted by the Knowledge Pit platform [8] which supports the organization of data mining competitions associated with data science-related conferences.

Following the success of our first competition, AAIA'16 Data Mining Challenge was also organized at Knowledge Pit. This time, however, the task was related to the problem of foreseeing periods of increased seismic activity, that may endanger miners working underground. The main motivation for organizing this challenge as an open on-line competition was the fact that such an approach allows to conveniently review and evaluate performance of the available state-of-the-art methods. It is also an objective way of verifying not only a viability of the predictive models but also whole analytic processes which include preprocessing, feature extraction, model construction and post processing of predictions (e.g. ensemble approaches). Additionally, a huge influence on the final shape of AAIA'16 Data Mining Challenge had our research interest in a severity of the cold start problem for prediction models. In the coal mining context, this problem appears in a situation when the exploration of a given working site has just started and there is very little historical data available that can be

978-83-60810-90-3/$25.00

utilized for a construction of the prediction model for the assessment of seismic hazards.

In the following Section II we reveal details regarding the organization of the data mining competition and then, in Section III, we describe its course and results, including a brief characteristic of the most interesting approaches among the submitted solutions. Next, in Section IV, we show how the competition results were used to conduct an analysis of the cold start problem in the prediction of seismic hazards. Finally, we conclude the paper in Section V by drawing our plans for a continuation of this study.

II. AAIA'16 Data Mining Challenge

AAIA'16 Data Mining Challenge: Predicting Dangerous Seismic Events in Active Coal Mines took place between October 5, 2015 and February 27, 2016. It was organized under auspices of 11^{th} International Symposium on Advances in Artificial Intelligence and Applications (AAIA'16, https://fedcsis.org/2016/aaia) which is a part of the FedCSIS conference series.

The task in this competition was related to the assessment of safety conditions in underground coal mines with regard to a seismic activity and early detection of seismic hazards. In particular, the data set provided to participants was composed of readings from sensors, such as geophones, that monitor the seismic activity perceived at longwalls of different coal mines and measure energy released by, so called, seismic bumps. Each case in the data was described by a series of hourly aggregated sensor readings from a 24 hour period. The provided data also contained information regarding the intensity of recent mining activities at the corresponding working site, coupled by the latest assessments of the safety conditions made by mining experts. Moreover, to further enrich the available data, for each working site that occur in the data set some additional meta-data were made available, such as an identifier of the mine, an identifier of a region where the working site is located or a working site's height.

Participants of the competition were asked to design a prediction model that would be able to accurately detect periods of increased seismic activity. In particular, the target attribute in the provided data (the decision) indicated cases for which the total energy of seismic bumps observed in a following 8 hour period exceeded a warning level of $5 * 10^4$ Joules (i.e. energy released in the period starting after the last hour of aggregated readings describing the case and ending 8 hours later).

In total, the provided data was described by 541 main attributes and 6 additional features related to particular working sites. The competition's data correspond to over 5 years of readings, which to best of our knowledge makes this research the most comprehensive study related to this domain, conducted anywhere in the world.

The data set was divided into a training part which was made available along with the corresponding decision labels and a test part. The labels for the test set were hidden from participants. The division of cases between the training and test sets was made based on a time stamps. In particular, the training data set corresponded to a period between May 5, 2010 and March 6, 2014. It consisted of a total of $133,151$ data rows, each corresponding to a different 24 hour period which were overlapping for consecutive cases. The test data covered the period between March 7, 2014 and June 24, 2015. Unlike for the training set, to facilitate an objective evaluation of solutions and prevent a common problem with, so called, data leakage [9], the test cases were not overlapping and provided in a random order. For this reason the test set used in the challenge was much smaller than the training data. It is important to notice, however, that even though it consisted of only $3,860$ cases, the test set covered a period of nearly 16 months.

Table I shows some basic characteristics of data from each working site that occurs in the competition data. It is worth noticing that not all working site that are present in training data also appear in the test set and there are a few working sites that are present in the test data but not in the training set. Such a situation reflects a real-life problem when the exploration of coal shifts to a new site for which there is no data available. A similar issue can also be identified within other domains, e.g. recommender systems, and is commonly referred to as *the cold start problem* [10]. Noticeable is also the fact that the distribution of cases with the *'warning'* decision label is quite uneven for different working sites.

TABLE I. Basic characteristics for data obtained from different working sites. The first column shows working sites ids, whereas the following columns present information regarding initial expert assessments of the working site's safety, number of data samples in the training and test sets, and the percentage of cases with the 'warning' decision label.

main working site ID	initial mining assessment	number of training cases	number of test cases	training warnings (percent)	test warnings (percent)
146	a	5591	98	0.0014	0.0000
149	b	4248	98	0.0718	0.0018
155	b	3839	98	0.1681	0.0094
171	a	0	49	0.0000	0.0000
264	b	20533	0	0.0039	0.0000
373	b	31236	0	0.0113	0.0000
437	b	11682	0	0.0041	0.0000
470	c	0	258	0.0000	0.0078
479	a	2488	35	0.0000	0.0000
490	a	0	160	0.0000	0.0500
508	a	0	58	0.0000	0.0172
541	b	6429	5	0.0087	0.0000
575	b	4891	253	0.0045	0.0012
583	b	3552	215	0.0021	0.0029
599	a	1196	363	0.0148	0.0289
607	b	2328	209	0.0000	0.0000
641	a	0	97	0.0000	0.0103
689	b	0	83	0.0000	0.1205
703	a	0	145	0.0000	0.0069
725	b	14777	330	0.0920	0.0021
765	a	4578	329	0.0000	0.0022
777	b	13437	330	0.0000	0.0009
793	b	2346	330	0.0000	0.0045
799	a	0	317	0.0000	0.0000
total	-	133151	3860	0.0226	0.0508

A. Evaluation of the uploaded solutions

Participants of the competition had to prepare their solutions in a form of predictions of a likelihood that a given record from the test set has the label 'warning' and send their solutions using the submission system of Knowledge Pit. Each of the competing teams could submit multiple solutions. Quality of the submissions was measured using Area Under the ROC Curve (AUC) [11], [12]. The submitted solutions were

evaluated on-line and the preliminary results were published on the competition Leaderboard. The preliminary score was computed on a subset of the test set, fixed for all participants. Size of this subset corresponded to approximately 25% of the test set and it was composed of data from four working sites with different characteristics. The final evaluation was conducted after completion of the competition using the remaining part of the test data.

Apart from submitting their predictions, each team was also obligated by competition rules to provide a brief report describing its approach. Only the final solutions from teams which sent a valid report could undergo the final evaluation and be published among the competition results. In this way we were able to collect a vast amount of information regarding the current state-of-the-art in predictive analysis of multivariate time series data and objectively verify different methods of preprocessing, feature extraction and post processing of the predictions (i.e. ensemble approaches [13]).

B. A course of a competition

Since one of the main objectives in organization of AAIA'16 Data Mining Challenge was to investigate the cold start problem in the domain of natural hazard detection, we designed this competition in an uncommon way. To gather comprehensive data about an impact of the size of available data on quality of predictions for a given working site, the training data set described above was divided into five separate parts and the course of the challenge was split into six phases. Table II shows some basic participation statistics related to each of the phase.

TABLE II. BASIC PARTICIPATION STATISTICS FOR EACH PHASE OF THE CHALLENGE. IN THE LAST PHASE ALL TRAINING DATA WAS MADE AVAILABLE TO ALL PARTICIPANTS, REGARDLESS OF THEIR ACTIVITY.

	training set size (cases)	number of submissions	best preliminary score	best final score
phase 1	79893	99	0.9296	0.9290
phase 2	93211	278	0.9412	0.9320
phase 3	106527	1377	0.9452	0.9405
phase 4	119839	363	0.9451	0.9375
phase 5	133151	513	0.9452	0.9379
phase 6	133151	505	0.9452	0.9439

After the start of the challenge only the first part of the training data was revealed to participants. The four consecutive parts were made available in approximately monthly intervals (each interval corresponded to a new competition phase), however, only active teams that submitted a required number of files with predictions could access the new data.

In the sixth phase, which lasted for the last two weeks of the competition, all training data parts were revealed to all participating teams regardless of their previous activity in the challenge. It was done to equalize winning chances for teams that decided to join the competition in its latest period.

III. OVERVIEW OF THE COMPETITION RESULTS

AAIA'16 Data Mining Challenge attracted many skilled data mining practitioners who managed to submit a variety of interesting solutions. In total, there were 203 registered teams with members from 31 different countries. The most of participating teams were from Poland (106), however, there were also many teams from countries such as India (14), United Kingdom (12), USA (12), Canada (9) and France (5).

Among the registered teams, 106 were active, i.e. submitted at least one solution to the Leaderboard. In total they submitted 3,236 solutions of which 3,135 were correctly formatted and successfully passed the evaluation procedure. Additionally, 50 teams provided a brief report describing their approach. These reports turned out to be a valuable source of knowledge regarding the state-of-the-art in the predictive analysis of time series data related to early detection of seismic hazards.

TABLE III. FINAL RESULTS AND NUMBER OF SUBMISSIONS FROM THE TOP RANKED TEAMS. THE LAST ROW SHOWS RESULTS OBTAINED SOLELY FROM ASSESSMENTS MADE BY MINING EXPERTS, WHICH WERE AVAILABLE IN THE DATA (ATTRIBUTES *latest_seismic_assessment* AND *latest_comprehensive_assessment*)

team name	rank	n of submission	final result
snm (organizers)	–	2	0.9396
tadeusz	1	31	0.9393
deepsense.io	2	111	0.9384
yata	3	54	0.9342
podludek	4	71	0.9336
jellyfish	5	1	0.9335
millcheck	6	80	0.9329
kkurach	7	32	0.9312
gabd	8	21	0.9299
basakesin	9	30	0.9297
rough	10	4	0.9269
...	...	...	...
experts	(18)	–	0.9196

Table III shows scores achieved by the top-ranked teams. It is worth to notice that the highest result in the final evaluation was obtained by a team involved in DISESOR project and organization of the challenge (team *snm*). Its solution was created using feature extraction methods developed for the purpose of the DISESOR system [7], combined with a rough set approach to reducing data dimensionality [14] and an ensemble learning approach. In order to construct their solution, authors were using only the data available to all participants, however, due to their organizational involvement, team *snm* was excluded from the final ranking. More details regarding this solution can be found in [15].

Among the ranked teams, the highest score was obtained by the team *tadeusz* which was also a subset of the second team in the ranking – *deepsense.io*. Their solution was also based on an ensemble technique. In their approach, authors carefully select a subset of the training data which they later use for constructing and validating the prediction models. Moreover, authors make a significant effort to develop a procedure for an unbiased performance evaluation for tuning parameters of their models and the resulting ensembles. The whole approach is comprehensively described in [16].

In general, the overview of the most successful approaches used by participants suggests that the key steps to achieving good results in this task were:

1) Extracting relevant features (computing a new data representation) that aggregate time series data and are robust with regard to a concept drift.
2) Designing an appropriate evaluation procedure for testing performance of used prediction models and tuning their parameters.
3) Using an ensemble learning techniques for blending predictions of simpler models.

Moreover, the results clearly showed that the proposed task proved to be a challenging one for the most of participants. From the 106 teams that submitted at least one solution only 18 were able to outperform in the final evaluation a simple scoring model that was based on safety assessments made by mining experts. These evaluations were available in the data as two attributes, namely *latest_seismic_assessment* and *latest_comprehensive_assessment*. Even though these features could take only four ordinal values ($a < b < c < d$), a simple logistic regression model that utilizes those two features achieves AUC score of 0.9196 on the final evaluation data (0.9028 on the preliminary test set).

The most likely reason for the weaker results of a large share of participants is over-tuning of their models to the preliminary evaluation set. In a case of many teams, preliminary results were much higher the final scores – the biggest difference was as high as 0.174 (over 17 percentage points). Noticeable is also the fact that in the preliminary evaluation 64 teams obtained a score which was higher than the score of the model based solely on the assessments of experts.

IV. Analysis of the cold start problem

The cold start problem is an important practical issue that is related to real-world applications of many decision support systems. In the case of coal mining, it typically appears when a system for monitoring natural hazards becomes operations for new, previously unexplored longwalls. One of our research objectives motivating the organization of AAIA'16 Data Mining Challenge was to investigate severity of this problem in the context of systems for early detection of periods of increased seismic activity.

For this reason the competition was divided into phases, as it was described in Section II (see Table II for details regarding availability of training data in consecutive phases). Since in each phase a new subset of training data was made available to active participants, we were able to verify the impact of this additional information by examining quality of solutions submitted in consecutive phases. Moreover, thanks to the competition rules that encouraged active participation, we received a large number of diverse solutions for analysis.

Figure 1 presents a distribution of evaluation scores obtained by submissions during the course of the competition. For this analysis we only used valid solutions with a reasonable quality (we disregarded 'random' submissions and those which obtained the preliminary score lower than 0.65). On that plot, black vertical lines denote dates on which additional parts of the training data set were released. Each solution on that plot is marked with a blue and red bar whose height corresponds to the obtained evaluation score. The level of red color in a bar indicates the final score, whereas the level of blue color marks the preliminary evaluation score.

A detailed analysis of the distribution of scores in time reveals some interesting observations. Firstly, in consecutive phases there is a quite conspicuous decrease in differences between the preliminary and final scores. In fact, in early phases of the competition preliminary scores tended to be much higher than the final ones, whereas in the last phase the trend was opposite. In order to confirm the statistical significance of this observation, we used a Wilcoxon rank sum test of preliminary and final scores in consecutive phases. The test confirmed that average differences in phases 1, 2 and 3 and significantly higher ($p-value << 0.01$) than for the phases 4, 5 and 6. Interestingly, in the last phase the differences become negative (final scores are usually higher than the preliminary ones). This phenomenon can be explained by the fact that in the last few days of every data mining competition participants tend to focus on maximizing their score by blending their previous solutions. For this reason we will exclude the last phase from our further analysis of the cold start problem. Table IV shows mean and standard deviation of evaluation scores for each of the competition phases.

TABLE IV. Mean and standard deviation of scores in each of the competition phases. Tha last column gives mean differences between the preliminary and final scores.

phase	prelim. mean	prelim. sd	final mean	final sd	mean diff.
phase 1	0.8590	0.0579	0.8251	0.0672	0.0339
phase 2	0.9059	0.0420	0.8851	0.0587	0.0207
phase 3	0.8683	0.0693	0.8307	0.1058	0.0376
phase 4	0.8868	0.0669	0.8772	0.0831	0.0096
phase 5	0.8943	0.0553	0.8857	0.0625	0.0086
phase 6	0.8820	0.0667	0.8942	0.0696	-0.0122

Another interesting observation related to analysis of the results shown on Figure 1 and displayed in Table IV is that the use of additional training data has a diminishing impact on performance of prediction models. For instance, if we compare average results from the second phase to results from the fourth or fifth phase, we see that the difference is minimal, even though in these phases we received a comparable number of submissions and the available training set data in, e.g. *phase 5*, was by nearly 43% larger than in *phase 2*. This was even less expected due to the fact that the data available in *phase 2* contained information about only 9 out of 21 main working sites present in the test data (these sites corresponded to $\approx 45\%$ of the test set), whereas in *phase 5* this number was much higher (13 out of 21 sites; $\approx 70\%$ of the test set).

To confirm the second observation, in each phase we analyzed the solutions with highest preliminary scores from teams that obtained scores higher than 0.85 – results of such teams better reflect performance of the state-of-the-art models. Figure 2 visualizes basic statistics (min,max,quantiles and mean values) for the preliminary and final evaluations of those submissions.

Conspicuous is the lack of significant differences in the best preliminary evaluation results in consecutive phases. The average final scores slightly increase from phase to phase, however, when we checked the statistical significance of the changes it turned out that a significant difference (p-value lower than 0.01) is only between results from the fifth and sixth phases. For other consecutive phases the p-value of Wilcoxon test was always higher than 0.175.

The above observations allow to formulate a hypothesis that having a sufficiently large data set it is possible to construct efficient prediction models for assessment of seismic hazards. The created models can outperform the currently used expert methods even for completely new working sites, as long as these sites have comparable geophysical properties and the same methodology is used for collecting new data. In order to verify this claim we decided to thoroughly investigate performance of top-ranked solutions submitted in each phase,

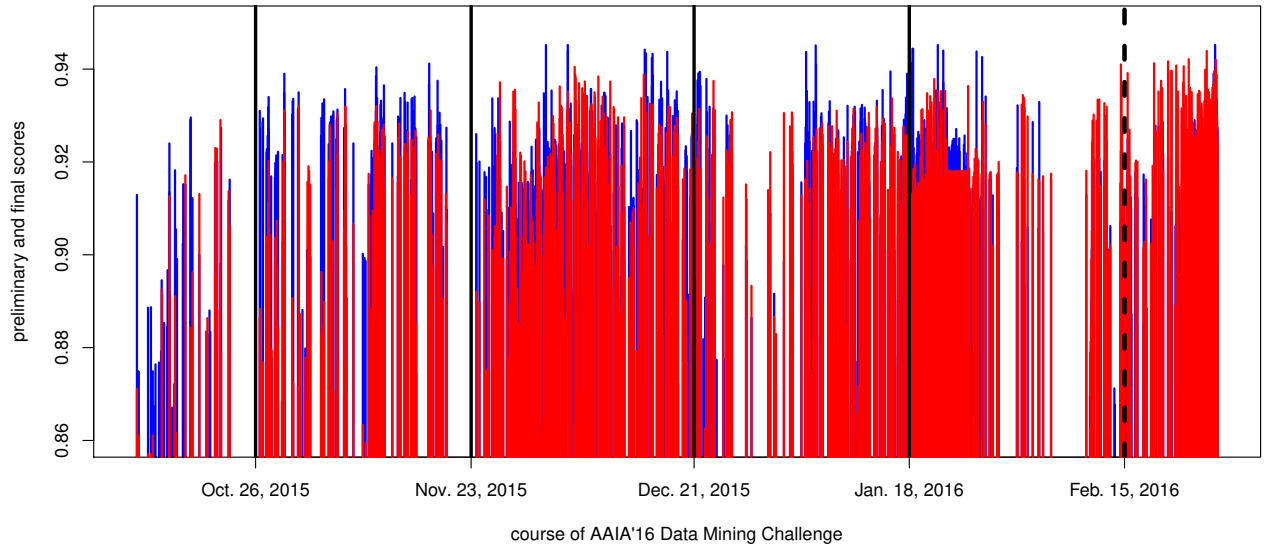


Fig. 1. Distribution of preliminary and final scores during the course of the competition. Blue bars show preliminary scores, wheras the corresponding red bars display final scores. The vertical black lines mark the dates which separate consecutive phases of the competition.

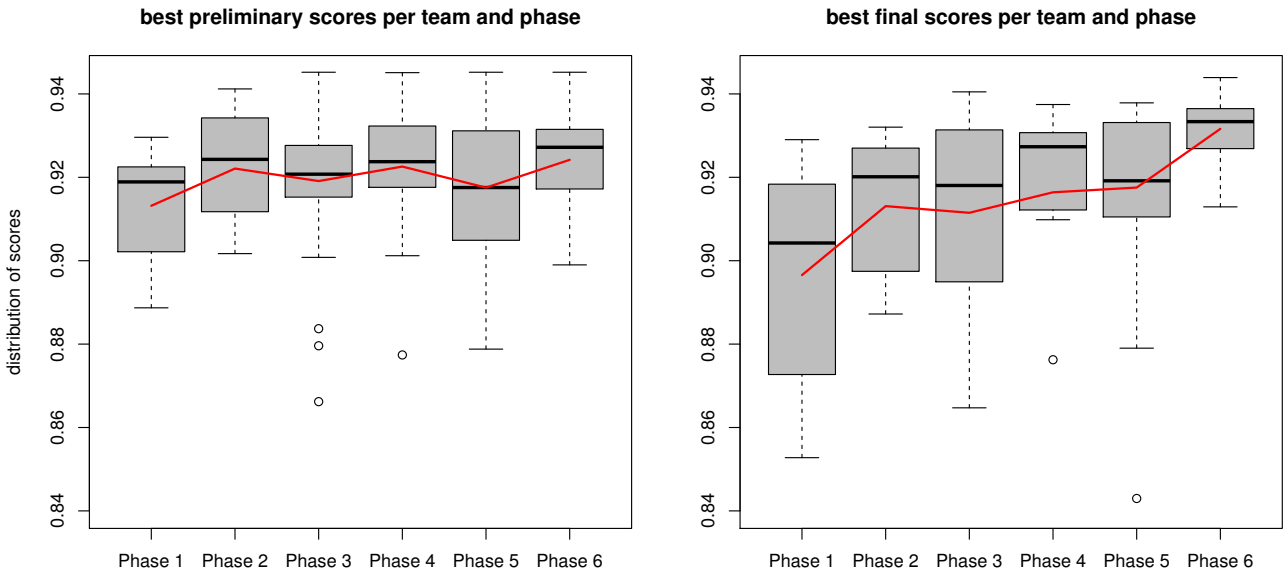


Fig. 2. Distribution of the best preliminary and final scores per team and competition phase. The red lines correspond to average values for given phases.

with regard to individual working sites.

For the purpose of this analysis we disregarded working sites for which there was no examples with the *'warning'* label in the test set. The reason for that was the inability to compute values of AUC on such data subsets. In this way, for the remaining part of our analysis there were 15 working sites left, which corresponded to $\approx 81.5\%$ of the test data.

From solutions submitted in each competition phase, we have chosen 6 with the scores in top 10% for a given phase. During the selection process we considered only solutions uploaded by teams actively participating in the competition, which fulfilled the criteria for obtaining all additional training data. Table V shows their average AUC values with respect to individual working sites. Additionally, the last two rows of the table give average values of AUC for working sites that are present in the training set and for those which are unavailable in the training data, respectively. Finally, the last column of Table V shows AUC values obtained for individual working sites using only the assessments made by experts.

For the most of working sites there is a statistically

TABLE V. AVERAGE SCORES OF TOP SOLUTIONS FOR INDIVIDUAL WORKING SITES, IN DIFFERENT PHASES OF THE COMPETITION. EVALUATIONS OF EXPERT ASSESSMENTS IS GIVEN FOR A COMPARISON IN THE LAST COLUMN. ADDITIONALLY, THE LAST TWO ROWS DISPLAY AGGREGATED VALUES (AVERAGES) FOR WORKING SITES WITH DATA IN THE TRAINING SET AND WITHOUT ANY AVAILABLE TRAINING DATA.

working site ID	phase 1	phase 2	phase 3	phase 4	phase 5	phase 6	expert assessments
149	0.8984	0.9056	0.8523	0.9062	0.8766	0.9005	0.9306
155	0.6578	0.7328	0.7492	0.7393	0.7242	0.7487	0.6845
470	0.9749	0.9922	0.9876	0.9935	0.9922	0.9964	0.9707
490	0.8013	0.8122	0.8340	0.8021	0.7892	0.8289	0.8109
508	0.9825	0.9971	1.0000	0.9942	0.9854	1.0000	1.0000
575	0.9348	0.9845	0.9859	0.9826	0.9820	0.9825	0.9723
583	0.9000	0.9419	0.9363	0.9388	0.9370	0.9401	0.9280
599	0.8391	0.8585	0.8678	0.8445	0.8670	0.8710	0.8020
641	0.9809	0.9983	1.0000	0.9965	1.0000	1.0000	1.0000
689	0.7723	0.8812	0.8523	0.8685	0.8582	0.8938	0.8884
703	0.9346	0.9792	0.9826	0.9699	0.9873	0.9722	0.9722
725	0.8968	0.9188	0.9251	0.9151	0.9176	0.9099	0.8955
765	0.7989	0.7911	0.7367	0.7608	0.7423	0.7808	0.7587
777	0.9118	0.9354	0.9242	0.9252	0.9175	0.9408	0.9444
793	0.9499	0.9545	0.9585	0.9538	0.9361	0.9468	0.8868
avail. in training	0.8653	0.8915	0.8818	0.8852	0.8778	0.8912	0.8670
unavail. in training	0.9077	0.9433	0.9428	0.9374	0.9354	0.9486	0.9404

significant improvement (tested using t-test with a confidence level of 0.05) of results from the later competition phases in comparison to the first phase. However, in nearly all cases the improvement between the second and later phases becomes marginal (one exception is the working site with ID 599). Interestingly, there are event sites (e.g. 689, 777) for which there is a noticeable drop in average quality of solutions between the second phase and phases 3, 4 and 5. Interesting is also the fact that the top solutions obtained consistently higher scores for working sites that were not present in the training data. Explanation of this fact require further analysis.

A comparison of the selected solutions to predictions that were based solely on assessments made by experts revealed that more complex models were able to quickly attain significantly higher scores for working sites with available training data. In the case of the remaining working sites the advantage of complex prediction models was not that clear. The average results for selected models in phase 6 were only slightly higher, however, for a part of investigated solutions the difference was much more favorable than for others.

V. CONCLUSIONS

In this paper we summarized AAIA'16 Data Mining Challenge: Predicting Dangerous Seismic Events in Active Coal Mines which was held at Knowledge Pit platform in association with 11^{th} International Symposium on Advances in Artificial Intelligence and Applications (AAIA'16). We explained research goals that motivated us to organize this competition. We also explained the task in the challenge and briefly described its course. Finally, we showed a detailed analysis of competition's results with an emphasis on the cold start problem.

The conducted analysis revealed several interesting findings regarding the influence of additional training data on performance of prediction models for assessment of seismic hazards. It showed that in order to train prediction methods that aim to work well for a wide range of locations, it is sufficient to provide training data for only several different working sites. Adding more data may have a minimal impact on prediction quality but it definitely helps in computing more reliable estimations of expected prediction performance, as well as in avoiding over-fitting of models to the training data.

Moreover, our analysis confirmed usefulness of the expert methods for assessment of natural hazards. Not only these assessments were able to robustly predict the seismic activity (they outperformed solutions of more than 80% of teams participating in the competition), but also they could be successfully applied to completely new working sites, without a need for using additional training data and complex algorithms.

This observation allows to formulate a general strategy for dealing with the cold start problem: for new working sites start predicting seismic hazards using the expert methods and concurrently gather data for training a more sophisticated prediction algorithm. Initiate your model using data from other working sites and then adjust it using the newly obtained data. Periodically compare performance of your model to results of the expert assessments and switch to your predictions when they become more accurate.

There are still several unanswered questions and research problems that we plan to investigate in our future work. For instance, the competition setting does not allow to study performance of incremental learning methods which can be applied to this problem. We would also like to more thoroughly analyze severity of the concept drift problem which in this context can be related to temporal nature of the data, as well as to changes in characteristics of different working sites. Another important issue is related to a development of methods for identification of good data subsets for training a prediction model for a given working site. Such methods could be based, for instance, on a comparison of similarities between different sites and choosing the data from those with the most similar characteristics.

Finally, in order to guarantee practical applicability of models for the mining industry it is important that mining experts could easily interpret and explain their predictions. For this reason, interpretability of a prediction model may be as important as its performance. The development of efficient algorithms that yield interpretable results is also directly related to a problem of extracting informative, yet compact representation of the training data. These two issues indicate prominent research directions for our future work.

ACKNOWLEDGMENTS

This research was supported by the Polish National Centre for Research and Development (NCBiR) grant PBS2/B9/20/2013 in the frame of the Applied Research Programme.

REFERENCES

[1] IBISWorld. (2016) Global coal mining: Market research report. [Online]. Available: http://www.ibisworld.com/industry/global/global-coal-mining.html

[2] A. Bifet and R. Kirkby, "Data stream mining: a practical approach," The University of Waikato, Tech. Rep., Aug. 2009.

[3] J. Kabiesz, B. Sikora, M. Sikora, and Ł. Wróbel, "Application of Rule-Based Models for Seismic Hazard Prediction in Coal Mines," *Acta Montanistica Slovaca*, vol. 18, no. 4, pp. 262–277, 2013.

[4] M. Kozielski, M. Sikora, and Ł. Wróbel, "Disesor - decision support system for mining industry," in *Proceedings of FedCSIS 2015*, M. Ganzha, L. A. Maciaszek, and M. Paprzycki, Eds., vol. 5. IEEE, 2015, pp. 67–74. [Online]. Available: http://dx.doi.org/10.15439/2015F168

[5] A. Janusz, M. Sikora, Ł. Wróbel, S. Stawicki, M. Grzegorowski, P. Wojtas, and D. Ślęzak, "Mining Data from Coal Mines: IJCRSâĂŹ15 Data Challenge," in *Proceedings of RSFDGrC 2015*, ser. LNCS, Y. Yao, Q. Hu, H. Yu, and J. W. Grzymala-Busse, Eds., vol. 9437. Springer, 2015, pp. 429–438.

[6] M. Boullé, "Tagging Fireworkers Activities from Body Sensors under Distribution Drift," in *Proceedings of FedCSIS 2015*, M. Ganzha, L. A. Maciaszek, and M. Paprzycki, Eds. IEEE, 2015, pp. 389–396.

[7] M. Grzegorowski and S. Stawicki, "Window-Based Feature Engineering for Prediction of Methane Threats in Coal Mines," in *Proceedings of RSFDGrC 2015*, ser. LNCS, Y. Yao, Q. Hu, H. Yu, and J. W. Grzymala-Busse, Eds., vol. 9437. Springer, 2015, pp. 452–463.

[8] A. Janusz, A. Krasuski, S. Stawicki, M. Rosiak, D. Ślęzak, and H. S. Nguyen, "Key Risk Factors for Polish State Fire Service: A Data Mining Competition at Knowledge Pit," in *Proceedings of FedCSIS'2014*, M. Ganzha, L. A. Maciaszek, and M. Paprzycki, Eds. IEEE, 2014, pp. 345–354.

[9] S. Kaufman, S. Rosset, C. Perlich, and O. Stitelman, "Leakage in data mining: Formulation, detection, and avoidance," *TKDD*, vol. 6, no. 4, p. 15, 2012. [Online]. Available: http://doi.acm.org/10.1145/2382577.2382579

[10] L. H. Son, "Dealing with the new user cold-start problem in recommender systems: A comparative review," *Information Systems*, vol. 58, pp. 87–104, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0306437914001525

[11] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, ser. Springer Series in Statistics. New York, NY, USA: Springer New York Inc., 2001.

[12] T. M. Mitchell, *Machine Learning*, ser. McGraw Hill series in computer science. McGraw-Hill, 1997.

[13] A. Janusz, "Combining multiple predictive models using genetic algorithms," *Intelligent Data Analysis*, vol. 16, no. 5, pp. 763–776, 2012. [Online]. Available: http://dx.doi.org/10.3233/IDA-2012-0550

[14] A. Janusz and D. Ślęzak, "Computation of approximate reducts with dynamically adjusted approximation threshold," in *Proceedings of ISMIS 2015*, F. Esposito, O. Pivert, M. Hacid, Z. W. Ras, and S. Ferilli, Eds., vol. 9384. Springer, 2015, pp. 19–28.

[15] M. Grzegorowski, "Massively Parallel Feature Extraction Framework Application in Predicting Dangerous Seismic Events," in *Proceedings of FedCSIS 2016*, M. Ganzha, L. A. Maciaszek, and M. Paprzycki, Eds. IEEE, 2016, in print September 2016.

[16] R. Bogucki, J. Lasek, J. K. Milczek, and M. Tadeusiak, "Early Warning System for Seismic Events in Coal Mines Using Machine Learning," in *Proceedings of FedCSIS 2016*, M. Ganzha, L. A. Maciaszek, and M. Paprzycki, Eds. IEEE, 2016, in print September 2016.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 213–220
DOI: 10.15439/2016F420
ACSIS, Vol. 8. ISSN 2300-5963

Early Warning System for Seismic Events in Coal Mines Using Machine Learning

Robert Bogucki, Jan Lasek, Jan Kanty Milczek, Michał Tadeusiak
deepsense.io,
Email: {robert.bogucki, jan.lasek, jan.milczek, michal.tadeusiak}@deepsense.io

***Abstract*—This document describes an approach to the problem of predicting dangerous seismic events in active coal mines up to 8 hours in advance. It was developed as a part of the AAIA'16 Data Mining Challenge: Predicting Dangerous Seismic Events in Active Coal Mines. The solutions presented consist of ensembles of various predictive models trained on different sets of features. The best one achieved a winning score of 0.939 AUC.**

I. Introduction

IN 2015, the mining industry in Poland reported 2158 dangerous incidents with 19 casualties and 12 severe injuries [1]. Underground mining work poses a number of threats including fires, methane outbreaks or seismic tremors and bumps. Monitoring and decision support systems might play an essential role in limiting the number of incidents and their prevention. Such systems, often based on machine learning or data mining techniques, can be effectively applied to lessen the danger to employees and prevent potential losses arising from lost and damaged equipment, see, e.g., [2], [3], [4].

In this paper, we present a model for predicting dangerous seismic events in coal mines. Using different machine learning models, we address the classification problem of whether the total seismic energy in the upcoming few hours is going to reach a warning level. The model was developed for the *AAIA'16 Data Mining Challenge: Predicting Dangerous Seismic Events in Active Coal Mines* and proved to be the most successful approach among the 203 teams participating in the challenge [3].

The paper is structured as follows: the first section outlines the problem and describes the main challenges. Next, we describe our approach, focusing on feature engineering, model optimization and evaluation. Finally, in the last section we conclude the work.

II. The challenge

In this section we introduce the problem and describe the provided data. We also make some preliminary remarks about the data and its nature.

A. Problem statement

The given problem is a classification task. The goal is to develop a prediction model that, based on the recordings from a 24-hour long period, predicts whether an energy warning level is going to be reached in the upcoming 8 hours. The warning is reached when the total energy of seismic bumps exceeds 50 000 J = 50 kJ. The accuracy of a model is determined with respect to the Area Under ROC curve (AUC) metric. This accuracy measure is defined as follows. Let $(\mathbf{x}_i, y_i) \in \mathbf{X}$ denote an instance from the dataset $\mathbf{X}$, i.e., $\mathbf{x}_i$ stands for the feature vector associated with a single measurement and $y_i \in \{0, 1\}$ stands for its label. Let f be a model that maps each instance to probability that it belongs to class '1' (or, more generally, a real-valued risk score). Then AUC is derived as

$$AUC(f, \mathbf{X}) = \frac{\sum_{i:y_i=0} \sum_{j:y_j=1} \mathbb{1}(f(\mathbf{x}_i) < f(\mathbf{x}_j))}{|\{y_i : y_i = 0\}| \cdot |\{y_j : y_j = 1\}|} \quad (1)$$

where $\mathbb{1}(\cdot)$ denotes an indicator function that returns 1 if a given condition is satisfied or 0 otherwise, and $|S|$ denotes the cardinality of set S. This accuracy measure returns values in the range range $[0, 1]$, where 1 is achieved by a perfect predictor. A random predictor yields values around 0.5.

B. Data

Two sets of observations were provided: training dataset with accompanying labels and the test set without them. The former was provided so that the problem could be approached from a machine learning angle, the latter serves for evaluation purposes. The competitors were asked to submit the likelihood of the label 'warning' for each record in the test set.

In total, the training set consists of 133 151 observations. Each observation (instance) is described by a set of 541 numbers. Below, we briefly introduce the data provided. For a more thorough description of the dataset please refer to the competition website [5].

The instances are described by a set of 13 features of different type and 22 time series over last 24 hours prior to the forecasting period. The time series' names are followed by 1, 2, . . . , 24, indicating consecutive hours of measurements (with the most recent hour prior to forecasting period being 24). Possible suffixes `_e`ξ, $\xi \in \{2, 3, 4, 5, 6\texttt{plus}\}$ refer to orders of magnitude of a given time series in a certain range, e.g. `sum_e3.5` stands for sum of energies within range $(10^2, 10^3]$ in the 5th hour of the time series. The series are listed below:

- `count_e2`, ..., `count_e6plus` - number of registered seismic bumps;
- `sum_e2`, ..., `sum_e6plus` - sum of energy of registered seismic bumps;
- `total_number_of_bumps`;
- `number_of_rock_bursts`;

978-83-60810-90-3/$25.00

- number_of_destressing_blasts;
- highest_bump_energy.

Additionally, for the most active geophones, the following series are provided:

- max_gactivity;
- max_genergy;
- avg_gactivity;
- avg_genergy;
- max_difference_in_gactivity;
- max_difference_in_genergy;
- avg_difference_in_gactivity;
- avg_difference_in_genergy.

There are also 4 *assessments* provided by experts. They are provided as categorical variables with four levels ranging from 'a' (the lowest risk) to 'd' (the highest risk):

- latest_seismic_assessment;
- latest_seismoacoustic_assessment;
- latest_comprehensive_assessment;
- latest_hazards_assessment.

Finally, several features which we will refer to as *general* are provided:

- total_bumps_energy;
- total_tremors_energy;
- total_destressing_blasts_energy;
- total_seismic_energy;
- latest_progress_estimation_l;
- latest_progress_estimation_r;
- latest_maximum_yield;
- latest_maximum_meter.

Metadata: For each observation we are given its location, i.e., a *longwall* in a particular coal mine that the measurement comes from. Each location is accompanied with additional information (metadata included in a separate file):

- main_working_id - ID of the main working site (at a longwall);
- main_working_name - name of the main working site;
- region_name - name of region where the main working site is located;
- bed_name - name of coal bed;
- main_working_type - type of the main working site;
- main_working_height - height of the main working site;
- geological_assessment - geological assessment of the main working site made by experts before the beginning of exploration (ordered categorical variable ranging from 'a' (lower risk) to 'c' (higher risk)).

Most of metadata were unique to the working sites, therefore were discarded early due to the reasons discussed later. The only information of potential use were main_working_height and geological_assessment, however they still had to be treated with caution:

- geological_assessment: A closer insight revealed that there is none mine assessed as 'd' and only one marked as 'c'. It was replaced by 'b'. Moreover, the proportion of 'a' assessments for longwalls in the training and test dataset varied significantly, 25% to 48%, respectively.
- main_working_height: many working sites had unique working heights - this posed a danger that the feature would be used by a model as a proxy for particular location rather than a potentially valuable information about the height. One solution, discussed later, could be to add extra noise, to diminish the relations between the mines and their heights.

The test set consists of 3 860 unlabeled observations. Approximately 25% of them were used for evaluation on the preliminary leaderboard, which was updated throughout the contest when participants submitted their solutions. The remaining observations were used for selection of the best solutions at the end of the competition.

We should also note that the observations in the test set were randomly selected events rather than time series as provided within the training set. More precisely, given a series consecutive observations, samples were uniformly drawn from them to form a test set. If two samples collected laid within the same window of 32 hours (for 24-hour long time series describing seismic activity plus 8 hour window for prediction), one of them was dropped so as to assure that the samples were approximately independent. This procedure removes a significant amount of observations hence the size of the competition test set was relatively modest in comparison to the amount of training data available. This resulted in a very unreliable leaderboard evaluation during the competition that was based on ca. 1 000 observations. Therefore, we put great emphasis and efforts to develop reliable evaluation methods given the available training data as discussed in the next section.

C. Initial remarks

When we approached the problem we quickly realized that the main challenge was to develop a prediction model that generalizes well to new locations. Table I presents the warning frequencies per location in the dataset. We observe that first of all, different locations vary considerably in terms of the frequency of warnings. Secondly, the sets of locations differ between the training and test dataset. Additionally, the test set in the competition originated from future recordings with respect to the training data available. This is the root of the problem. Hence a proposed model should be both location and time independent in the sense that it yields unbiased predictions for instances with no regard to their origin and time they are collected. We also see that the number of instances originating from different locations varies considerably. These preliminary observations should be carefully considered during model building and evaluation steps. We elaborate on this in the next section.

III. The solution

In this section we describe in detail our solution to the given problem. We discuss different sets of features that were proposed, various evaluation methods, models and their set up.

A. Feature engineering

In our experiments we created several feature sets for model training. For the sake of simplicity and completeness, we describe them under consecutive headers and denote as $\mathbf{FS_n}$ which stands for the n-th *feature set* we proposed. These feature sets were developed independently by members of our team. Inevitably, there are significant overlaps between them.

$\mathbf{FS_1}$: The processing of the data focused mainly on aggregation, aiming to reduce the number of the hourly measurements as the majority of them were just zeros (for the training set, about 66% of all numbers were 0). The feature extraction step ended up with 133 features, over 4 times less than the original set. From the original features we kept:

- all *general* features;
- all seismic assessments converted to consecutive integers and their average;
- number of bumps (`count_e*`) and their energies (`sum_e*`) summed over all 24 hours, together with mean energies resulting from division (if `count_e*` was 0, then we were substituting the result by 0);
- number of bursts and the highest bump energies were just summed.

We also aggregated the remaining time series related to most active geophones (8 time series), however this time we introduced some aggregations over subsets of hourly measures based on their relative importance. The process is described below.

In order to assess the impact of features we used a functionality provided by the implementation of Gradient Boosting Trees available in the **XGBoost** [6] package. The library allows building a tree classifier and assessing the importance of particular features by providing the number of times the feature was used in a split. The more often a feature is used, the more separation gain it offers and therefore the more important it is. We used an XGBoost classifier with 150 trees (other parameters were default). Fig. 1 presents an example of such feature importance analysis for `avg_genergy`. It seems that features are gaining importance towards the end of the time-series - it agrees with the intuition that the measurements closer to the forecasting period are more informative. Therefore in this case, apart from the entire time-series statistics, we are also interested in statistics based on the last five hours (they stand out from the preceding hours). Also, we keep the measurements from the very last hour as a separate feature. Having applied analogous analysis to the above feature groups, we selectively compute statistics such as:

- average and average over absolute values;
- standard deviation;
- max and max over absolute values;

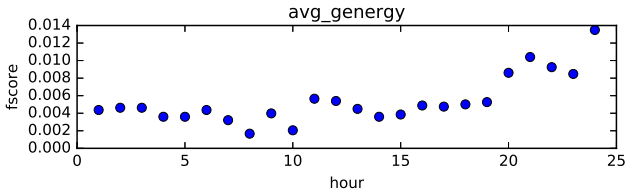


Fig. 1. Importance of hourly measurements of `avg_genergy`.

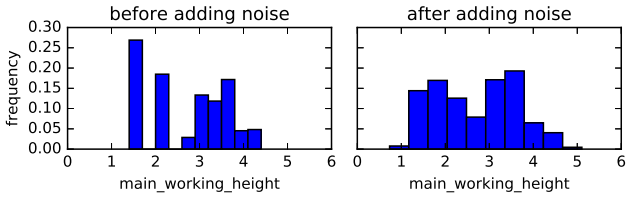


Fig. 2. Distribution of `main_working_height`, before and after adding noise.

- average over last γ hours (were γ varies from 1 to 6);
- standard deviation over last γ hours;
- slope of a linear regression over last 5 hours with respect to time.

As mentioned above, competitors were also provided with the metadata describing specific mine sites. Most of them were discarded. The only metadata used here were the *main working height*, but only after adding Gaussian noise ($\sigma = 0.2$) resulting in more even distribution, see Fig. 2. This step was performed to prevent a model from recognizing a particular location by its height.

In addition to the above features, we produced a vast set of more than 6 000 interactions between them, i.e., pairwise products of features. This is obviously an exhaustive number and we were not planning to use all of them. However, some interactions proved to be valuable. We applied an iterative process of selecting the most promising subsets of features and their interactions. We will come back to them when describing the final model (Section III-C Model$_1$).

$\mathbf{FS_2}$: In constructing this set, at first we decided to drop time series describing maximum statistics (`max*`) since they were highly correlated with corresponding average records. For the series `count_e*`, `sum_e*`, `number_of_rock_bursts`, `number_of_destressing_blasts`, `avg_gactivity`, `avg_genergy` we extracted the following features:

- minimum,
- maximum,
- standard deviation,
- indicator variable, if there is a non-zero value in the series,
- hours elapsed from the last non-zero observation.

Moreover, these statistics were derived over the window of the last 2, 4, 8 and 24 hours prior to the forecasting period in order to describe the most recent data in greater

detail. These features were appended to data with the original time series that they were computed for. Additionally, for series `avg_difference_in_gactivity` and `avg_difference_in_genergy` maximal absolute value was derived over the last 2 and 24 hours. Finally, the categorical variables were converted to binary features using one-hot encoding, i.e., for each possible value of a categorical a separate column was created which indicates that a given observation has this particular category. These operations produced a feature set with a total number of 700 features.

FS_3: In this set of features, we first derived several new series based on the original ones

1) `log_max_avg_diff_genergy` which was derived as difference in `max_genergy` and `avg_genergy`, with application of logarithm hereafter. Analogous operation was performed for other series `avg` and `max_difference_in_genergy` and a corresponding series for `gactivity`
2) `log_ave_energy` series was produced by computing average energy based on `sum` and `count` series.

In addition to features enumerated for **FS_2**, we derived the following statistics:

- 0.25, 0.5 and 0.75-quantiles,
- number of times a series increased in comparison to the previous hour's recording,
- number of positive values in a series,
- indicator variable, if there is a non-zero value in the series.

The statistics were computed on 4, 8 and 24 hours window. Furthermore, we computed the coefficient, intercept and R^2 statistic for a fit of linear model of series `avg_difference_in_gactivity`, `avg_gactivity`, `log_ave_energy` to an independent hourly temporal variable (1, 2, . . . , 24). Finally, we computed correlations between `avg_difference_in_gactivity` and `avg_difference_in_genergy` as well as `avg_gactivity` with `avg_genergy`. After extracting features, constant features were dropped from the feature set. Also, if there were features that were correlated over 0.99 (according to Pearson correlation coefficient), one of them was removed. For categorical variables, they were one-hot encoded for logistic regression model or converted to integers (with higher risk categories being assigned a higher integer) for tree-based models.

These steps produced a training set of 426 features (for the integer encoding of categorical features).

FS_4: The feature set that was created with the goal of being simple and as such leaving little room for overfitting. Out of the basic (not time-based) features, `main_working_id` was dropped. Out of the metadata, only `geological_assessment` was used. The time-based features were ran through maximum and standard deviation functions on 8-hour time periods with 4 hour increments, only the features concerning quantities and maximums were used (features listing averages and sums were left out).

B. Evaluation procedures

Evaluation methodology is a crucial part of creating a successful application of a model. Below we list different validation techniques that we employed to assess the accuracy of a model. Here, the issue of overfitting a model to particular locations and time-frames of samples is considered in detail.

k-fold cross validation (k-CV): This is one of the basic validation procedures. It is performed by assigning each example in the training set randomly to one of k folds (in our application we used $k = 10$ or 20). Note that due to temporal alignment of instances in the training data, this evaluation procedure tends to produce overly optimistic evaluation scores (we observed that during the contest by, e.g., large discrepancy in local evaluation and leaderboard scores). This is because consecutive instances are likely to share the same label. If some of them pertain to a training fold and the others to test fold, then a classifier has a relatively easy task to assign this instance to the proper label.

Leave one location out (LOLO): This evaluation method was chosen to estimate the model's performance on mining sites not included in the training data (see Table I). It was supposed to promote models that overfit less and filter out those whose good performance was actually based on data leaks. We have decided to not use the three largest locations (with IDs 264, 373 and 437) for testing. These three locations constitute to a large portion of the total training data (48%) and were not appearing in the test set. Locations that had no 'warning' labels were also not used as validation data, as AUC could not be computed for them. This approach resulted in a 8-fold cross-validation that gave much lower scores than the other ones (not even the best models could exceed 0.9 AUC), and the scores varied between folds (from as low as 0.6 to as high as 0.999) but it should not be percieved as a flaw — it was intended behavior.

Train and test split #1 ($TrTs_1$): This evaluation methodology was devised to reflect the way the leaderboard was constructed. It is based on multiple train and test splits of the data. It proceeds in two steps:

1) 5 series are chosen at random and included in the validation set,
2) among series that have not been selected in 1) we include the first 70% observations in the training sample and the other 30% in the validation set.

Moreover, in each of those 70%-30% splits, 32 observations between the split point were removed to assert approximate independence between the training and validation set (by introducing a gap of 32 hours between them). Again, data from locations with IDs 264, 373, 437 where included only in the training set.

In order to arrive at a reliable error estimate, this evaluation was repeated 25 times and consecutive measurements were

averaged. With that many iterations we arrived at stable results for mean AUC value.

Train and test split #2 ($\mathbf{TrTs_2}$): The evaluation was based on multiple train and test splits (20 in the final model) with some restrictions. By comparing the `total_seismic_energy` (TSE) of mines (which turned out to be linearly correlated with the frequency of appearances of warnings) we tried to make the split, so the TSE in the inferred test sets resembled the level of energies in the private test set.

TABLE I
NUMBER OF INSTANCES ORIGINATING FROM DIFFERENT LOCATIONS IN THE TRAINING AND TEST SET ALONG WITH AVERAGED TOTAL SEISMIC ENERGIES (TSE) AND FREQUENCY OF WARNINGS (NOT AVAILABLE FOR TEST SET CASES).

Mine ID	Train set			Test set	
	Instances	Mean TSE [J]	Warnings Frequency	Instances	Mean TSE [J]
373	31236	81002	1.1%	-	-
264	20533	7563	0.4%	-	-
725	14777	190232	9.4%	330	106741
777	13437	0	0.0%	330	29061
437	11682	4727	0.4%	-	-
541	6429	9397	0.9%	5	324
146	5591	678	0.1%	98	1
575	4891	9775	0.5%	253	7503
765	4578	136	0.0%	329	51265
149	4248	48357	7.3%	98	72749
155	3839	322021	17.2%	98	527229
583	3552	2595	0.2%	215	73302
479	2488	5548	0.0%	35	102
793	2346	0	0.0%	330	11547
607	2328	6027	0.0%	209	9470
599	1196	29932	1.9%	363	39962
171	-	-	-	49	33
470	-	-	-	258	10701
490	-	-	-	160	13698
508	-	-	-	58	32183
641	-	-	-	97	10672
689	-	-	-	83	63889
703	-	-	-	145	44031
799	-	-	-	317	8

Table I presents averaged TSE for each mine grouped over train and test datasets, together with the frequencies of warnings in the training dataset. It is worth to point out significant discrepancies between the activity levels of mines in both sets. For mine 765, the activity in the training set is mere 136 J, with no warnings. In the test set, the average activity is above 50 kJ, so there must have been several warnings emitted. A closer look reveals that there are some abnormalities in the training set. Fig. 3 presents the TSE of mines 155 and 765. While the activity of the former looks realistic, 765 is mute for majority of the time, only to exhibit a few spikes towards the end of the time series. On the other hand, its activity in the test set greatly increased. Some mines do not exhibit any activity in the training set, i.e. TSE equals zero (mines 777, 793). This is one of the reasons we have to avoid producing models that would be able to recognize the mines, the classifiers should generalize correctly from the activity records, regardless any behavior specific to certain mines. Also, it poses a problem - whether to consider the suspicious

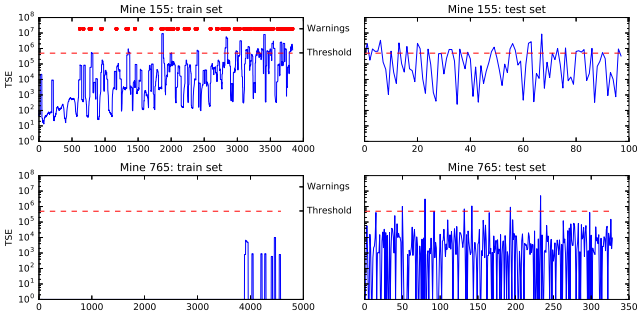


Fig. 3. Total seismic energy over time.

mines during the training or not. It is rather unusual for a mine to have a zero seismic activity and supposedly the data in these periods might be corrupted.

The final train and test splits were based on the above TSE analysis and were produced in the following way:

1) mines 777, 793 were excluded due to their suspicious lack of any activity in the training set (although they represented a significant amount of data);
2) in every split, five randomly selected mines were left only for testing (to evaluate the generalization properties of classifiers);
3) from the remaining sites we were taking 20% of samples for testing. For mines 146 and 599 samples were drawn from the beginning (due to corresponding energy levels in the test set), for the remaining - from the end;
4) in some cases (mines 373, 437) only samples where TSE were nonzero were taken into account.

The process was repeated several times to obtain multiple train/test splits. The final evaluation was based on the average score over 20 splits.

C. Model training and optimization

There are several models that we employed in creating the final solution to the problem. We used the implementation of models available in Python's **scikit-learn** package for machine learning (ver. 0.17.1) [7], [8] and **XGBoost** package for tree boosting models (ver. 0.4) [6]. Throughout the paper, for brevity, we use the following abbreviations for the model names: Linear Discriminant Analysis - LDA, Logistic Regression - LR, Extra Trees Classifier - ETC (all from **scikit-learn** library) and Extreme Gradient Boosting Classifier - XGB (from **XGBoost** library).

$\mathbf{Model_1}$: The fist model was built using $\mathbf{FS_1}$ and $\mathrm{TrTs_2}$ evaluation method. Several models were considered, apart from XGB and ETC, also logistic regressions and neural networks, finally only the first two were used in the final blend. They were performing particularly well in spite of rather large number of features.

First, we ran learning on all the features and interactions we produced. Based on the importance scores provided by XGB (described in Section III-A $\mathbf{FS_1}$) we kept the first 982

interactions and all *individual* features. Then, a grid search returned sets of parameters scoring the highest:

The optimal parameters for XGB model were (otherwise default):

- `n_estimators = 100`
- `max_depth = 2`
- `learning_rate = 0.08`

The optimal parameters for ETC were (otherwise default):

- `n_estimators = 1000`
- `max_depth = 7`
- `criterion = entropy`

It is worth to note, that trees, by their design, are relatively powerful in discovering interactions between features. However, in their case the interactions are not discovered concurrently, but rather in a multilevel manner, between consecutive splits. By explicitly using interactions as features, they can be made use of directly.

Having obtained well performing hyperparameters, we ran a randomized search for best features' subsets. In each iteration we were randomly selecting from 20 to 40 individual features (out of 133) and additionally up to 10 interactions (out of 982). We ran several thousands evaluations on XGB and several hundreds on ETC, tracking their validation scores.

The idea was to produce many models built only on subsets of features and to take advantage of assembling them which reduces variance of predictions and minimize the risk of overfitting to anomalies in particular features. This is a powerful method for increasing the performance of the model [9].

The final blend was composed of:

- single ETC of 10 000 trees using all 133 features and 20 best interactions;
- single XGB using the same features;
- a blend of 20 ETCs built on 20 best subsets of features;
- a blend of 20 XGBs built on 20 best subsets of features;

The final submission scored 0.9199 on the public leaderboard. The score in the final evaluation reached 0.9393 and turned out to be the best in the competition.

Model$_\mathbf{2}$: The second model involved the following classifiers: two linear models (LDA and LR) as well as the tree-based ETC model.

The first part of the solution was the LDA model trained on **FS$_\mathbf{2}$** using k-CV evaluation procedure. The regularization shrinkage parameter selection was done in an automated way (i.e., the parameter `shrinkage` set to "`auto`") in **scikit-learn**'s LDA implementation. The other models were LR and ETC trained on **FS$_\mathbf{3}$** using TrTs$_1$ evaluation method. The parameter values were set using grid search. The optimal values for LR model were:

- `penalty = l1`
- `C = 0.003`

The optimal parameters for ETC were:

- `n_estimators = 1000` (number of trees)
- `max_depth = 3`
- `max_features = 200`
- `min_samples_split = 3`
- `class_weight = 10` (for label '1').

The three models were blended by averaging their predictions with equal weights to produce a solution. Prior to averaging, the model predictions were standardised so that their standard deviations would equal 1. This step aims to convert the probabilities yielded by individual models to the same scale. Note that the mean values of predictions are irrelevant since AUC is invariant to monotonic transformations of output, see Equation 1. On the competition test set, the model yielded 0.9385 and 0.9340 of preliminary and final evaluation score, respectively.

Model$_\mathbf{3}$: This model used only **FS$_\mathbf{4}$** and was meant to be more universal than the other models and thus was tuned on LOLO validation. The algorithms used were ETC, XGB and logistic regression. For each algorithm, many sets of predictions were generated (using the top results from a grid search). This model achieved 0.928 and 0.933 on preliminary and final evaluations, respectively.

Below we list the best parameters found for each algorithm:
ETC

- `min_samples_leaf = 5`
- `n_estimators = 40 000`

XGB

- `subsample = 1.0`
- `num_round = 200`
- `max_depth = 10`
- `objective = binary:logistic`
- `base_score = 0.05`
- `eta = 0.04`
- `colsample_bytree = 0.8`

LTR

- `solver = sag` (Stochastic Average Gradient, just for speed)
- `C = 1.0`

D. Model ensemble

We have decided to use sorted order position averaging (as the AUC assessment method considers only the rank of predicted likelihoods and not the values) of the three presented models' predictions with the final weights being 1, 3 and 2 for models 1, 2 and 3 respectively. The averaging was employed in order to leverage various approaches and come up with yet a better predictor for the given task. The weights for the ensemble were chosen basing on individual model's scores on the preliminary leaderboard. The ensemble produced a model scoring 0.933 and 0.938 on preliminary and final leaderboard, respectively. All in all, it turned out that model 1 outperformed the full ensemble by a small margin (0.939 to 0.938). However, it might be caused by the relatively small test set size.

E. Things we tried that did not work

Throughout the process of creating the most successful model we tested a couple of ideas that turned out not to

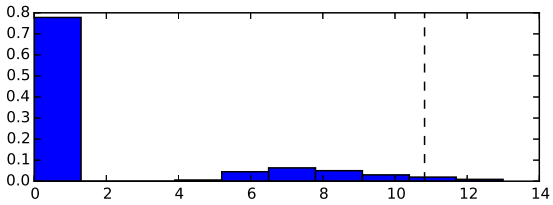


Fig. 4. Histogram of the total energy within 8 next hours (after with application of $\log(x+1)$ transformation). The dashed line indicates the warning level of $\log(5 \cdot 10^4 + 1)$ J.

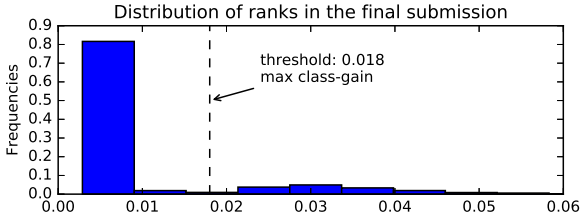


Fig. 5. Histogram of likelihoods returned by the winning model.

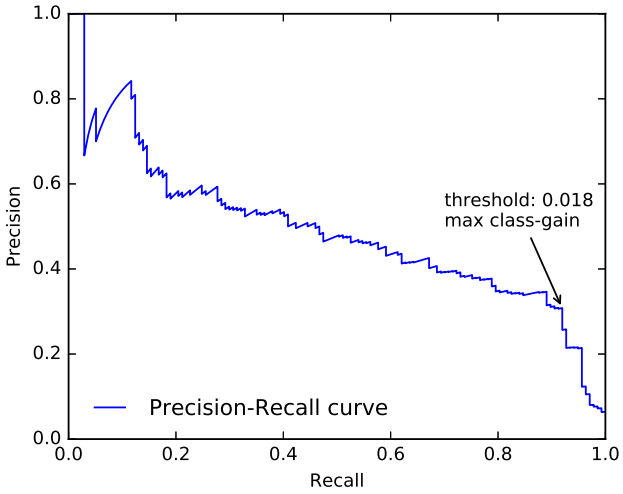


Fig. 6. Precision-recall curve.

improve our results. First of all, we framed the problem given in the competition as a regression task. Because the observations in the training data were given as time series, it was possible to retrieve the energy level for the next hours, see Fig. 4. This allowed us to forecast energy levels within the target window of 8 hours. Note that predictions from a regression model may be directly evaluated using AUC accuracy measure as it can be considered as a *risk scoring* model for high values of energy. For the original classification problem, we tried to modify the energy levels and train the models on an enriched set of labels, e.g., we assumed 30 kJ (and a couple of other values) as the warning level and estimate the model.

We also experimented with undersampling of training instances pertaining to class 0 so that the proportion of '1' in the training data increases. We also tried to reduce samples from locations 264, 373 and 437 in the training set by undersampling or assigning them a lower sample weight (in, e.g., LR model).

However, in this particular application, our efforts were not successful as the performance of the models (in terms of evaluation scores) was not improving.

F. Model performance on the final test set

After the competition we were provided with the true labels used during the final evaluation. We were able to compute different metrics than AUC. The winning model's predictions had a strongly skewed distribution (Fig. 5), corresponding to total energies seen in Fig. 4. The distribution has two modes, however of a much lower mode related to predicted warnings - this is due to imbalance of classes. Depending on the threshold beyond which we consider predictions as warnings we can derive the confusion matrix:

		True warning	
		1	0
Predicted warning	1	$TP = 126$	$FP = 284$
	0	$FN = 11$	$TN = 2390$

Based on the matrix we can compute several useful accuracy measures of the model:

$$\texttt{precision} = \texttt{TP} / (\texttt{TP} + \texttt{FP}) \tag{2}$$

$$\texttt{sensitivity (recall)} = \texttt{TP} / (\texttt{TP} + \texttt{FN}) \tag{3}$$

$$\texttt{specificity} = \texttt{TN} / (\texttt{TN} + \texttt{FP}) \tag{4}$$

$$\texttt{F1} = 2 \cdot \frac{\texttt{precision} \cdot \texttt{recall}}{\texttt{precision} + \texttt{recall}} \tag{5}$$

$$\texttt{Class-gain} = \texttt{specificity} + \texttt{recall} - 1 \tag{6}$$

The threshold maximizing the class-gain score is 0.018 (see Fig. 5) and yields the following accuracy on the final test set:

precision	recall	specificity	F1	class-gain
0.31	0.92	0.89	0.46	0.81

The entire precision-recall curve can be seen in Fig. 6.

In order to assess our results we looked for research addressing similar problems as the one considered here. In [4] we found an approach to solve the same problem, however the results are not directly comparable since they are based on different datasets. Our results prove to outperform results reported there for all presented classifiers (the highest class-gain reported is 0.75). Also, in the cited paper there was no inter-coal-mine validation, while the models described in our work were cross-validated on separate coal mines. Therefore, the models proposed are designed to generalize well and should be applicable also to working sites with no historical data available.

IV. Summary

Given that the dataset originates from working mine sites, with the entire measurement infrastructure already installed,

we hope that the approach presented in this paper could be implemented and serve as a valuable tool for alerting about dangerous seismic events early. This hopefully should result in preventing possible accidents which pose a threat to employees and generate losses from damaged coal mine infrastructure and machinery.

Even though the models presented here have outperformed the other models in the competition, we recommend they be ensembled with other high-scoring models, because properly combined efforts of multiple participants are expected to yield better results than individual solutions.

Lastly, we would like to thank the organizers for the opportunity to solve a real-life problem and the contestants for creating such a competitive environment.

REFERENCES

[1] Wyższy Urząd Górniczy, "Wypadkowość w górnictwie od 1 stycznia 2015 do 31 grudnia 2015," 2015, in Polish, last accessed 18 April 2016. [Online]. Available: http://www.wug.gov.pl/bhp/Statystyki_archiwalne_2015

[2] A. Zagorecki, *Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing: 15th International Conference, RSFDGrC 2015, Tianjin, China, November 20-23, 2015, Proceedings*. Cham: Springer International Publishing, 2015, ch. Prediction of Methane Outbreaks in Coal Mines from Multivariate Time Series Using Random Forest, pp. 494–500. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-25783-9_44

[3] A. Janusz, M. Sikora, Ł. Wróbel, and D. Ślęzak, "Predicting Dangerous Seismic Events: AAIA16 Data Mining Challenge," in *Proceedings of FedCSIS 2016*. IEEE, 2016, in print September 2016.

[4] M. Sikora, "Induction and pruning of classification rules for prediction of microseismic hazards in coal mines," *Expert Systems with Applications*, vol. 38, no. 6, pp. 6748–6758, 2011.

[5] Knowledge Pit, a host platform for data challenges, 2016, last accessed 18 April 2016. [Online]. Available: https://knowledgepit.fedcsis.org/

[6] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," *arXiv preprint arXiv:1603.02754*, 2016.

[7] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[8] L. Buitinck, G. Louppe, M. Blondel, F. Pedregosa, A. Mueller, O. Grisel, V. Niculae, P. Prettenhofer, A. Gramfort, J. Grobler, R. Layton, J. VanderPlas, A. Joly, B. Holt, and G. Varoquaux, "API design for machine learning software: experiences from the scikit-learn project," in *ECML PKDD Workshop: Languages for Data Mining and Machine Learning*, 2013, pp. 108–122.

[9] T. K. Ho, "The random subspace method for constructing decision forests," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 20, no. 8, pp. 832–844, 1998.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 221–224
DOI: 10.15439/2016F21
ACSIS, Vol. 8. ISSN 2300-5963

Predicting Dangerous Seismic Events in Coal Mines under Distribution Drift

Marc Boullé
Orange Labs,
2 avenue Pierre Marzin, 22300 Lannion, France
http://www.marc-boulle.fr/
Email: marc.boulle@orange.com

***Abstract*—We describe our submission to the AAIA'16 Data Mining Competition, where the objective is to devise a reliable prediction model for detecting periods of increased seismic activity in coal mines. Our solution exploits a selective naive Bayes classifier, with optimal preprocessing, variable selection and model averaging, together with an automatic variable construction method that builds many variables from time series records. One challenging part of the competition is that the input variables are not independent and identically distributed (i.i.d.) between the train and test datasets, since the train data and test data rely on different coal mines and different times periods. We apply a drift-aware methodology to alleviate this problem, that enabled to get a final score of 0.9246 (team marcb), less than 0.015 from the challenge winner.**

I. Introduction

The AAIA'16 Data Mining Competition[1] is related to a problem of prediction of dangerous seismic events in coal mines. Coal mines are equipped with seismic sensors that register bumps and energy. Sensor readings are available as times series of 24 records, with hourly statistic summaries (such as number of bumps, sum, mean or max of energy...). Train data consists of 79,893 samples from a first period, whereas the test data contains 3,860 samples coming from a second period. In the test data, the time periods do not overlap and are in random order, which is not the case in the train data. In this competition, active participants are rewarded by up to four additional train datasets, provided that they make enough submissions. Altogether, the most active participants can obtain up to a total of 133,151 train samples. The objective is to predict if the total seismic energy perceived with 8 hours after the period covered by a data sample exceeds a warning threshold, and the evaluation criterion is the Area Under the ROC Curve (AUC).

In this paper, we present our submission to the challenge. It exploits a Selective Naive Bayes classifier together with an automatic variable construction method (Section II). A good classifier trained on the train data obtains a disastrous leaderboard score. This is not caused by over-fitting, but by a severe distribution drift between train and test data. We suggest in Section III a methodology to alleviate this problem, and apply it in Section IV to elaborate our submissions to the challenge. Finally, Section V summarizes the paper.

[1]https://knowledgepit.fedcsis.org/contest/view.php?id=112

II. Supervised Classification Framework

We summarize the Selective Naive Bayes (SNB) classifier[2] introduced in [1]. It extends the Naive Bayes classifier owing to an optimal estimation of the class conditional probabilities, a Bayesian variable selection and a Compression-based Model Averaging. We also describe the automatic variable construction framework presented in [2], used to get a tabular representation from times series.

A. Optimal preprocessing

Numerical variables are preprocessed using supervised discretization [3] to evaluate the class conditional probabilities. In the MODL approach [4], the discretization is turned into a model selection problem and solved in a Bayesian way. Using a hierarchical prior distribution on the discretization parameters, the Bayes formula is applicable to derive an exact analytical criterion to evaluate the posterior probability of a discretization model. A 0-1 normalized version of this criterion provides a univariate informativeness evaluation of each input variable. Similarly, categorical variables are preprocessed using supervised value grouping [5].

B. Bayesian Approach for Variable Selection

The naive independence assumption can harm the performance when violated. In [1], the Selective Naive Bayes (SNB) classifier [6] is trained using a Bayesian model selection approach to select the best subset of variables [7]. Efficient search heuristics with super-linear computation time are proposed, on the basis of greedy forward addition and backward elimination of variables.

C. Compression-Based Model Averaging

Instead of taking the best subset of variables, the method introduced in [1] averages all the classifiers resulting from different subsets of variable, using a logarithmic smoothing of the posterior distribution of the trained classifiers. The weighting scheme on the models reduces to a weighting scheme on the variables, and finally results in a single Naive Bayes classifier with weights per variable.

[2]Available as a shareware at http://www.khiops.com

D. Automatic Variable Construction for Multi-Table

Variable construction [8] has been less studied than variable selection in the literature. It is all the more necessary in the case of relational data to obtain a flat input data table with tabular representation. It implies a large amount of work for the data analyst and heavily relies on domain knowledge to construct new potentially informative variables. Learning from relational data has recently received an increasing attention in the literature, since the introduction of Multi-Relational Data Mining (MRDM) in [9], [10]. In this paper, we exploit the automatic variable construction framework presented in [2]. It relies on a formal description of the data structure, with a root table and several secondary tables in 0 to 1 or 0 to n relationship and a set of construction rules (*Count*, *CountDistinct*, *Mode*, *Min*, *Max*, *Mean*, *Median*, *StdDev*, *Sum*, *Selection*). The space of variables that can be constructed is virtually infinite, which raises both combinatorial and over-fitting problems. These problems are solved by introducing a prior distribution over all the constructed variables, as well as an effective algorithm to draw samples of constructed variables from this distribution.

III. A Methodology to Reduce the Drift Problem

Statistical learning relies on identically and independently distributed (i.i.d.) data. Given this assumption, models trained from a train dataset can be deployed on a test dataset, with some guarantees of performance. This i.i.d. assumption does not hold in many real world cases, for example in case of time series data, in the marketing field where a model (churn, fraud, cross-selling...) is trained on a past period and deployed on a future period, ergonomics where a model is trained from a panel of few volunteers... In these cases of *drift* between the train and deployment datasets, as the data are not i.i.d, obtaining good classification performance on the train data does not guarantee good performance on the test data.

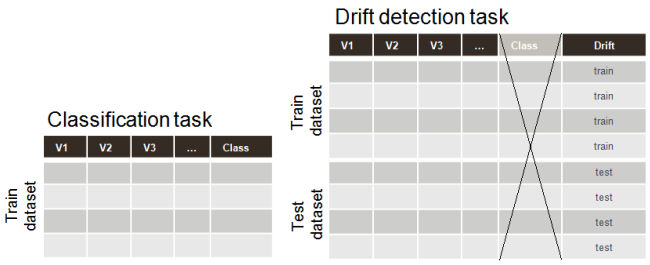


Fig. 1. Classification and drift detection tasks

In [11], [12], we have investigated this issue and proposed a methodology to reduce the drift problem. Let us assume that we have a classification task, a train dataset with class labels and a test dataset that potentially comes from a different distribution. The objective is to train a classifier and to predict the test class labels as accurately as possible whatever be the drift. Let us then consider two tasks: classification and detection of the drift. The drift detection task can be turned into a classification task as in [13], by merging the train and test datasets and using the dataset label ('train' or 'test') as the target variable (see Figure 1).

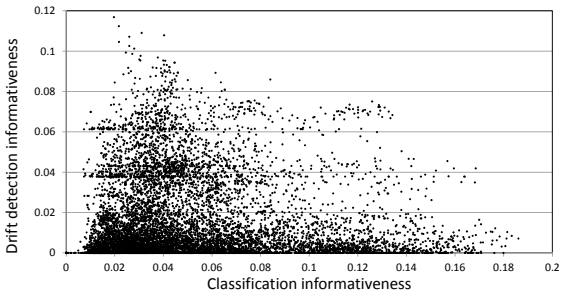


Fig. 2. Informativeness of 10,000 input variables

The initial representation is then evaluated using the pre-processing method summarized in Section II-A, both for the classification and the drift detection tasks. Intuitively, if we are able to select an input representation with good classification performance on the train data but poor drift detection, we expect that our classifier will be less sensitive to drift and its performance drop on the test dataset will be reduced.

The objective is then to explore varying input representations and select the one with the best classification performance together with the poorest drift detection. To illustrate this, we represent in Figure 2 the informativeness of 10,000 input variables for the classification and drift detection tasks (borrowed from the AAIA'2015 challenge [11]). The results show that there are variables with large drift informativeness and small classification informativeness (top-left of the figure), or on the contrary variables with small drift informativeness and large classification informativeness (bottom-right). The interesting variables are those on the right and close to the X axis, with small drift informativeness. Using these information, we can select interesting variables, either automatically (as in the AAIA'2015 challenge [11]) or manually with a focus on interpretability (as in the IJCRS'2015 challenge [12]). With interesting variables only, the classification performance may slightly decrease in the train dataset (because only part of the available variables are exploited), but the performance is likely to be more resilient to drift, with a better performance on the test dataset.

IV. Challenge Submissions

A. Applying the Framework for the Challenge

Coal mines are represented using a multi-table schema:

- root table that contains the identifier of the main working site (coal mine) and 12 other characteristics related to the whole period of 24 hours,
- secondary table (0-n) for the time series of 24 hourly summarized seismic sensor readings,
- secondary table (0-1) that contains some meta-data per working site.

Using the data structure presented in Figure 3 and the construction rules introduced in Section II-D, one can for

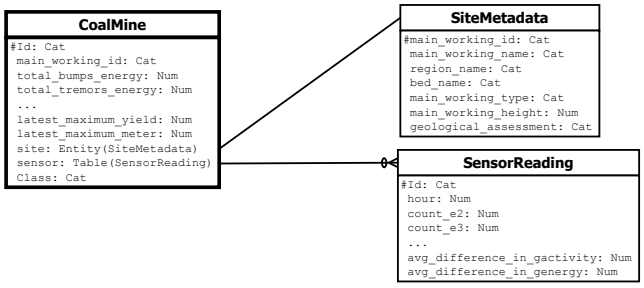


Fig. 3. Multi-table representation for the data of the AAIA'16 challenge

example construct the following variables ("name": comment) to enrich the description of a *CoalMine*:

- "Mean(sensor.sum_e2)": mean of the sensor sum_e2 readings,
- "Count(sensor) where sum_e2 $>$ 0.5": number of sensor readings where the sum_e2 value is greater than 0.5,
- "Max(sensor.sum_e2) where highest_bump_energy $>$ 50": max of the sum_e2 value from sensor readings where highest_bump_energy is greater than 50.

The number of variables to construct is the only user parameter. An input flat data table representation is then obtained from the set of all automatically constructed variables. All these variables are then preprocessed using the optimal discretization method (cf. Section II-A) to assess their informativeness and evaluate their class conditional probabilities, before training the SNB classifier.

For each experiment, 1000 variables are built using the automatic variable construction framework summarized in Section II-D.

B. Preliminary experiments

We first perform some explanatory analysis to better understand the data, without any submission on the leaderboard.

1) Evaluation of the expected performance: Using a 70%-30% train-test split of the train dataset, we obtain a train AUC of 0.99 and a test AUC of 0.97. The performance are both accurate and reliable. However, prediction of increased seismic activity should not be so easy, and we suspect that the good performance might be caused by some bias in the dataset.

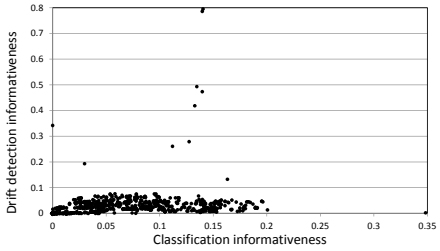


Fig. 4. Informativeness of 1000 input variables for the AAIA'2016 challenge

2) Evaluation of the drift: To evaluate whether the train and test dataset are i.i.d, we apply the drift detection methodology described in Section III. This drift detection task achieves an almost perfect performance with an AUC of 0.995, meaning that the train and test data can be well separated. The most informative variables for the drift detection task are the identifier of the main working site, as well as all the meta-data variables per working site. These drift informative variables are far above most other input variables in Figure 4. As the data are not i.i.d, obtaining good classification performance on the train data does not guarantee good performance on the test data.

3) Distribution of coal mines in train and test datasets: To further investigate on the observed drift, we collect the identifiers of the coal mines in the train and test datasets. Overall, 24 coal mines are used: 7 in the initial train dataset, 16 with all the additional train datasets and 21 in the test dataset. The distribution of the coal mines is heavily unbalanced in the train dataset, whereas is is more balanced in the test dataset.

4) Distribution of the target labels: The target class is heavily unbalanced, with 1171 *warning* (around 1.5%) and the rest as *normal*. Furthermore, 2 among the 7 initial train coal mines are never labeled as *warning*.

According to the challenge organizers, the time periods in the test data do not overlap and are in random order. We then assume that in the train data, the time periods overlap and are in sequential order. This overlapping causes an additional problem of non i.i.d data, with the train data being over-sampled compared to the test data.

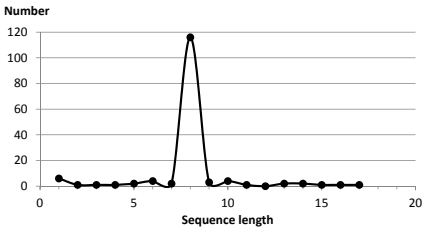


Fig. 5. Number of sequences of train warnings per sequence size

To evaluate the over-sampling factor we collect the sequences of consecutive *warning* in the train dataset. The results, displayed in Figure 5, show that most sequences are of length 8.

C. First submission

The preliminary explanatory analysis summarized in Section IV-B shows that there is a severe drift between the train and test datasets, caused by different mines, different time periods and different sampling rates. To reduce this drift problem, we apply the following protocol:

- remove any variable that identifies the coal mines (*main_working_id* plus the additional meta-data variables per working site),
- re-sample the train mines by keeping at most 5% of instance per mine, so as to get a more balanced distribution as in the test dataset (21 test mines, thus $1/21 \approx 5\%$),
- sub-sample the remaining train instances by a factor of $1/8 \approx 12\%$), so as to get approximatively the same sampling factor as in the test dataset.

This first trial, submitted one day after entering the challenge, obtains a leaderboard AUC of 0.9239, which is very competitive ($\approx 1\%$ from the leader at the submission time).

D. Second and final submission

Although the first obtained results are quite good, they exploit only a small subset of the train instances (around 3% after applying the sampling strategies). To better exploit all the available train data, we repeat the train protocol described previously 100 times (based on different random samples) and average the predictions. This second trial (our final one) obtains almost the same leaderboard AUC (0.9243), but we expect the averaging strategy may lead to more reliable predictions (the leaderboard AUC is evaluated on only 25% of the test data).

Furthermore, this averaging over 100 train samples provides additional insights w.r.t. the variance of the results, which amounts to around 1%. We expect that the variance of leaderboard AUC is still higher and that the best participant submissions (over potentially hundred of submissions) are likely to over-estimate the true test AUC. Thus getting a leaderboard AUC within the variance of the leader leaves few room for further improvement.

E. Additional experiments

First, as a sanity check, we submitted the first prediction obtained using all the available train data (all variables and instances: see Section IV-B). As expected, the drift effect is disastrous, and our 0.97 train AUC dropped down to a 0.60 leaderboard AUC.

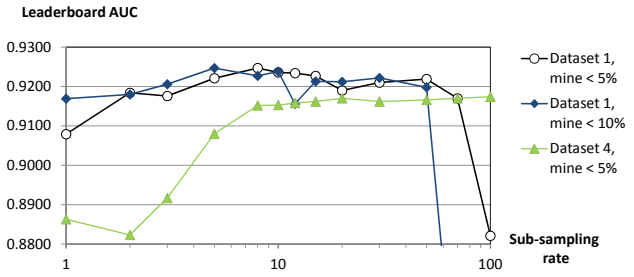


Fig. 6. Leaderboard AUC per sampling rate for three scenarios

We also performed sensitivity analysis, with varying the train mine re-sampling rate, sub-sampling rate, averaging strategy and number of constructed variables. We also experimented with using the additional train datasets, with some manual or automatic feature selection as well as finer anti-drift approaches (see [12]). For example, Figure 6 displays the leaderboard AUC using 1 or all 4 additional train datasets, with train mine resampling rate of at most 5% or 10%, and sub-sampling rate varying from 1% to 100%. This shows that the initial train mine re-sampling rate (5%) looks stable in a wider range (than 10%), that the initial sub-sampling rate (12%) is in a plateau of good performance (between 8% and 20% looks fine). Finally, using all the 4 additional train datasets, the performance was slightly less accurate, but as this is within the expected variance, this is not significant.

Overall, the obtained leaderboard results showed that the preliminary chosen protocol parameters (see Sections IV-B, IV-D) were quite stable, and the results dropped down only for significant changes in the re-sampling and sub-sampling rates (worse performance for at least half or twice the initial value of the parameters). The additional data did not provide any further improvement, but there was not much room for such improvement. In the end, we choose to keep our second submission, that went from a 0.9243 leaderboard AUC to a 0.9246 final AUC.

V. Conclusion

In the AAIA'16 Data Mining Competition, the train and test data are not i.i.d, which causes a dramatic drop of the test performance, even for accurate and reliable trained classifiers. After preliminary explanatory analysis, we identified several causes of drift between the train and test data: different distributions of coal mines, different sampling rate and different period. To be more robust to drift, we proposed a methodology based on removing variables too sensitive to drift, re-sampling to get a more balanced distribution of the train mines and sub-sampling to achieve approximately the same sampling rate. Applying this methodology, 100 classifiers were trained, each exploiting sub-samples of only 3% of the trained instances, and the averaged predictions obtained a 0.9246 final AUC.

References

[1] M. Boullé, "Compression-based averaging of selective naive Bayes classifiers," *Journal of Machine Learning Research*, vol. 8, pp. 1659–1685, 2007.

[2] ——, "Towards automatic feature construction for supervised classification," in *ECML/PKDD 2014*. Springer-Verlag, 2014, pp. 181–196.

[3] J. Dougherty, R. Kohavi, and M. Sahami, "Supervised and unsupervised discretization of continuous features," in *Proceedings of the 12th International Conference on Machine Learning*. Morgan Kaufmann, San Francisco, CA, 1995, pp. 194–202.

[4] M. Boullé, "MODL: a Bayes optimal discretization method for continuous attributes," *Machine Learning*, vol. 65, no. 1, pp. 131–165, 2006.

[5] ——, "A Bayes optimal approach for partitioning the values of categorical attributes," *Journal of Machine Learning Research*, vol. 6, pp. 1431–1452, 2005.

[6] P. Langley and S. Sage, "Induction of selective Bayesian classifiers," in *Proceedings of the 10th Conference on Uncertainty in Artificial Intelligence*. Morgan Kaufmann, 1994, pp. 399–406.

[7] I. Guyon, S. Gunn, M. Nikravesh, and L. Zadeh, *Feature Extraction: Foundations And Applications*. Springer, 2006.

[8] H. Liu and H. Motoda, *Feature Extraction, Construction and Selection: A Data Mining Perspective*. Kluwer Academic Publishers, 1998.

[9] A. J. Knobbe, H. Blockeel, A. Siebes, and D. Van Der Wallen, "Multi-Relational Data Mining," in *Proceedings of Benelearn '99*, 1999.

[10] S. Kramer, P. A. Flach, and N. Lavrač, "Propositionalization approaches to relational data mining," in *Relational data mining*, S. Džeroski and N. Lavrač, Eds. Springer-Verlag, 2001, ch. 11, pp. 262–286.

[11] M. Boullé, "Tagging fireworkers activities from body sensors under distribution drift," in *Federated Conference on Computer Science and Information Systems*, 2015. doi: 10.15439/2015F423 pp. 389–396.

[12] ——, "Prediction of methane outbreak in coal mines from historical sensor data under distribution drift," in *Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing - 15th International Conference, RSFDGrC*, 2015. doi: 10.1007/978-3-319-25783-9 pp. 439–451.

[13] A. Bondu and M. Boullé, "A supervised approach for change detection in data streams," in *Proceedings of International Joint Conference on Neural Networks*, 2011, pp. 519–526.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 225–229
DOI: 10.15439/2016F90
ACSIS, Vol. 8. ISSN 2300-5963

Massively Parallel Feature Extraction Framework Application in Predicting Dangerous Seismic Events

Marek Grzegorowski
Faculty of Mathematics, Informatics and Mechanics, University of Warsaw,
Banacha 2, 02-097 Warsaw, Poland, Email: M.Grzegorowski@mimuw.edu.pl

***Abstract*—In this paper we introduce an automated mechanism for knowledge discovery from data streams. As a part of this work, we also present a new approach to the creation of classifiers ensemble based on a wide variety of models. Furthermore, we describe an innovative, highly scalable feature extraction and selection framework designed to work with the MapReduce programming model and the application of designed framework to build an ensemble of classifiers which takes into account both the quality and the diversity of individual models. The effectiveness of the solution has been verified through a participation in an open data mining competition which concerned the problem of predicting periods of increased seismic activity causing life-threatening accidents in coal mines. The submitted solution obtained the highest AUC score of all the solutions uploaded by 106 participating research teams.**

I. Introduction

FOR SOME time, we can observe a massive shift in technology that makes sensors are more and more available and common. On the other hand, we can notice a significant grow in popularity of stream analytics as well as a decline in prices of data storage. One of the main beneficiaries of the aforementioned changes are monitoring, threats detecting and decision supporting systems. An exemplary application of which could be active monitoring of coal extraction [21] to provide protection for people underground or supporting fire commanders in decision making [14].

Nowadays, an increasing number of business technology objectives is related with comprehensive data analytics. Meanwhile, many researchers recognize feature engineering [8] as a major step in the process of knowledge discovery, necessary to obtain good results of the analysis. In this paper we propose an innovative approach to data analysis that, in order to provide high quality assessment, implies creation of an ensemble [3] of classifiers using a wide variety of models based on various subsets of attributes, in this way, resulting not only in enhancement of the quality of indications, but also minimizing the impact of concept drift on the final evaluation of results.

The continuous collection and analysis of multiple reading streams from a large network of sensors located underground raises a problem of long lasting and usually very complex preparation of data. Therefore, in order to simplify and speed up the feature extraction we introduced a novel framework designed to process streams of numerical readings from multiple sensors. The developed framework is ready to operate in production environments and, hence, is tailored for incremental processing of the emerging data based on the sliding window technique and the concept of parallelization presented in [5].

In the following sections we present the extension of our former solution [6], [7] with additional features and describe modification of the architecture allowing the work both with incremental data as well as with highly-scalable batch computations via MapReduce [2] programming model. The assessment of the solution was carried out on the basis of real life problems related to the streaming data [22].

The effectiveness of the framework has been confirmed in the analysis of several significantly different problems within the data analysis competitions. The first, concerned the recognition of the activity and posture of firefighter based on readings from multiple motion and vital sensors as a part of AAIA'15 Data Mining Competition: Tagging Firefighter Activities at the Fire Scene[17]. The second concerned the prediction of dangerous concentration of methane in the atmosphere of the mine sidewalks as a part of IJCRS'15 Data Challenge: Mining Data from Coal Mines[10]. The third concerned the prediction of increased seismic activity in mines as a part of AAIA'16 Data Mining Challenge: Predicting Dangerous Seismic Events in Active Coal Mines [9]. In all competitions, the results were very promising. It is worth noting that in the case of the seismic activity analysis, the solution based on the elaborated framework received the highest score in terms of Area Under ROC Curve (AUC) measure, equal to 93.96%, while in the methane concentration level analysis it achieved the second highest result with AUC equal to 94.73%.

The paper is organized as follows. In Section II we present the description of the data challenge problem. In Section III and IV we provide detailed information about the elaborated feature engineering framework, including insights of feature extraction and selection. Next, in Section V, we describe the conduct of the experiments and resulted ensemble model. Finally, in Section VI we summarize the work.

II. AAIA'16 Data Challenge Problem Description

Providing safety of miners working underground is the fundamental requirement for the coal mining industry in Poland. Coal mining companies are obligated by the law to introduce many safety measures to secure proper working conditions of their underground personnel. The task in the competition was to devise a reliable prediction model for detecting periods of increased seismic activity that could endanger miners.

More precisely, the tasks of the data challenge was to predict likelihood of the 'warning' label for the records from the test

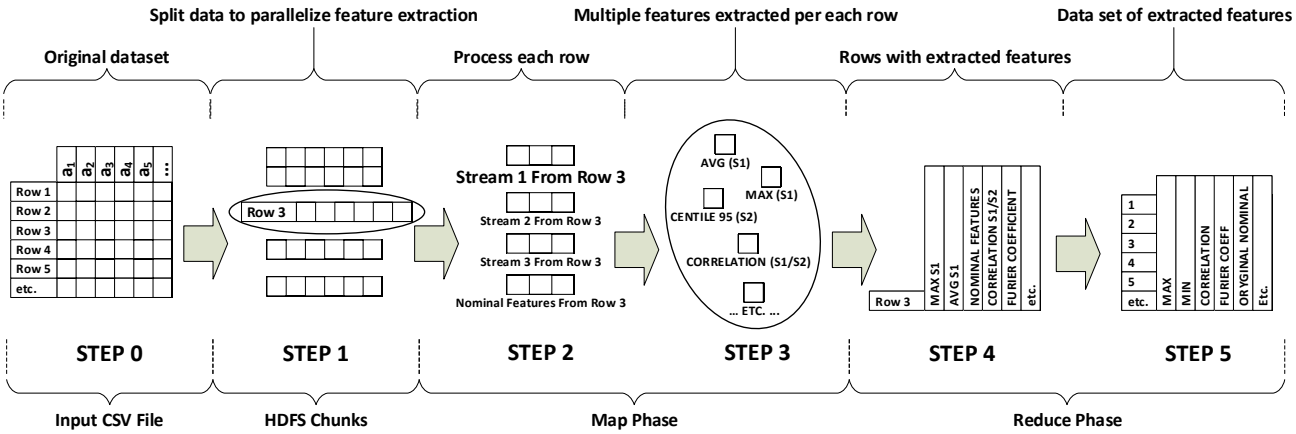


Figure 1. This diagram presents the high level overview of the feature extraction process broken down into individual steps. The curly braces at the top of the diagram indicate goals achieved in each processing step. The curly braces at the bottom of the diagram show how the individual processing steps were implemented. The 'original dataset' in STEP 0 corresponds to dataset provided by organizers of AAIA'16 Data Mining Challenge: Predicting Dangerous Seismic Events in Active Coal Mines. Features $a_1, a_2, a_3, ..$ correspond to attributes presented in Table I. STEP 1 was designed to partition of the original dataset into individual rows in order to parallelize calculations - this step was implemented by internal mechanisms of the MapReduce framework. STEP 2 was to split each row into nominal and stream data.In STEP 3 the feature extraction framework was applied to each data stream (e.g. time series containing 24 values corresponding to the average energy of the most active geophone - see Table I) and all features described in Tables II and III were created. In STEPs 4 and 5 all features, the nominal attributes as well as attributes newly extracted from streams, were put together.

set. The real number corresponding to the predicted likelihood was not expected to be in any particular range, however, higher numerical value should have indicated a higher chance of the 'warning' label. The final assessment of solutions were calculated using the Area Under the ROC Curve (AUC).

The data set, which was provided by Research and Development Centre EMAG, consisted of hourly aggregated readings from seismic sensors that count the number of seismic bumps perceived at longwalls and measure their total energy. Data records were composed of 24 consecutive hours of such readings coupled with the most recent assessments of the conditions at longwalls made by mining experts. All the attributes of the data set are described in Table I. The data sets were well prepared, cleaned of malformed and erroneous values, without missing attributes.

Table I
THE TABLE PRESENTS ATTRIBUTES OF THE MAIN DATA FILES.

no.	feature description
1	id of the main working site where the measurements were taken
2	total energy of: seismic bumps, major seismic bumps, destressing blasts and all types of bumps registered in the last 24h
3	latest progress in the mining from, both, left and right side
4	latest seismic, comprehensive and seismoacoustic (standard and alternative method) hazard assessments made by experts (a/b/c/d): a - no hazard; b - moderate hazard; c - high hazard; d - dangerous
5	maximum yield from the last meter of the small-diameter drilling
6	depth at which the maximum yield was registered
7	five time series containing 24 values (one per hour 1..24) each corresponding to a number of seismic bumps with energy in the following ranges: $(0, 10^2]$, $(10^2, 10^3]$, $(10^3, 10^4]$, $(10^4, 10^5]$ and $(10^5, Inf)$ aggregated per hour (1..24)
8	five time series containing 24 values (one per hour 1..24) each corresponding to sum of energy of registered seismic bumps with energy in the following ranges: $(0, 10^2]$, $(10^2, 10^3]$, $(10^3, 10^4]$, $(10^4, 10^5]$ and $(10^5, Inf)$ aggregated per hour (1..24)
9	four time series, each containing 24 values (one per hour 1..24) corresponding to the number of: seismic bumps, rock bursts, destressing blasts and to energy of the strongest seismic bump
10	four time series, each containing 24 values (one per hour 1..24) corresponding to maximum activity, maximum energy, average activity and average energy of the most active geophone
11	four time series, each containing 24 values (one per hour 1..24) ccorresponding to the maximum difference and average difference in, both, activity and energy registered by the most active geophone

In total, the training file contained 133150 records provided in a tabular format with 541 columns. The label indicates whether a total seismic energy perceived during 8 hours after the period covered by a data record exceeded the warning threshold of $5 * 10^4$ Joules. There were 2963 examples labeled as 'warning' and 130187 'normal' cases in the training set.

III. FEATURE EXTRACTION

In the course of feature extraction we generated a large number [26] of potentially relevant characteristics [16], [23] and applied the feature selection in the next step. The feature extraction was based on the sliding window [24] method and was configured to accept on its input a data set containing readings from multiple streams [5]. According to the submitted configuration each stream, stored in a row of the csv file (training and test set), was divided into three non-overlapping frames. During the process of moving a sliding window through the time series a number of aggregating functions were applied. The Table II presents features extracted form a single time series.

The statistics indicated in Table II were supplemented by Kendall's correlation between each pair of data streams for every row in csv. Furthermore, because there were more than one window generated for each time series we extracted inter-window statistics, that is, a set of values that express

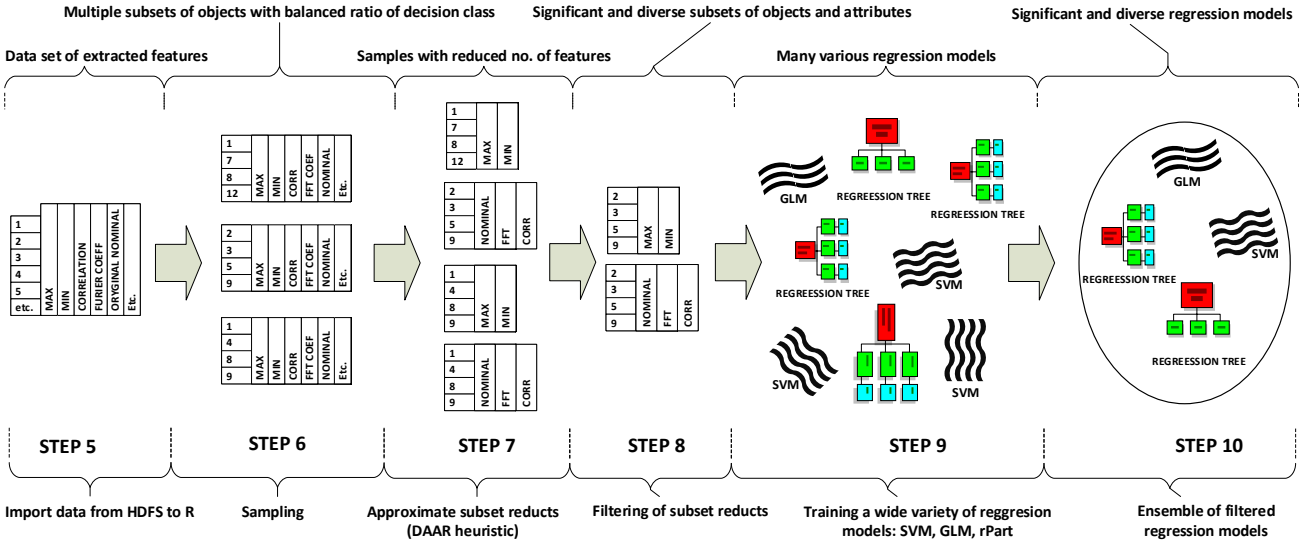


Figure 2. This diagram presents the course of feature selection, regression models training and construction of the final ensemble of regression models broken down into individual steps. The curly braces at the top of the diagram indicate goals achieved in processing steps. The curly braces at the bottom of the diagram show how each processing step was implemented. All the processing steps from 6 to 10 were implemented in R environment for statistical computing. Consecutive processing steps were designed to: STEP 6 - was to drew a number of random samples of objects with balanced decision classes. In STEP 7 the reduced attribute subsets were calculated - this step was implemented basing on the concept of approximate decision reducts derived from the theory of rough sets. In STEP 8, a number of attribute subsets were obtained. Each subset was derived by merging 2 or 3 reducts, however, only significantly different subsets were maintained for the purpose of model training in the following phase of processing. In STEP 9 previously obtained subsets of objects and attributes were used to train regression models based on selected algorithms: rPart, SVM, GLM. Lastly, in STEP 10 the most important models were used to form an ensemble. In the process of models selection for the purpose of ensemble construction we took into consideration, both the quality of the reggresion model as well as the degree of diversity in relation to already selected models. The course of steps 9 and 10 has been described as Algorithm 1.

the changes between pairs of same statistics in consecutive sliding windows. The inter-window stats are presented in Table III.

In order to optimize time-effectiveness of the experimentation process the framework implementation was modified so that it could be run in several modes, including: incremental stream processing[5], single-threaded mode and the map-reduce mode. In order to allow multi-mode framework operation, the feature extraction mechanism was isolated as a separate tool with respect to the runtime environment.

The solution for the competition was calculated in batch mode and launched on a cluster of Docker[2] containers with installed Hadoop[3] software library as an implementation of MapReduce protocol. The scheme of the whole feature extrction process is depicted in Figure 1. In the "Map" phase (compare steps 2 and 3 in Figure 1) every data row was divided into: sub-streams of numerical readings from various sensors, set of nominal and aggregated features and, in case of the training set, also label. Labels, nominal and aggregated attributes were transferred to the "Reduce" phase unchanged while the numerical streams were subjected to an additional feature extraction described above. In the "Reduce" phase

Table II
THE TABLE PRESENTS FEATURES WHICH ARE CALCULATED TO REPRESENT THE TIME SERIES IN A SLIDING WINDOW.

feature	description
max	a maximum value of the readings in the window
min	a minimum value of the readings in the window
maxMinDiff	a difference between the max and min
mean	a mean value of readings in the window
percentileX	a Xth percentiles for the readings, where: $X \in \{2,5,10,15,20,25,30,50,70,75,80,85,90,95,98\}$
percentiles5Diff	a subtraction of the percentiles 95% and 5%
stdDev	a standard deviation of the readings
variance	a variance deviation of the readings
fftCoeffSet	a set containing first 5 Fourier transform coefficients
kurtosis	a Kurtosis measure [1]
skewness	a measure of the asymmetry

Table III
INTER-WINDOW STATISTICS THAT EXPRESS THE CHANGES BETWEEN A PAIR OF EQUIVALENT FEATURES IN CONSECUTIVE WINDOWS.

feature	description
maxDiff	a difference between *max* stats in the consecutive sliding windows
meanDiff	a difference between *mean* stats in the consecutive sliding windows
minDiff	a difference between *min* statistics in the consecutive sliding windows
percentileXDiff	a difference between *Xth* percentile statistics in the consecutive sliding windows, where: $X \in \{5,25,50,75,95\}$

[2]See https://www.docker.com
[3]See http://hadoop.apache.org

(steps: 4 and 5 in Figure 1) all the attributes obtained for each row were combined back together.

IV. Feature Selection

The experimentation was significantly affected by the fact that the training data set was very unbalanced, rows labeled as 'warning' represented only 2,3% of all objects. Therefore, in the first step we drew a number of random samples that contained between 10 000 and 20 000 objects out of 133 150 all objects of the training set. Created samples differed in the number of objects of "warning" class (minority class), each contained a minimum of 1000 and a maximum of 2000 objects of this class. Objects within a particular sample were unique, however they could be repeated between different samples.

The generated data samples were randomly divided into two disjoint groups: group A - containing object subsets for the purpose of the feature selection, group B containing samples for the purpose of training of regression models. It should be noted that in order to use the chosen feature selection algorithms, before generating the 'A' samples the training set was subjected to discretization of a numerical attributes using local (univariate) version of the algorithm described in [18].

The selection of attributes [12] was carried out on the basis of a filter method derived from the theory of rough set[19]. Using the R[4] language and environment for statistical computing with installed RoughSets [20] package we calculated approximate decision reducts[13] with DAAR [11] heuristic. Approximate decision reducts are relatively small, thus, we decided to merge few as a single attribute subset. As a result of feature selection process (depicted in Figure 2) we prepared a number of significantly different attribute subsets.

V. Models Training and Ensambling

The task of maximizing AUC measure, posed by the organizers of the competition, prompted us to apply regression algorithms to identify the probability of the 'warning' label. The final solution is based on the concept of building an ensemble of diverse regression models which interpret 'warning' label as 1, while 'normal' label as 0. To provide a diversity of models we trained them on various subsets of attributes, what was important in order to provide a variety of regression models because the analyzed data set had few very dominant attributes. An additional effect of using different features for different models was to protect the ensemble against significant concept drift [1] on the part of the attributes between the training and test sets. This approach was also expected to protect the model against over-fitting [15] and, hence, against the significant decrease in the quality of prediction on the test set.

The machine learning were conducted on pre-prepared samples of objects with reduced number of attributes. The ultimate ensemble consisted of 8 significantly different regression models which were calculated with three various algorithms, including: regression trees (calculated by the algorithm from the rPart[5] package), SVM regressor (computed using the algorithm from the e1071[6] package) and the glm[7] function from R language to fit a generalized linear model. The following list presents models included in the final ensemble:

- Five simple regression tree models calculated with rPart.
- Two SVM models with different kernel functions
  - SVM_1 - regression, kernel: linear, cost: 1, gamma: 0.1, eps: 0.1 , Number of Support Vectors: 2968
  - SVM_2 - regression, kernel: radial, cost: 1, gamma: 0.07143, eps: 0.1, Number of Support Vectors: 7171
- One generalized linear model

Algorithm 1: The construction of regressors ensemble.

Data:
- *attSubsets* - pre-calculated subsets of attributes - approximate reducts
- *objectSamples* - pre-calculated object samples
- *testSet* - test set
- *regressionAlgorithms*, default: { rPart, SVM, glm }
- *allowedAttempts*, default: 3
- *minimalQualityTreshold* - minimal quality treshold

Result:
- *ensemble of regression models*

```
/* Initialization of variables                    */
ensemble ← ∅;  weakAttempts ← 0
alg ← regressionAlgorithms.removeFirst
while TRUE do
    a1, a2 ← attSubsets.drawAndRemoveTwo
    b1, b2 ← objectSamples.drawAndRemoveTwo
    /* Every model is trained and
       validated on various samples               */
    model ← alg.trainAndEvaluate(a1, b1, a2, b2)
    score ← model.score(testSet)
    /* The ensemble is expanded if the
       model meets the quality threshold
       and there is no similar model              */
    if model.evaluation > minimalQualityTreshold ∧
     ¬ensemble.containsSimilar(model, score) then
        ensamble ← ensamble ⋃ {model ⊕ score}
    else
        weakAttempts ← weakAttempts + 1
        if weakAttempts < allowedAttempts then
            continue;
        if regressionAlgorithms ≠ ∅ then
            alg ← regressionAlgorithms.removeFirst
              weakAttempts ← 0
        else end of experimentation
            break;
return ∑_{s∈ensemble.scores} s;
```

[4]See https://www.r-project.org

[5]See https://cran.r-project.org/web/packages/rpart

[6]See https://cran.r-project.org/web/packages/e1071

[7]See https://stat.ethz.ch/R-manual/R-patched/library/stats/html/glm.html

The whole phase of machine learning were carried out in the R statistical environment. The ensemble was successively extended with new models based on one of three designated algorithms, starting from the the rPart algorithm which in the initial assessment achieved the most promising results. In each step of the experiment: a sample of objects from those available in group 'B' and a subset of the attributes form those obtained in the phase of feature selection were drawn. The prepared subset of the training set was used to train a single regression model which was added to the ensemble under two conditions. First, the evaluation of the results had to exceed the satisfactory quality threshold. Second, the results of the regression for the test set had to be significantly different from any of the models already added to the ensemble. A detailed description of the experiment is presented in the Algorithm 1.

VI. Summary

The paper introduces an automated framework of extraction and selection of attributes designed to work with big data using MapReduce programming model. The article presents the proof of concept application of the framework to build an ensemble of classifiers based on a simple heuristic indicating the extension of the ensemble which takes into account both the quality and the diversity of the ultimate solution. The effectiveness of the developed solution has been verified by the participation in an open knowledge discovery competition in which it obtained the highest score in terms of AUC (93.96%) of all solutions submitted by 106 participating research teams.

VII. Acknowledgements

This research was partially supported by Polish National Centre for Research and Development (NCBiR) grant PBS2/B9/20/2013 in frame of Applied Research Programme.

References

[1] M. Boullé. Tagging fireworkers activities from body sensors under distribution drift. In Ganzha et al. [4], pages 389–396.

[2] J. Dean and S. Ghemawat. Mapreduce: simplified data processing on large clusters. *Commun. ACM*, 51(1):107–113, 2008.

[3] T. G. Dietterich. Ensemble methods in machine learning. In *Proceedings of the First International Workshop on Multiple Classifier Systems*, MCS '00, pages 1–15, London, UK, UK, 2000. Springer-Verlag.

[4] M. Ganzha, L. A. Maciaszek, and M. Paprzycki, editors. *2015 Federated Conference on Computer Science and Information Systems, FedCSIS 2015, Lódz, Poland, September 13-16, 2015*. IEEE, 2015.

[5] M. Grzegorowski. Scaling of complex calculations over big data-sets. In D. Ślęzak, G. Schaefer, S. T. Vuong, and Y. Kim, editors, *Active Media Technology - 10th International Conference, AMT 2014, Warsaw, Poland, August 11-14, 2014. Proceedings*, volume 8610 of *Lecture Notes in Computer Science*, pages 73–84. Springer, 2014.

[6] M. Grzegorowski and S. Stawicki. Window-Based Feature Engineering for Prediction of Methane Threats in Coal Mines. In Yao et al. [25], pages 452–463.

[7] M. Grzegorowski and S. Stawicki. Window-Based Feature Extraction Framework for Multi-Sensor Data: A Posture Recognition Case Study. In Ganzha et al. [4], pages 397–405.

[8] I. Guyon and A. Elisseeff. An introduction to variable and feature selection. *J. Mach. Learn. Res.*, 3:1157–1182, Mar. 2003.

[9] A. Janusz, M. Sikora, Ł. Wróbel, and D. Ślęzak. Predicting Dangerous Seismic Events: AAIA16 Data Mining Challenge. In M. Ganzha, L. A. Maciaszek, and M. Paprzycki, editors, *Proceedings of FedCSIS 2016*. IEEE, 2016. In print September 2016.

[10] A. Janusz, M. Sikora, L. Wróbel, S. Stawicki, M. Grzegorowski, P. Wojtas, and D. Ślęzak. Mining Data from Coal Mines: IJCRS'15 Data Challenge. In Yao et al. [25], pages 429–438.

[11] A. Janusz and D. Ślęzak. Random probes in computation and assessment of approximate reducts. In M. Kryszkiewicz, C. Cornelis, D. Ciucci, J. Medina-Moreno, H. Motoda, and Z. W. Ras, editors, *Rough Sets and Intelligent Systems Paradigms - Second International Conference, RSEISP 2014, Held as Part of JRS 2014, Granada and Madrid, Spain, July 9-13, 2014. Proceedings*, volume 8537 of *Lecture Notes in Computer Science*, pages 53–64. Springer, 2014.

[12] A. Janusz and D. Ślęzak. Rough set methods for attribute clustering and selection. *Applied Artificial Intelligence*, 28(3):220–242, 2014.

[13] A. Janusz and D. Ślęzak. Computation of approximate reducts with dynamically adjusted approximation threshold. In F. Esposito, O. Pivert, M. Hacid, Z. W. Ras, and S. Ferilli, editors, *Foundations of Intelligent Systems - 22nd International Symposium, ISMIS 2015, Lyon, France, October 21-23, 2015, Proceedings*, volume 9384 of *Lecture Notes in Computer Science*, pages 19–28. Springer, 2015.

[14] A. Krasuski, A. Jankowski, A. Skowron, and D. Ślęzak. From sensory data to decision making: A perspective on supporting a fire commander. In *2013 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology, Atlanta, Georgia, USA, 17-20 November 2013, Workshop Proceedings*, pages 229–236. IEEE Computer Society, 2013.

[15] P. Lameski, E. Zdravevski, R. Mingov, and A. Kulakov. SVM parameter tuning with grid search and its impact on reduction of model over-fitting. In Yao et al. [25], pages 464–474.

[16] J. Lasek and M. Gagolewski. The winning solution to the AAIA'15 data mining competition: Tagging firefighter activities at a fire scene. In Ganzha et al. [4], pages 375–380.

[17] M. Meina, A. Janusz, K. Rykaczewski, D. Ślęzak, B. Celmer, and A. Krasuski. Tagging firefighter activities at the emergency scene: Summary of AAIA'15 data mining competition at knowledge pit. In Ganzha et al. [4], pages 367–373.

[18] H. S. Nguyen. On efficient handling of continuous attributes in large data bases. *Fundam. Inform.*, 48(1):61–81, 2001.

[19] Z. Pawlak. Rough sets. *International Journal of Parallel Programming*, 11(5):341–356, 1982.

[20] L. S. Riza, A. Janusz, C. Bergmeir, C. Cornelis, F. Herrera, D. Ślęzak, and J. M. Benítez. Implementing algorithms of rough set theory and fuzzy rough set theory in the R package "roughsets". *Inf. Sci.*, 287:68–89, 2014.

[21] M. Sikora and B. Sikora. Improving prediction models applied in systems monitoring natural hazards and machinery. *International Journal of Applied Mathematics and Computer Science*, 22(2):477–491, 2012.

[22] J. Stefanowski, A. Cuzzocrea, and D. Slezak. Processing and mining complex data streams. *Inf. Sci.*, 285:63–65, 2014.

[23] S. Wawrzyniak and W. Niemiro. Clustering approach to the problem of human activity recognition using motion data. In Ganzha et al. [4], pages 411–416.

[24] A. Wieczorkowska, J. Wroblewski, P. Synak, and D. Ślęzak. Application of temporal descriptors to musical instrument sound recognition. *J. Intell. Inf. Syst.*, 21(1):71–93, 2003.

[25] Y. Yao, Q. Hu, H. Yu, and J. W. Grzymala-Busse, editors. *Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing - 15th International Conference, RSFDGrC 2015, Tianjin, China, November 20-23, 2015, Proceedings*, volume 9437 of *Lecture Notes in Computer Science*. Springer, 2015.

[26] A. Zagorecki. A versatile approach to classification of multivariate time series data. In Ganzha et al. [4], pages 407–410.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 231–234
DOI: 10.15439/2016F116
ACSIS, Vol. 8. ISSN 2300-5963

Fisher's Linear Discriminant Analysis Based Prediction using Transient Features of Seismic Events in Coal Mines

Başak Esin Köktürk Güzel, Bilge Karaçalı
Department of Electrical and Electronics Engineering
İzmir Institute of Technology
İzmir, Turkey
Email:{basakkokturk,bilge}@iyte.edu.tr

***Abstract*—Identification of seismic activity levels in coal mines is important to avoid accidents such as rockburst. Creating an early warning system that can save lives requires an automated way of predicting. This study proposes a prediction algorithm for the AAIA′16 Data Mining Challenge: Predicting Dangerous Seismic Events in Active Coal Mines that is based on transient activity features along with average indicators evaluated by a Fisher's linear discriminant analysis. Performance evaluation experiments on the training datasets revealed an accuracy level of around 0.9438 while the performance on the test dataset was at a level of 0.9297. These results suggest that the proposed approach achieves high accuracy in predicting danger seismic events while maintaining low complexity.**

I. Introduction

ONE OF the most important subjects in coal mining is to detect specific gas emissions and seismic activity rates. The miners can suddenly find themselves in dangerous situations due to methane explosions or rockburst [1]. The most common accident cause in coal mines are cave-ins (roof, rock and coal) which account for 70.6 of all injuries [2]. Safety of miners' lives substantially depends on an early warning mechanism that can potentially be constructed using specific alert measurements for seismic activities as well as physical conditions of the mine. In order to create such an early warning mechanism, warning signals can be triggered if the measured values or energy levels, taken measurement from reference points of the mine, exceed a preset hazard threshold. However, seismic activity datasets that are observed from coal mines are very high dimensional and hard to process due to the fact that they are measured from a wide range of points and for a long duration. Since the dataset is very complex and high dimensional, expert knowledge-based systems can fail for foresight of the dangerous activities. The automation of early warning systems in coal mines has vital importance to prevent interpretation differences between mining experts and make analysis more rapid.

In the literature, there are several automated methods that have been proposed to recognize hazardous seismic activity patterns. Neural networks are the most popular method for prediction of seismic events in coal mines [3]. Identification of neural network parameters and layer numbers, however, is complicated, and entails substantial cost because of its "black-box" structure [4]. As a simpler and practical alternative, we propose a hazardous seismic ebent activity prediction method based on Fisher′s Linear Discriminant Analysis [5] that operates on an encoding of transient seismic activity are on 24 hour period along with average seismic activity parameters and conventional risk assessment methods. Performance evaluation of the method on the AAIA′16 Data Mining Challenge Dataset suggest that the approach offers accurate prediction of hazardous seismic activity around %92.97 levels.

In the next section, we provide a detailed description of the dataset and explain the proposed approach. In the third section, we present performance evaluation results of our method on both initial training dataset as well as the additional training datasets. At the conclusion section, we summarize our algorithm and discuss the results.

II. Material and Methods

The dataset used in this study was provided by Research and Development Centre EMAG for AAIA′16 Data Mining Challenge: Predicting Dangerous Seismic Events in Coal Mines. The dataset consists of total energy measurements for 24 hour period from different sensors and the counts for seismic bumps perceived at longwalls. In addition, the dataset contains hourly readings for 24 consecutive hours that are related with the most recent assessments of the conditions determined by mining experts. In the dataset each sample has one ID of the main working site where the measurements were taken and 540 features which contains 12 average risk parameters and 528 risk assessments measures. Finally, the respective labels ("normal" or "warning") are provided for each individual sample to assist on the training.

We propose a method that evaluates the average risk parameters along with the risk assessment measures separately from the hourly measurements of the provided parameters over a 24-hour period for the prediction task at hand. To this end, we extracted the hourly measurements of the 22 different parameters provided in the training dataset

978-83-60810-90-3/$25.00

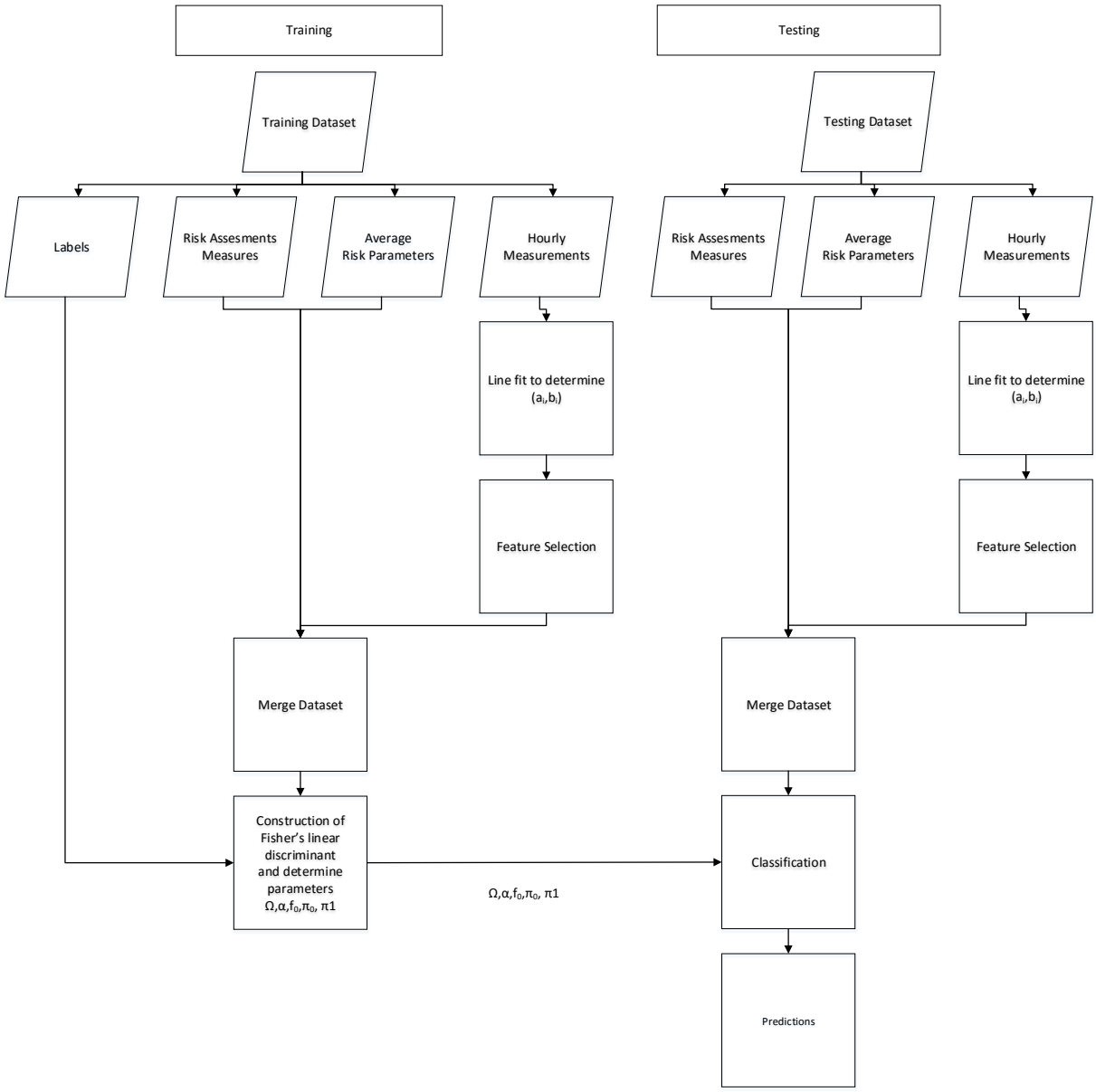


Fig. 1. Block Diagram of the Proposed Method

in indices from 14 through 541 and performed a line fit to determine the parameters (a_i, b_i) such that the fit error

$$\frac{1}{2}\sum_{h=1}^{24}(p_{i,h}-(a_i h+b_i))^2 \tag{1}$$

is minimal for each parameter p_i with hourly measurements $p_{i,h}$, for $i = 1,2,\ldots,22$. This resulted in a time evolution dataset with 44 features. To further refine this dataset, we have calculated the Kolmogorov-Smirnov statistic [6] between the empirical cumulative probability distributions of the two groups over each of the 44 time evolution parameters (a_i, b_i) for $i = 1,2,\ldots,22$, and ranked the parameters in the order of decreasing statistic value, with the understanding that the larger values of the statistic indicate more pronounced separation between the groups. Next, we have carried out Fisher′s linear discriminant analysis [7] on the top ranked $1,2,\ldots,44$ time evolution parameters, and calculated the area under the resulting receiver operating characteristics curve obtained on the original training dataset and the associated labels. This analysis identified 39 time evolution parameters with the greatest Kolmogorov-Smirnov statistic providing the highest area under the curve on the original training dataset, that were then collected to form the

calculated time evolution dataset containing the transient features.

The final prediction was obtained by merging the average risk parameters and the risk assessment measures provided in indices 2 through 13 in the training dataset with the calculated time evolution dataset, and constructing a Fisher's linear discriminant over the merged dataset. This entailed calculating the average vectors and covariance matrices

$$\mu_0 = \frac{1}{\ell_0} \sum_{j \in J_0} x_j \tag{2}$$

$$\mu_1 = \frac{1}{\ell_1} \sum_{j \in J_1} x_j \tag{3}$$

$$\Sigma_0 = \frac{1}{\ell_0 - 1} \sum_{j \in J_0} (x_j - \mu_0)(x_j - \mu_0)^T \tag{4}$$

$$\Sigma_1 = \frac{1}{\ell_1 - 1} \sum_{j \in J_1} (x_j - \mu_1)(x_j - \mu_1)^T \tag{5}$$

over the parameter vectors of the merged dataset $\{x_j\}$ with respect to the index sets J_0 and J_1 defined by

$$J_0 = \{j | y_j = 0\} \tag{6}$$

and

$$J_1 = \{j | y_j = 1\} \tag{7}$$

in terms of the training labels $\{y_j\}$ with $\ell_0 = |J_0|$ and $\ell_1 = |J_1|$. This allowed expressing the discriminant function $f(x)$ for a new parameter vector x through

$$f(x) = w^T x \tag{8}$$

with

$$w = (\Sigma_0 + \Sigma_1)^{-1}(\mu_1 - \mu_0). \tag{9}$$

As the final step of the analysis, we have identified the parameters α and f_0 to convert the values of the discriminant function into an empirical log-likelihood ratio for the two groups via the expression

$$L(x) = \alpha(f(x) - f_0) \tag{10}$$

so that the collection of values $\{L(x_j)\}$ over the merged training dataset $\{x_j, y_j\}$ achieved the smallest average training errors on the two groups with respect to a threshold of 0, or the average of the Type I and Type II training error rates, and the corresponding empirical posterior probabilities given by

$$P_1(x_j) = \frac{\pi_1}{\pi_0 e^{-L(x_j)} + \pi_1} \tag{11}$$

satisfied

$$\frac{1}{\ell_0 + \ell_1} \sum_{j=1}^{\ell_0 + \ell_1} P_1(x_j) = \pi_1 \tag{12}$$

with π_0 and π_1 denoting the prior probabilities of the respective groups in the training dataset. This was carried out by finding the f_0 value that achieved the equality above for a specific value of α for $\alpha = 2^{-5}, 2^{-4.5}, 2^{-4}, \ldots, 2^5$. Calculating the errors over the resulting (α, f_0) pairs identified the best values for α and f_0 for the smallest training error while maintaining the required prior probabilities.

In the next section, we present the results that we have obtained on different training and test dataset combinations.

III. Results

We have been tested our proposed method using the five different dataset and respective warning level labels that were provided by AAIA′16 Challenge committee. These datasets were named as: `training dataset`, `additional training dataset 1`,`additional training dataset 2`, `additional training dataset 3` and `additional training dataset 4`. Different combinations of these datasets were used to train the algorithm and the others were used for testing its performance.

Firstly, we used the original training dataset and we estimated the posterior probabilities for both additional training datasets as well as the original training dataset. The receiver operating characteristic (ROC) curves for this trial is shown in Figure 2.

The greatest AUC was obtained on the additional train dataset 1 at 0.9619 followed by at 0.9422 and at 0.9345 and at 0.9088 and at 8943. Next, we merged each additional

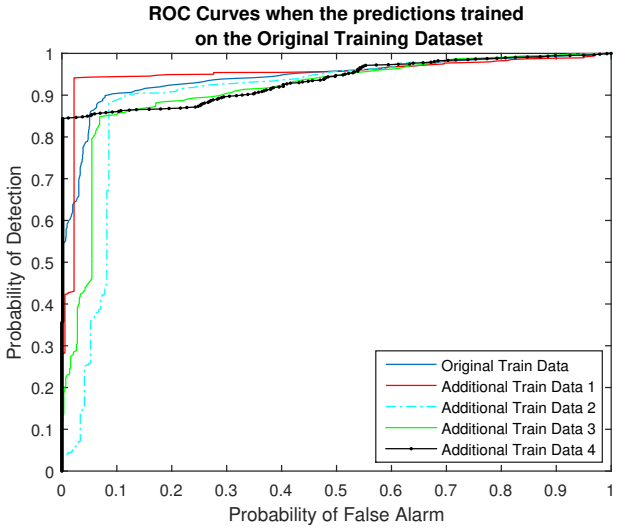


Fig. 2. The ROC curves when the algorithm trained by original training dataset. Best prediction is performed on additional training dataset 1

training dataset with the original training dataset and used the combined data to train the algorithm, and tested the resulting prediction on all datasets. The average area under curve (AUC) values are shown in Table I.

The area under the receiver operating characteristics curve obtained on the test dataset was 0.9297 as reported by the evaluation committee of the AAIA'16 Challenge when the training data was combination of the original training data and additional training data 1. The ROC curves are shown in Figure 3 for each dataset.

TABLE I
AVERAGE AREA UNDER CURVE VALUES FOR DIFFERENT TRAINING SET COMBINATIONS

Train Data	Average AUC Value
Original Train Data	0.9264±0.0240
Original Train Data , Additional Train Data 1	0.9311±0.0292
Original Train Data , Additional Train Data 2	0.9280±0.0253
Original Train Data , Additional Train Data 3	0.9298±0.0240
Original Train Data , Additional Train Data 4	0.9290±0.0224
Original Train Data , Additional Train Data 1, Additional Train Data 2	0.9320±0.0283
Original Train Data , Additional Train Data 1, Additional Train Data 3	0.9334±0.0261
Original Train Data , Additional Train Data 1, Additional Train Data 4	0.9337±0.0275
Original Train Data , Additional Train Data 2, Additional Train Data 3	0.9336±0.0250
Original Train Data , Additional Train Data 2, Additional Train Data 4	0.9303±0.0228
Original Train Data , Additional Train Data 3, Additional Train Data 4	0.9327±0.0220

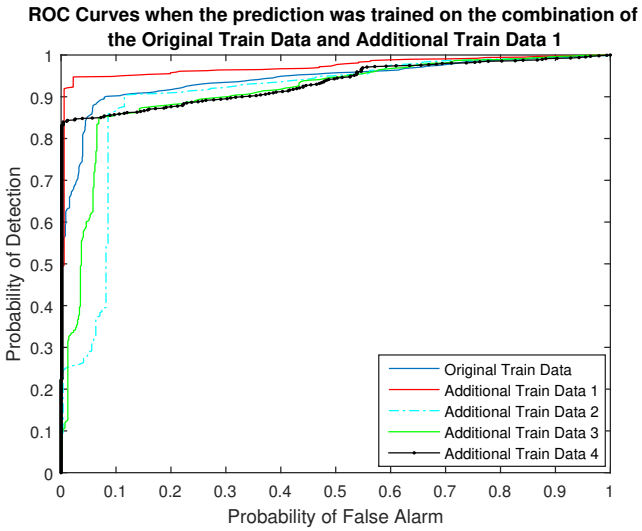


Fig. 3. The ROC curves when the algorithm was trained on the merge of the original training dataset with the additional training dataset 1. Best prediction performance is achieved on the additional training dataset 1

IV. CONCLUSION

In this paper, we have proposed a prediction algorithm for dangerous seismic events in coal mines using a combination of existing risk assessment parameters, average seismic energy measurements as well as hourly seismic activity measurements and Fisher's linear discriminant analysis. The method fits a line to capture the seismic activity information provided by hourly measurements and uses the two line-fit parameters as features for the ensuing prediction subjected to feature selection using Kolmogorov-Smirnov statistics and area under the curve measures on the training data. In order to produce the final predictions, we have applied a mathematical conversion on the outputs of the discriminant function to produce empirical posterior probabilities that ranged between 0 and 1, indicating the likelihood of a future seismic event. At an additional level of complexity, we have also evaluated the performance of the predictions subject to different training datasets, as training datasets themselves vary in the level at which they represent the actual prediction problem. In the performance comparison tests over the training data, we observed varying accuracy levels for the different training datasets used, and submitted the best performing configuration to the AAIA'16 challenge, that achieved an area under the curve level of 0.9297 on the test data that was withheld from the challenge participants. The strengths of our proposed method lie first in the manner with which the hourly seismic activity measurements are evaluated and merged with the average measurements as well as existing risk assessment parameters. In addition, the simplicity of the Fisher's linear discriminant function offers a greater potential for generalizability of the demonstrated high performance to other seismic activity prediction cases as it minimizes the risk for overtraining. Finally, the conversion of the prediction results into posterior probabilities allows processing the results in conjunction with other probabilistic insights that one may have on the prediction problem at hand such as site-specific conditions and associated risks no reflected in the measurements. This also reflect the weakness of the method proposed here as it does no take into account any site-specific information, though this can be remedied in future applications.

REFERENCES

[1] A. Janusz, M. Sikora, Ł. Wróbel, S. Stawicki, M. Grzegorowski, P. Wojtas, and D. Ślęzak, "Mining data from coal mines: Ijcrs'15 data challenge," in *Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing*. Springer, 2015, pp. 429–438.

[2] M. Sari, H. S. B. Duzgun, C. Karpuz, and A. S. Selcuk, "Accident analysis of two turkish underground coal mines," *Safety Science*, vol. 42, no. 8, pp. 675–690, 2004. doi: http://dx.doi.org/10.1016/j.ssci.2003.11.002

[3] J. Van Zyl and C. W. Omlin, "Prediction of seismic events in mines using neural networks," in *Neural Networks, 2001. Proceedings. IJCNN'01. International Joint Conference on*, vol. 2. IEEE, 2001. doi: http://dx.doi.org/10.1109/IJCNN.2001.939568 pp. 1410–1414.

[4] J. V. Tu, "Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes," *Journal of clinical epidemiology*, vol. 49, no. 11, pp. 1225–1231, 1996. doi: http://dx.doi.org/10.1016/S0895-4356(96)00002-9

[5] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Annals of eugenics*, vol. 7, no. 2, pp. 179–188, 1936. doi: http://dx.doi.org/10.1111/j.1469-1809.1936.tb02137.x

[6] B. Bagwell, "A journey through flow cytometric immunofluorescence analyses—finding accurate and robust algorithms that estimate positive fraction distributions," *Clinical Immunology Newsletter*, vol. 16, no. 3, pp. 33–37, 1996. doi: http://dx.doi.org/10.1016/S0197-1859(00)80002-3

[7] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification*. John Wiley & Sons, 2012.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 235–238
DOI: 10.15439/2016F122
ACSIS, Vol. 8. ISSN 2300-5963

Utilizing an ensemble of SVMs with GMM voting-based mechanism in predicting dangerous seismic events in active coal mines

Łukasz Podlodowski
National Information Processing Institute
al. Niepodległości 188b 00-608 Warsaw, Poland
Email: lukasz.podlodowski@gmail.com

***Abstract*—This paper presents an application of a Gaussian Mixture Model-based voting mechanism for an ensemble of Support Vector Machines (SVMs) to the problem of predicting dangerous seismic events in active coal mines. The author proposes a method of preparing an ensemble of SVMs with different parameters and using the "wisdom of the crowd" for a classification problem. Experiments performed during the research showed an improvement in the quality of the classification after the mixture of Gaussian distributions was applied as votes distribution. The author also proposes a method of data selection for long sequences of measurement arranged chronologically with highly unbalanced occurrence of the positive class in the two-class classification problem. Finally, using the proposed model to solve the problem defined by the organizers of AAIA'16 DM showed an increase in the stability of the ensemble classifier and an improvement in the quality of the classification problem solution.**

I. Introduction

THE AIM of this paper is to present a solution to the problem introduced in AAIA'16 Data Mining Challenge: Predicting Dangerous Seismic Events in Active Coal Mines [1]. The task was related to the issue of predicting periods of increased seismic activity that may cause life-threatening accidents in underground coal mines. The task was divided into a classification problem of risk states with low hazard called "normal" and the state with high hazard called "warning". Application of hybrid methods of machine learning for a similar problem was presented in [2], but instead of a two-class problem, the authors of the experiments proposed a tripartite division: "normal", "warning" and "hazard", and focused on the medium-term (several minutes) forecasting of the maximum methane concentration at the wall end area. They also pointed out that "mathematical models correlating methane emission with methane content of a seam, ventilation method and geological features of mine workings facilitate overall prediction of average methane concentrations during exploitation of a working, nevertheless they cannot be applied for a direct short- or medium-term prediction of methane concentration" [2]. The problem discussed at AAIA'16 Data Mining Challenge focused on a different scope of time. Granulation of data is adjusted to one-hour windows. During this time, various kinds of information are accumulated, i.e. the number of registered seismic bumps of a specific energy level, or the average activity of the most active geophone. This problem does not allow to use information on the dynamics of changes within the hour during which signals were collected, and forces the participants of the challenge to process coarse-grained information about general characteristics of the signals.

Another solution for a similar problem with regression rule learning was described in [3]. The main objective of research presented in [4] was to reduce the number of forecasting errors during monitoring natural hazards and machinery in coal mines, achieved by the application of the regression rule induction, the k-nearest neighbors method, and the time series ARIMA forecasting.

A. Proposed solution

I propose a solution based on an ensemble of Support Vector Machines (SVM) of the kind described in [5][6], and on a voting mechanism based on the Gaussian Mixture Model described in [7]. Common approach based on ensemble of classifiers like boosting and bagging focus on preparing different training data set for each member of ensemble [8][9]. Instead of that my solution focus on receiving different information by changing parameters of SVMs in ensemble. Each SVM was trained on the same data. The GMM voting-based mechanism allows to extract correlation between SVMs outputs and evaluate likelihood of class occurrence. Process of preparing solution is illustrated in Fig. 1

The solution ranked 4th in AAIA'16 Data Minning Challenge with a final result 0.934. The final results were evaluated in accordance with Area under Curve values on a specially curated testing set. Finding parameters of model was mostly leaded by data driven strategy. Preparing solution focused on achieving balance between quality of solution scored by AUC value evaluated in cross-validation procedure and computation complexity.

II. Preparing data

The data set included 133 151 records, each corresponding to a 24-hour measurement. Vectors had 541 columns. Values stored in a single record can be divided into two separate parts. The first part consists of an identifier of the main working site and 12 other characteristics related to the whole period of 24 hours described by the record. The second part is composed of

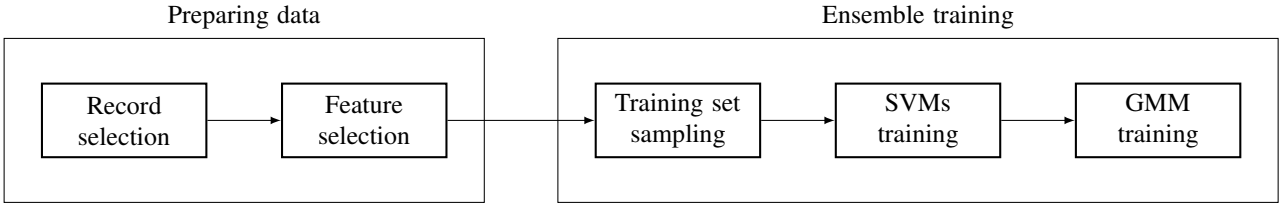


Fig. 1. Procedure of preparing data and training GMM-SVM classifier. Ensemble training was repeated 300 times and the best classifier was chosen based on cross-validation results.

hourly aggregated measurements, thus for each characteristic it includes 24 consecutive values associated with readings of geophones [1].

Measurement was labelled as "normal" or "warning" which indicate whether a total seismic energy perceived with 8 hours after the period covered by a data record exceeds the warning threshold, in correspondence with the classes prepared for solving the classification task. The distribution of classes was highly unbalanced. The "normal" class covered 130 187 of all records, while the "warning" class only corresponded to 2 962 "warning" measurements. Records were sorted chronologically, which highlighted the tendency of "warning" states to occur in short sequences. The testing set provided by AAIA'16 Data Mining Challenge consisted of 3 860 records.

A. Record selection

Before any operations on the data, the set vectors were linearly normalized to $[-1; 1]$. The first step towards the proposed solution was to choose the data vectors that would provide the most discriminative information. I assumed that this information was kept in boundary periods of an occurrence of "warning" measurement. Because of that, "normal" records were limited to the period corresponding to six hours before and to six hours after the "warning" sequence occurred. This approach permits focusing on the most important records and to solving the problem of highly unbalanced classes occurrence in the training set. In the next part of this paper all references to the training set are corresponding to the set prepared in this way.

B. Preparing base model

The next step was to prepare a base classification model which allows future data analysis. Since the problem is a high-dimensional one, it required a model with high resistance to high dimensionality of data. Another criterion of choosing the model was the fact that almost all attributes in a vector was represented by floating or integers values and represented measurements of signal sensors which suggested that data could be naturally represented in a Hilbert space. The chosen model was SVM. The procedure of adjusting parameters of hyperplane splitting space into areas corresponding to two classes is not highly sensitive to high dimensionality of space. Moreover, expanding dimensionality of space with a constant number of data records could enable the SVM to simplify finding optimal hyperplane splitting space into two subspaces problem [8]. The SVM was also designed to solve the problem of classification of two classes. Finally, proven high effectiveness of SVM for a classification problem [10] [11] made it a natural candidate to being the base classification model for this problem. Radial Basis Function (RBF) was chosen as kernel function, for it allows SVM to map a non-linear relationship between attributes and outputs. This ability is not approachable for the linear kernel. In [12], it has been shown that the linear kernel is a special case of the RBF kernel.

Before proceeding further, the SVM was trained on a set with all attributes. Parameters of the SVM were adjusted based on the grid search procedure. In the end, the base SVM took parameters $\gamma = 0.015625$ and $C = 2$.

C. Feature selection

Firstly, record attribute corresponding to the record ID was excluded from the training set. 529 attributes of training set records was grouped in a sequence of 24 elements corresponding to values in 24-hour period of measurement. A backward elimination was performed to drop out some of these sequences. The evaluation was based on 2-fold cross-validation method. The influence of removing the whole 24-hours sequence was investigated on each step. The procedure showed that removing 12 sequences significantly increased the accuracy of the base classificator and limited vector to 252 attributes. Four attributes corresponding to the latest available hazard assessment prepared by experts took on ordinal values, which represented the level of hazard were transformed to choose only from two values "no hazard" or "hazard".

III. Classification model

After preparing the data, a classifier based on the generated ensemble of SVMs classifier was proposed. Creation of this ensemble could be divided into two steps: choosing appropriate SVMs classifiers and preparing the voting mechanism.

A. Choosing SVMs

At this step, a grid search was performed and the best result it yielded was collected. I decided to use the base SVM described in the previous section of this paper and two additional SVM classifiers.

Since the high value of C allows the SVM to select more vectors as support vectors, the method could overfit. Thus, I took the first additional SVM $\gamma = 0.015625$ and $C = 1$. A second additional SVM was selected by getting a smaller value of γ, because a high value γ parameter could prevent the

SVM from finding the boundaries which allow to generalize the "shape" of the area covered by class. Based on the cross-validation results, I have chosen the SVM with parameters $\gamma = 1.22 \times 10^{(-4)}$ and $C = 1024$. Because procedure of training ensemble was repeated 300 times to avoid underfitting problem number of additionally SVMs was limited for computation time reduction.

Choosing base SVM and two additional SVMs for ensemble allow to achieve balance between computation complexity and quality of classification.

B. Voting mechanism

Instead of a simple voting mechanism, the GMM was used to represent a distribution of voting for each class. Since the testing set contains only 3 860 records, it was extremely probable that the sample would not have the same a priori distribution as the training set. It limits the possibility of a correct application of Bayes theorem in the classification model based on the estimated priori distribution. The priori likelihood of occurrence for all classes was assumed as equal in the testing data. In order to make the data represent the priori, I have limited the training set of the "normal" class for GMM to four hours before and four hours after the "warning" sequence occurs. The experiments showed that the results of the model has improved, since the data information about working wall, where measurement had been collected, was added to the vector. The working walls could be correctly identified by their IDs, but IDs do not fit well into the normal distribution âĂ¿ the base of the GMM. Hence, the ID was replaced with the height of a working wall.

The m parameter, representing the number of Gaussian components in the GMM for each class, was estimated on the basis of choosing the best results of the cross-validation procedure. If we describe parameters of a distribution as θ, a density function of the mixture distribution of features is described by:

$$f(x) = \sum_{i=1}^{m} w_i p(x \mid \theta), \tag{1}$$

where $p(x \mid \theta) = \mathcal{N}(x \mid \mu_i, \Sigma_i)$ corresponds to normal distribution that:

$$p(x \mid \theta) = \frac{1}{2\pi^{\frac{d}{2}}\sqrt{|\Sigma_i|}} \exp\left(-\frac{1}{2}(x-\mu_i)^T \Sigma_i^{-1}(x-\mu_i)\right). \tag{2}$$

Σ_i, μ_i and w_i are covariance matrix, mean vector and w_i represent the weight of the ith component in the mixture and d is a number of dimension in the modeled space. Parameters were estimated in EM procedure [13]. If y is assumed to be an unobserved data, then EM method takes the form of:

E-step: calculate the expectation of the unobserved data $E_{f(y|x,\theta^{(t)})}\left[\log f(x, y \mid \theta^{(t)})\right]$

M-step: find $\theta^{(t+1)}$ such that:
$\theta^{(t+1)} = \underset{\theta}{\operatorname{argmax}} E_{f(y|x,\theta^{(t)})}\left[\log f(x, y \mid \theta)\right]$

For n elements in the training set and an unobserved variable y_i^j takes:

$$y_i^j = \begin{cases} 1, & \text{if } i\text{th element was generated by } j\text{th component,} \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

To estimate parameters of the jth Gaussian component estimators take the form of:

$$w_j^{(t+1)} = \frac{1}{n}\sum_{i=1}^{n} E(y_i^j \mid x_i, \theta^{(t)});$$

$$\mu_j^{(t+1)} = \frac{\sum_{i=1}^{n}\left(E(y_i^j \mid x_i, \theta^{(t)}) x_i\right)}{\sum_{i=1}^{n} E(y_i^j \mid x_i, \theta^{(t)})}; \tag{4}$$

$$\Sigma_j^{(t+1)} = \frac{\sum_{i=1}^{n}\left(E(y_i^j \mid x_i, \theta^{(t)})(x_i - \mu_j^{(t+1)})(x_i - \mu_j^{(t+1)})^T\right)}{\sum_{i=1}^{n} E(y_i^j \mid x_i, \theta^{(t)})}.$$

The mixture of Gaussian components was prepared as a distribution of feature occurrences under the condition of class occurrence. Finally, the likelihood of the "warning" class was predicted on the basis of a posterior likelihood, which was evaluated basing on the Bayes theorem [14].

As Table I shows, the GMM with only one Gaussian component obtained the best results. It suggests that the SVMs' results tend to evaluate likelihood unanimously, which reduced the GMM-based voting mechanism to a single Gaussian component discriminant analysis.

C. Learning classifier

The training set for the ensemble of SVMs was different than for the GMM. Ensemble of SVMs was trained on a random sample of 30 percent of the training data. The objective of this mechanism was to avoid the overfitting problem.

Since the ensemble of SVMs was trained only on 30 percent of the training data, the risk of underfitting increased considerably. In order to solve this problem, the procedure of learning classifier GMM-SVMs was repeated. The model was learnt 300 times and the best train set for SVMs was chosen basing on the cross-validation results.

IV. Experimental results

Application of the SVM to the proposed solution was based on a LIBSVM implementation [15]. The organizers of AAIA Data Mining Challenge provided the training set in two steps. Initially, only half of the training data was available for participants. The rest of the training data was divided into four parts and supplied after some conditions, associated with the number of submissions, were met. My experiments focused on the quality of the classification performed by different approaches and the variance of results of SVM ensembles. For research variance of different strategy of voting I had collected results of an average voting, which corresponds to the voting based on simple mean of generated outputs of each SVM in

the ensemble and a GMM-based voting mechanism times 50 and then estimated the mean and standard deviation of the results.

The experiments concerned with data preparation were performed on the training set provided in the first stage. The quality of solution was described by AUC and evaluated in cross-validation procedure. Other experiments, concerning the classifier quality, covered the whole training data. All cross-validation procedures were performed with fold = 2. As shown in the Table I, GMM-SVM provided the best results. I compared the average voting strategy with the GMM voting mechanism. SVM ensemble based on a simple average voting achieved worse results, but still better than a single SVM classifier.

As shown in Table II, the results of an average voting are not stable. One reason may be that all SVMs lacked the ability to cope with underfitting, which resulted in higher standard deviation of the AUC results. Since the GMM learning method calculates parameters to fit the best maximum likelihood of the prediction based on the training set, it has the ability to obtain information on the errors of each SVM and the correlation between their outputs. This allows for the use of information about the areas in space that were problematic for some of the ensemble classifiers.

TABLE I
EXPERIMENTAL RESULTS - CLASSIFICATION QUALITY

Data set	Experiment	AUC %
First part of training data	Raw data	69.098
	After backward elimination	71.923
Whole training data	SVM	71.346
	Average vote	73.428
	GMM-SVM m = 1	**75.346**
	GMM-SVM m = 2	74.474

TABLE II
EXPERIMENTAL RESULTS - STABILITY OF VOTING MECHANISMS

Voting mechanism	Mean AUC %	Standard deviation of AUC %
Average vote	54.537	17.055
GMM-SVM m = 1	73.746	0.746
GMM-SVM m = 2	72.963	0.831

V. SUMMARY

This paper presents an application of the GMM-based voting mechanism for an ensemble of SVMs for the problem of predicting dangerous seismic events in active coal mines. The problem defined by the organizers of AAIA Data Mining Challenge allows for the successful use of GMM-SVM model of classification. My experiments showed that using the GMM voting instead of the average of outputs allows to decrease model variance. The GMM also makes obtaining information about classifier errors in the ensemble possible.

REFERENCES

[1] Aaia'16 data mining challenge: Predicting dangerous seismic events in active coal mines. [Online]. Available: https://knowledgepit.fedcsis.org/contest/view.php?id=112

[2] M. Sikora, Z. Krzystanek, B. Bojko, and K. Śpiechowicz, "Application of a hybrid method of machine learning for description and on-line estimation of methane hazard in mine workings," *Journal of Mining Science*, vol. 47, no. 4, pp. 493–505, 2011. doi: 10.1134/S1062739147040125. [Online]. Available: http://dx.doi.org/10.1134/S1062739147040125

[3] M. Kozielski, A. Skowron, Ł. Wróbel, and M. Sikora, *Beyond Databases, Architectures and Structures: 11th International Conference, BDAS 2015, Ustroń, Poland, May 26-29, 2015, Proceedings*. Cham: Springer International Publishing, 2015, ch. Regression Rule Learning for Methane Forecasting in Coal Mines, pp. 495–504. ISBN 978-3-319-18422-7. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-18422-7_44

[4] M. Sikora and B. Sikora, "Improving prediction models applied in systems monitoring natural hazards and machinery," *International Journal of Applied Mathematics and Computer Science*, vol. Vol. 22, no. 2, pp. 477–491, 2012. doi: 10.2478/v10006-012-0036-3. [Online]. Available: http://www.degruyter.com/view/j/amcs.2012.22.issue-2/v10006-012-0036-3/v10006-012-0036-3.xml

[5] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297. doi: 10.1007/BF00994018. [Online]. Available: http://dx.doi.org/10.1007/BF00994018

[6] V. Kecman, *Support Vector Machines: Theory and Applications*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, ch. Support Vector Machines – An Introduction, pp. 1–47. ISBN 978-3-540-32384-6. [Online]. Available: http://dx.doi.org/10.1007/10984697_1

[7] S. E. Jelali, A. Lyhyaoui, and A. R. Figueiras-Vidal, "Applying emphasized soft targets for gaussian mixture model based classification." in *IMCSIT*. IEEE, 2008. doi: 10.1109/IMCSIT.2008.4747229. ISBN 978-83-60810-14-9 pp. 131–136. [Online]. Available: http://dblp.uni-trier.de/db/conf/imcsit/imcsit2008.html#JelaliLF08

[8] T. J. Hastie, R. J. Tibshirani, and J. H. Friedman, *The elements of statistical learning : data mining, inference, and prediction*, ser. Springer series in statistics. New York: Springer, 2009. ISBN 978-0-387-84857-0 Autres impressions : 2011 (corr.), 2013 (7e corr.).

[9] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2006. ISBN 0387310738

[10] D. Meyer, F. Leisch, and K. Hornik, "The support vector machine under test," *Neurocomputing*, vol. 55, no. 1âĂ¿2, pp. 169 – 186, 2003. doi: http://dx.doi.org/10.1016/S0925-2312(03)00431-4 Support Vector Machines. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0925231203004314

[11] M. Ochab and W. Wajs, "Bronchopulmonary dysplasia prediction using support vector machine and libsvm," in *Proceedings of the 2014 Federated Conference on Computer Science and Information Systems*, ser. Annals of Computer Science and Information Systems, M. P. M. Ganzha, L. Maciaszek, Ed., vol. 2. IEEE, 2014. doi: 10.15439/2014F111 pp. pages 201–208. [Online]. Available: http://dx.doi.org/10.15439/2014F111

[12] S. S. Keerthi and C.-J. Lin, "Asymptotic behaviors of support vector machines with gaussian kernel," *Neural Computation*, vol. 15, no. 7, pp. 1667–1689, Jul. 2003. doi: 10.1162/089976603321891855. [Online]. Available: http://dx.doi.org/10.1162/089976603321891855

[13] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society: Series B*, vol. 39, pp. 1–38, 1977. doi: 10.2307/2984875. [Online]. Available: http://web.mit.edu/6.435/www/Dempster77.pdf

[14] S. J. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, 2nd ed. Pearson Education, 2003, ch. 13. Uncertainty, pp. 466–486. ISBN 0137903952

[15] C.-C. Chang and C.-J. Lin, "Libsvm: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 27:1–27:27, May 2011. doi: 10.1145/1961189.1961199. [Online]. Available: http://doi.acm.org/10.1145/1961189.1961199

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 239–243
DOI: 10.15439/2016F134
ACSIS, Vol. 8. ISSN 2300-5963

Predicting Dangerous Seismic Activity with Recurrent Neural Networks

Karol Kurach
University of Warsaw
Email: kk236085@mimuw.edu.pl

Krzysztof Pawlowski
University of Warsaw
Email: kpawlowski236@gmail.com

***Abstract*—In this paper we present a solution to the AAIA'16 Data Mining Challenge. The goal of the challenge was to predict, from multivariate time series data, periods of increased seismic activity which may cause life-threatening accidents in underground coal mines. Our solution is based on Recurrent Neural Network with Long Short-Term Memory cells. It requires almost no feature engineering, which makes it easily applicable to other domains with multivariate time series data. The method achieved the 5th place in the AAIA'16 competition, out of 203 teams.**

I. Introduction

UNDERGROUND coal mine workers are exposed to a life-threatening danger in a form of seismic events. To improve workers' safety, it is crucial to predict those phenomena in advance. However, knowledge-based safety monitoring systems that are currently deployed in coal mines sometimes fail to forecast such occurrences early enough. The goal of the AAIA'16 Data Mining Challenge: Predicting Dangerous Seismic Events in Active Coal Mines competition [12] was to design methods that could improve reliability of seismic activity prediction.

The task is an instance of a classification problem with unbalanced data provided in a form of multivariate, non-stationary time series. We present a solution based on Recurrent Neural Network with Long Short-Term Memory cells. The proposed model is generic and does not rely on the domain knowledge. It requires only minimal feature preprocessing and no feature engineering or feature selection steps. The solution achieved a competitive 5th place in the AAIA'16 competition.

The rest of the paper is organized as follows. In Section II we give an overview of the related work. The details of the AAIA'16 challenge are described in Section III. Section IV gives a brief introduction to Recurrent Neural Networks and Long Short-Term Memory cells. In Section V we describe the details of our architecture, training and model selection. Finally, Section VI summarizes the paper.

II. Related Work

Seismic hazard and rock bursts pose a threat to miners' lives and overall safety of the coal mining operation. One of the techniques for addressing this problem is to monitor the sensor readings' with automated algorithms. Originally, natural earthquake seismology approaches have been used to deal with the problem [3]. Mine-induced seismicity can be assessed using mechanisms of mine tremors, such as magnitude, moment, stress drop and seismic efficiency [16] or using seismic tomography [10]. More recently, typical machine learning approaches have been used – such as Random Forests [4] and other nonlinear methods [5], including Support Vector Machines or Naive Bayes Classifier.

The recent IJCRS 2015 Data Challenge competition [13] provided an opportunity to compare different approaches on the data set coming from coal mining. Although the goal was a bit different (to predict dangerous levels of methane concentration), the data shared similar characteristics – being an example of a time series, multivariate prediction problem with concept drift. Most of the top solutions relied heavily on feature engineering, either manual or automatic, such as: automatic variable construction [1], window-based feature engineering [8], hand-crafted features [15] or thousands of automatically generated features [21][22].

III. Challenge description

A. Data

The aim of the AAIA'16 competition was to predict relative likelihood of seismic events in coal mines based on the recorded measurements. It is an instance of supervised learning classification task, with most of the data given in a form of non-stationary multivariate time series. The data is split into 5 training sets and a single test set. All the training sets together contain $133,151$ records, while the test set contains $3,860$ records. Each record describes a period of 24 hours and consists of:

- an identifier of the main working site and 12 characteristics related to the whole period of 24 hours, such as total energy of seismic bumps registered in the last 24 hours,
- 22 times series with 24 numeric per-hour aggregated measurements, such as energy of the strongest seismic bump within a given hour.

Thus, in total each record contains 541 values. As mentioned previously, the records are grouped into 5 training sets and a single test set. The subsequent training sets correspond to later periods, adjacent records in them overlap by 1 hour and are given in a chronological order. The test set contains records that come from period later than the last training set, its records are non-overlapping and given in random order.

A label is given for each record in the training set while for the test set such label is missing – it is the goal of the

978-83-60810-90-3/$25.00

competition to forecast those values. The label is a categorical variable that can be either *normal* or *warning*. Value *warning* indicates that a total seismic energy measured within 8-hour period after the time covered by the record exceeded the warning threshold of $50,000$ Joules. For each record in the test set, the numeric predictions are to be made about those (hidden) values.

Additionally, there is an extra „meta-data" set that describes main working sites included in the training and test sets. It contains information such as the height of the main working site or the latest geological assessment. We note that the training and test sets are highly unbalanced, with respect to both the *main working site* attribute (Figure 1) and the labels (Table I).

B. Evaluation

The competition score is defined as an *area under the ROC curve*. It is calculated based on predictions of label values, made for all $3,860$ test set records. Each prediction is a number, where a higher value denotes that the true label value is more likely to be *warning*.

The contestants submit their predictions during the competition. However, before the competition is concluded, the contestants know only the score computed over *preliminary test set* – a part of the whole test set that contains approximately 25% of the data. This subset is chosen randomly by the organizers and is the same for all the contestants. It is not revealed to the participants which of the test records belong to it. The contestants can select a single final solution, possibly guided by the scores obtained on the preliminary test set. The final score, however, is computed over the *final test set*, which consists of the remaining approximately 75% of the test data. This score is shown only after the end of the contest and is used to compute the final standings – the highest-scoring team is declared the winner.

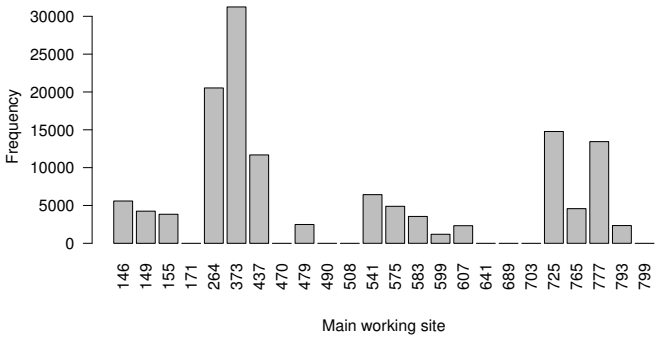


Fig. 1. The frequencies of different main working sites in the training sets. Some sites appear only in the test set.

TABLE I
DISTRIBUTION OF LABELS ACROSS THE TRAINING SETS

	tr. set 1	tr. set 2	tr. set 3	tr. set 4	tr. set 5
normal	78722	13137	13047	12744	12538
warning	1171	181	269	568	774

IV. Deep Recurrent Neural Network

A. Recurrent Neural Networks

Recurrent Neural Network (RNN) is a type of artificial neural network in which dependencies between nodes form a directed cycle. This allows the network to preserve a state between subsequent time steps. We focus on a simple RNN with a single, self-connected hidden layer.

RNNs process all elements from a sequence one-by-one, and the output at every time step depends on all previous inputs. This is a fundamental difference from feedforward networks, where the network's output depend only on the current element. It has an important theoretical implication: RNNs are capable of approximating arbitrary well any measurable sequence-to-sequence mapping [9].

Since RNNs contain loops, the standard backpropagation algorithm does not work. Instead, a *backpropagation through time* algorithm is used [20]. The idea behind this method is to unroll the network over N time steps, and copy the parameters N times. The RNN parameters are shared across all time steps, which makes them trainable and allows generalization.

Since the number of unrolled steps can be arbitrary, RNNs are particularly suited for modeling sequential data, where the length of the input is not fixed or can be very long. Recurrent nets have shown impressive results in many NLP tasks. One particularly successful variant of RNN is a recurrent network with LSTM cells, which we describe below.

B. Long Short-Term Memory

One important problem with training RNNs is the *vanishing gradient*, which can occur when values smaller than 1.0 are multiplied at each time step during the backpropagation through time. For some activation functions, the maximal value of the derivative is small. For example, the derivative of commonly used sigmoid function is never bigger than 0.25. As a result, after N time steps the gradient is multiplied by a value less than or equal to 0.25^N, which quickly becomes very small as N increases. While using some activation functions (eg. ReLU [17]) can reduce the likelihood of vanishing gradients, there is a special architecture designed to address this problem: Long Short-Term Memory (LSTM).

The LSTM is better at storing and accessing information than standard RNN [11]. The LSTM block consists of a self-connected memory cell and 3 gates named: input, output and forget. The gates control the access to the cell and can be interpreted as "read", "write" and "reset" operations in the standard computer's memory. The network learns to control the gates and decides to update and/or use the value at any given time step. Since all the components are built from

differentiable functions, the gradients can be computed for the whole system and it is possible to train it end-to-end using backpropagation. There are several variants of LSTM that slightly differ in connectivity structure and activation functions. Below we describe the definitions of the input, output and forget gates that we used.

Let $h_t \in \mathbb{R}^n$ be a hidden state, $c_t \in \mathbb{R}^n$ be a vector of memory cells of the network and let x_t be the input at the time step t. Let W_i, W_f, W_u, W_o be matrices and b_i, b_f, b_u, b_o the respective bias terms. We define LSTM as a transformation that takes 3 inputs $(h_{t-1},\ c_{t-1}, x_t)$ and produces 2 outputs (h_t and c_t). In all equations below $\odot$ is element-wise multiplication. We assume also that $\oplus$ is an operation that aggregates h_{t-1} and x_t. We used plain sum, but concatenation of vectors is also commonly used.

The *forget gate* which decides how much of the information should be removed from the cell is defined as:

$$f_t = \text{sigm}(W_f * [h_{t-1} \oplus x_t] + b_f) \tag{1}$$

The *input modulation* gate value i_t and the cell update u_t are defined as:

$$\begin{aligned} i_t &= \text{sigm}(W_i * [h_{t-1} \oplus x_t] + b_i) \\ u_t &= \tanh(W_u * [h_{t-1} \oplus x_t] + b_u) \end{aligned} \tag{2}$$

Intuitively, input modulation decides how much of the u_t should be added to the memory at step t. For example, if x_t can be ignored, i_t will be close to 0. Knowing the values above, the new cell value c_t is computed as:

$$c_t = f_t \odot c_{t-1} + i_t \odot g_t \tag{3}$$

The last step is to compute h_t, the output passed to the next LSTM's time step. It is controlled by the *output gate* o_t:

$$\begin{aligned} o_t &= \text{sigm}(W_o * [h_{t-1} \oplus x_t] + b_o) \\ h_t &= o_t \odot \tanh(c_t) \end{aligned} \tag{4}$$

The LSTM networks have been successfully applied to real-world problems, including language modeling [18], handwriting [7] or speech [6] recognition, and machine translation [19].

V. Model

A. Preprocessing

Recall from Section II that most of the solutions to the previous challenge depend heavily on feature engineering. Such approach, while effective in practice, makes the model less generalizable as the feature engineering steps depend on the problem at hand. Our goal was to create a model that learns everything from the raw data and does not rely on the domain knowledge. To this end, we limit our preprocessing only to the following two operations:

- **Data normalization**, in regards to *mean* and *standard deviation*. This is a standard Machine Learning procedure, and as such it should be applicable to almost any problem. The normalization makes easier both optimization of the loss function and the regularization, because all feature values are at the same scale.
- **Upsampling positive examples**. As presented in Table I, the ratio of positive to negative examples is highly skewed. To make it more balanced, we sample with repetition from the set of positive examples and add them to the training set. We experimented with different upsampling ratios and achieved best results for increasing the number of positives by $10 - 20$ times.

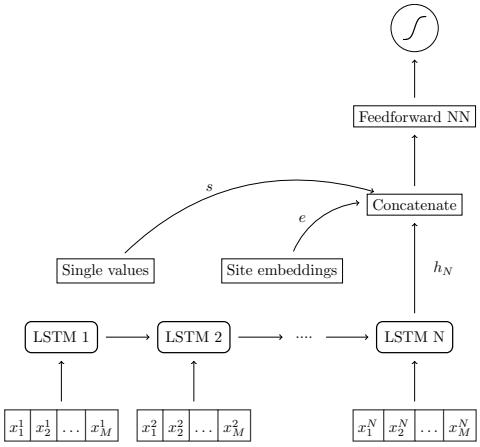


Fig. 2. Overview of the architecture. A core component is a single-layer LSTM unrolled for $N = 24$ time steps that processes $M = 22$ per-hour measurements. The ith per-hour measurement is marked as x^i. After processing N time steps, the last hidden state $h_N \in \mathbb{R}^{50}$ of the LSTM encodes information about all per-hour measurements. Then, h_N is concatenated with vector $s \in \mathbb{R}^{12}$ of per-record characteristics and the vector $e \in \mathbb{R}^{10}$ representing the working site id embedding.

B. Architecture

The overview of the architecture is presented in Fig. 2. We use a single-layer LSTM model that is processing 24 hourly aggregated measurements. At every time step, the hidden state of the LSTM ($h_i \in \mathbb{R}^{50}$) is connected to the previous state h_{i-1} and the normalized measurement values from the ith hour (x^i in the picture). After processing the whole sequence, the network's final hidden state h_N encodes all measurements in the order in which they appeared.

The vector h_N is concatenated with 2 other vectors: s and e. The vector s contains 12 per-record characteristics described in Section III-A. The vector e is an 10-dimensional embedding of the working site id. The values of the embedding vectors are initialized randomly and learned from the training data.

On top of the concatenation layer we build a standard supervised classifier (2-layer feedforward network in this case). We apply sigmoid on the network's output to ensure the predicted value is in the range $[0, 1]$ and can be interpreted as the probability of the *warning* label. The Binary Cross Entropy loss is used as the cost function.

TABLE II
WORKING SITES CHARACTERISTICS

site id	region name	bed name	assessment	mapped to
146	Partia F	416	a	N/A
149	Partia F	418	b	N/A
155	Partia H	502	b	N/A
171	Partia F	409	a	146
264	Z	405/2	b	N/A
373	G-1	707/2	b	N/A
437	G-1	712/1-2	b	N/A
470	Z	405/2	c	264
..	.....	....	.....	....
777	9	504	b	N/A
793	0	405	b	N/A
799	9	504	a	777

C. Training

We initialize all model's parameters by sampling uniformly from $[-0.1, 0.1]$. The optimization of the loss function is done using Adam algorithm [14] with a learning rate of 0.0005 and ε parameter equal to 10^{-10}. The training is run for 5 full passes (epochs) over the training data. After each epoch, the learning rate is multiplied by 0.63 and the training set is randomly shuffled.

We apply standard l_2 regularization of the weights with $\lambda = 0.01$. To avoid *exploding gradient* problem, the gradients are clipped globally to the value of 1. The model was implemented in Torch [2] and trained using a single GPU.

D. Model selection

Model selection was a significant challenge in the AAIA'16 competition. Recall from Section III-A that the time periods in the training data are overlapping. As a result, the standard cross-validation on a random split of the data tends to be over-optimistic. Also, there is a significant concept drift between the 5 provided training sets. The k-th training set was collected in a time period right after the set $(k-1)$th. We also know that the last training set was collected before the test set.

To address the problem of overlapping periods and to make the local evaluation as close to the final one as we can, we decided to use 5-fold cross-validation, with one training file being one fold. The average of the AUC scores was the final score we assigned to the model. We completely ignored the leaderboard score, as it proved to be very misleading in the past for this type of data [13].

E. Dealing with unknown sites

As described in Section V-B, our architecture computes embedding vectors for every working site id. However, recall from Fig. 1 that some of those identifiers exist only in the test data and not in the training data. We used the following method to fix this problem: we looked at the working site metadata and manually mapped 8 missing ids to existing ids that share similar characteristics. An example is presented in Table II which contains a subset of the metadata file. The id 171 is mapped to 146 because the region name and geological assessment is the same for those two sites. Similarly, id 799 is mapped to 777 because of the same region and bed names.

The above approach can be automated by joining (in SQL sense) the training data with the metadata on working site id attribute and then embedding different categorical variables (region/bed name, etc). This would remove the need of manual mapping of the missing ids and potentially improve the quality as well. However, we did not manage to try this approach during the competition.

F. Ensembling

From the begin of the competition, our main design decision was to create a competitive solution that consist of only one model trained from the raw data. However, the practice of machine learning competitions shows that ensembling of many different models is an easy way of improving the final score. We decided to do a simple rank average ensembling with a logistic regression model. More precisely, we use two models: RNN (described in section V) and LR (logistic regression) to evaluate records from the test set in the following way.

Let $\mathrm{rank}_X(\mathrm{record})$ denote the rank of the prediction given to the record by model X, among all predictions by model X on the test set. Then, the final prediction is computed as follows:

$$pred(record) = \frac{rank_{RNN}(record) + rank_{LR}(record)}{2}$$

By employing this technique we moved our solution one place up on the leaderboard.

VI. CONCLUSION

In this paper we presented a solution to AAIA'16 data mining challenge based on a Recurrent Neural Network with LSTM cells. It achieved a competitive score of 0.934 and the 5th place in the competition.

Compared to other methods (see Section II), our solution does not rely heavily on many hand-crafted features. Instead, it learns feature representation from the raw sensor data with a minimal feature engineering. It is a similar method to the one that we used in the previous IJCRS'15 competition, where our model achieved the 6th place. Top performance in both competitions suggests that our approach is versatile and can be successfully applied to different multivariate time series problems.

REFERENCES

[1] Marc Boullé. Prediction of methane outbreak in coal mines from historical sensor data under distribution drift. In *Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing*, pages 439–451. Springer, 2015.

[2] Ronan Collobert, Koray Kavukcuoglu, and Clément Farabet. Torch7: A matlab-like environment for machine learning. In *BigLearn, NIPS Workshop*, number EPFL-CONF-192376, 2011.

[3] C Allin Cornell. Engineering seismic risk analysis. *Bulletin of the Seismological Society of America*, 58(5):1583–1606, 1968.

[4] Long-jun Dong, Xi-bing Li, and PENG Kang. Prediction of rockburst classification using random forest. *Transactions of Nonferrous Metals Society of China*, 23(2):472–477, 2013.

[5] Longjun Dong, Xibing Li, and Gongnan Xie. Nonlinear methodologies for identifying seismic event and nuclear explosion using random forest, support vector machine, and naive bayes classification. In *Abstract and Applied Analysis*, volume 2014. Hindawi Publishing Corporation, 2014.

[6] Alan Graves, Abdel-rahman Mohamed, and Geoffrey Hinton. Speech recognition with deep recurrent neural networks. In *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, pages 6645–6649. IEEE, 2013.

[7] Alex Graves, Marcus Liwicki, Santiago Fernández, Roman Bertolami, Horst Bunke, and Jürgen Schmidhuber. A novel connectionist system for unconstrained handwriting recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(5):855–868, 2009.

[8] Marek Grzegorowski and Sebastian Stawicki. Window-based feature engineering for prediction of methane threats in coal mines. In *Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing*, pages 452–463. Springer, 2015.

[9] Barbara Hammer. On the approximation capability of recurrent neural networks. *Neurocomputing*, 31(1):107–123, 2000.

[10] David R Hanson, Thomas L Vandergrift, Matthew J DeMarco, and Kanaan Hanna. Advanced techniques in site characterization and mining hazard detection for the underground coal industry. *International journal of coal geology*, 50(1):275–301, 2002.

[11] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[12] Andrzej Janusz, Marek Sikora, Łukasz Wróbel, and Dominik Ślęzak. Predicting Dangerous Seismic Events: AAIA16 Data Mining Challenge. In *Proceedings of FedCSIS 2016*. IEEE, 2016. In print September 2016.

[13] Andrzej Janusz, Marek Sikora, Łukasz Wróbel, Sebastian Stawicki, Marek Grzegorowski, Piotr Wojtas, and Dominik Ślęzak. Mining data from coal mines: Ijcrsˇ201915 data challenge. In *Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing*, pages 429–438. Springer, 2015.

[14] Diederik Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[15] Petre Lameski, Eftim Zdravevski, Riste Mingov, and Andrea Kulakov. Svm parameter tuning with grid search and its impact on reduction of model over-fitting. In *Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing*, pages 464–474. Springer, 2015.

[16] A McGarr. Some applications of seismic source mechanism studies to assessing underground hazard. In *Rockbursts and Seismicity in Mines.*, pages 199–208, 1984.

[17] Vinod Nair and Geoffrey E Hinton. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, pages 807–814, 2010.

[18] Martin Sundermeyer, Ralf Schlüter, and Hermann Ney. Lstm neural networks for language modeling. In *INTERSPEECH*, 2012.

[19] Ilya Sutskever, Oriol Vinyals, and Quoc VV Le. Sequence to sequence learning with neural networks. In *Advances in neural information processing systems*, pages 3104–3112, 2014.

[20] Paul J Werbos. Generalization of backpropagation with application to a recurrent gas market model. *Neural Networks*, 1(4):339–356, 1988.

[21] Adam Zagorecki. Prediction of methane outbreaks in coal mines from multivariate time series using random forest. In *Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing*, pages 494–500. Springer, 2015.

[22] Adam Zagorecki. A versatile approach to classification of multivariate time series data. In *2015 Federated Conference on Computer Science and Information Systems, FedCSIS 2015, Lódz, Poland, September 13-16, 2015*, pages 407–410, 2015.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 245–248
DOI: 10.15439/2016F152
ACSIS, Vol. 8. ISSN 2300-5963

Automatic Feature Engineering for Prediction of Dangerous Seismic Activities in Coal Mines

Eftim Zdravevski, Petre Lameski, Andrea Kulakov
Faculty of Computer Science and Engineering
Ss.Cyril and Methodius University, Skopje, Macedonia
Email: {eftim.zdravevski,petre.lameski,andrea.kulakov}@finki.ukim.mk

***Abstract*—In this paper we present our submission to the AAIA'16 Data Mining Challenge, where the objective was to predict dangerous seismic events based on hourly aggregated readings from different sensor and recent mining expert assessment of the conditions in the mine. During the course of the competition we have exploited a framework for automatic feature extraction from time series data that did not require any manual tuning. Furthermore, we have analyzed the impact of overlapping of input data on model robustness. We argue that training an ensemble of classifiers with distinct (i.e. non-overlapping) chronological data rather than one classifier with all available data can produce more reliable and robust prediction models. By doing that, we were able to avoid overfitting and obtain the same score performance on the evaluation and test datasets, despite the significant data drift in the datasets.**

***Keywords*—feature engineering, feature selection, time series classification, temporal data mining, drift detection**

I. Introduction

The task in the AAIA'16 Data Mining Competition [1] is to devise a reliable prediction model for detecting periods of increased seismic activity that endangers miners working underground in coal mines. The data set consists of hourly aggregated readings from seismic sensors that count the number of seismic bumps perceived at longwalls and measure their total energy. Data records are composed of 24 consecutive hours of such readings coupled with the most recent assessments of the conditions at the longwalls made by mining experts. The target attribute in the data corresponds to information whether in a following period of 8 hours the total energy of seismic bumps exceeds a warning level. The full training dataset contains 133151 records with 541 columns, each corresponding to 24 hours of measurements. The evaluation metric of the competition was Area Under the ROC Curve (AUC).

The challenges of the competition were versatile: quite different distribution of working sites in the training and tests dataset that indicated of potential distribution drift in the data; the training class distribution was imbalanced; the final training set was 40 times larger than the test set; and the class distribution in the test set was unknown. After the competition ended it was disclosed that 2% of the records in the training and 5% in the test set were warnings. These challenges required very robust features and prediction models that would prevent over-fitting to the training set. In this paper we describe how our submission to the challenge addressed these challenges.

II. Cross-validation for model selection

In order to evaluate the models and feature sets locally, a way that will take into consideration the distribution of working sites, classes and train/test dataset size is required. To address this challenge, we developed a strategy that splits the training set keeping in mind the following parameters: general class distribution in the training set; class distribution per working site in the training set; and ratio of unknown versus known working sites in the test set.

Using these parameters, we tried to replicate a split of the train set into two sets, one for training and one for local testing, that would resemble the relation between the real training and test sets. When splitting the training set, we iteratively choose whether to add the working site to the train or test split based on its working class distribution, the current and desired class distribution in the splits, and the known vs unknown ratio in the test split. Using this approach, we were able to get a more realistic estimate of the feature set performance than the cross-validation schemes. We considered cross-validation scores to be more realistic if they resembled the leaderboard score, and thus were lower than the ones obtained with regular cross-validation. However, even with different feature subsets it was always 6% greater than the leaderboard score (always over 0.98).

Even though it was promising approach, we did not have time to develop a stratified or leave-one-subject-out cross-validation strategy based on it, so we have eventually abandoned it and relied for performance estimation based on the leaderboard score. In addition, this strategy based on the assumption that the training and test sets have similar class distribution, which was not specified in the task description. After the competition ended and the test labels were disclosed, it turned out that this assumption did not hold indeed.

III. Feature Engineering Framework

This competition coincided with our work on a framework for automated feature extraction from time series data. Therefore, it presented a good opportunity to test an early prototype of the framework on this dataset.

A. Feature generators

Using a systematic approach, the system is able to generate a variety of features that can robustly describe the dataset. A recent data mining competition for posture recognition of firefighters [2] inspired different feature engineering approaches

that are very effective [3, 4, 5]. Using the proposed approaches there, from each series of readings the system generates the following types of features: basic statistics (minimum, maximum, range, arithmetic mean, harmonic mean, geometric mean, median, mode, standard deviation, variance, skewness, kurtosis, signal-to-noise ratio, energy, etc.); curve fitting parameters [4]; equal-width histogram features [5]; percentile based features (first quartile, median, third quartile, inter-quartile range, amplitude, etc.) [3]; auto-correlation of the signal with several types of correlations (signal processing auto-correlation and Pearson, Spearman and Kendall correlation) [4]; and inter-correlations between each pair of raw time series values using the aforementioned types of correlation coefficients.

B. Time series generators

The feature extraction framework is able to generate new time series and then based on them to extract new sets of features, just like from an original time series. Authors of [3, 4, 5] demonstrated that this approach can further enhance the predictive performance of the system. Therefore, the framework uses the following time series generators (TSG): first derivatives of the original series [5]; amplitudes, frequencies and magnitudes obtained with fast Fourier transformation [3]; delta series (the deviations of the original values from the mean of the particular block of the time series), which can remove the seasonality in the data; sliding window time series [6]; and combining two time series by multiplying, subtracting or dividing their values [4].

C. Learning algorithms

For estimating the informativeness of individual features and the predictiveness of the whole problem the framework uses Random Forest (RF) [7] and Extremely Randomized Trees (ERT) [8] with high number of trees. They are used with the default parameters, as we have noticed that tuning them does not improve the performance dramatically (unlike with SVMs, for example). ERT models are very similar to RF in terms of predictive performance, but quite faster. We have noticed that when using the same number of trees ERT is significantly faster (over 50%) than RF, especially when the number of features is large (over 500).

D. Feature extraction and selection heuristics

Applying all possible feature transformations would generate very large number of features and would make learning based on them practically impossible. To mitigate this, the feature engineering and selection processes are interleaved, so generation of new time series and features is heuristically guided.

The algorithm for feature extraction and selection is shown in Figure 1. After the initialization and configuration of which features should be computed, the processing starts, one dataset instance (record) at a time. When the features are extracted from the whole dataset, it estimates the predictive performance and calculates feature importances in order to prepare a baseline for the feature selection loop that follows.

To improve the performance of the model under data drift, the framework performs greedy wrapper feature selection, inspired by the idea proposed in [9]. First it merges the training and validation or test dataset and generates an artificial "drift" class, denoting whether the instance is from the training or the test dataset. Then it estimates the importance of each feature when predicting the regular class (normal/warning) and the drift class (train/test). Afterwards, it calculates a set of feature importance thresholds for the regular class and the drift class. Next, in a loop it evaluates the performance different sets of features based on whether they fall within these importance thresholds.

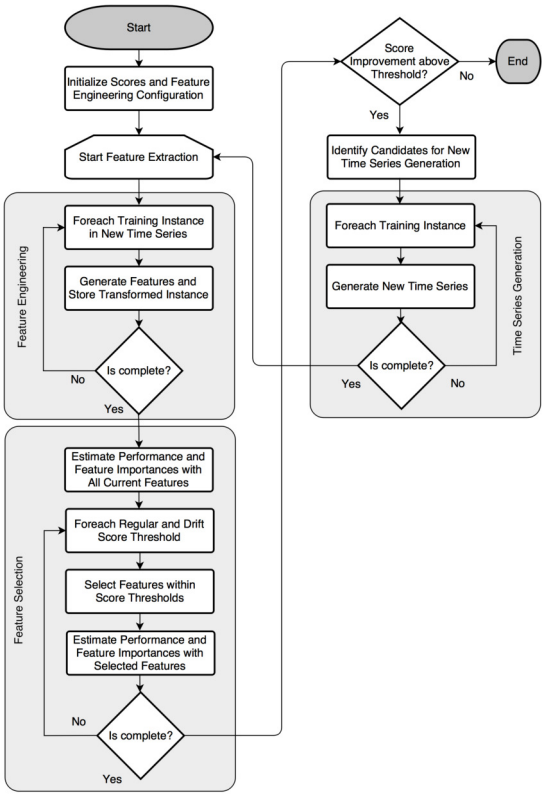


Fig. 1. Algorithm for feature extraction and selection

Next, by averaging the regular informativeness of the retained features for a particular time series the framework estimates which time series are informative for prediction of the target class. Additionally, it calculates correlations between different time series. Based on these heuristics, it identifies candidates for new which new series should be generated. After the new series are generated, the algorithm returns to the feature engineering phase. The second time it only generates features with only those transformations that resulted in retained features in the initial run. Afterwards, it merges the new feature set with the selected features in the previous iteration. The feature engineering and selection loop ends when generating new features does not improve the best performance by a considerable margin. For this competition, the we have only used the initial and one additional loop of generation of features.

Given that we were not able to replicate the training/test split for performing local evaluations, we decided to use the leaderboard scores for selecting the feature importance thresholds. By default, the framework uses cross-validation.

IV. The Feature Engineering Framework in Practice

The framework implementation it is still not publicly available. It has been implemented in Python on top of the scikit-learn Library [10]. In order to use the framework, we only need to specify the data set information: number and indexes of nominal features and numeric features, the number of time series and samples per series, the CSV files containing the description of columns and series. Then the framework takes over. It starts by calculating the basic statistics, inter-correlations, autocorrelations, histograms and quantiles. Based on them it evaluates which time series are informative. In this case, it discovered that the time series can be ranked as shown in Table I (columns IRI and IR - initial relative importance and initial rank, respectively).

This drift detection mechanism turned out to be very helpful for feature reduction. Namely, depending on the particular feature set size, removing the features that are good drift predictors improved our leaderboard score by 1-3 %. In our final solution the threshold 30 for both scores turned out to give good results thus removing almost 60% of the generated features. After their removal, the relative importance of different time series changed dramatically compared to the initial ranking, as can seen in Table I (columns IR and FR - initial and final rank, respectively).

A. Importance of time series

Some of the time series were significantly less informative than others, so they were discarded, keeping only the top 16 series. When evaluating the performance locally without the leaderboard, the system was able very quickly to find a feature set with AUC ROC score over 0.99 with stratified and regular cross-validation. When submitting those predictions to the public leaderboard we were able to get to a score of about 0.91. This dramatic drop of performance was a clear evidence of drift in the training and test datasets, which was somewhat expected due to the different mining sites.

TABLE I. Time series importance before and after importance evaluations. IRI=Initial Relative Importance, IR=Initial Rank, FRI=Final Relative Importance, FR=Final Rank

Time Series Name	IRI	IR	FRI	FR
max_gactivity	1.000	1	1.000	1
avg_difference_in_gactivity	0.998	2	0.788	9
max_difference_in_gactivity	0.962	3	0.786	10
avg_difference_in_genergy	0.954	4	0.771	11
max_difference_in_genergy	0.938	5	0.766	12
avg_genergy	0.916	6	0.886	4
max_genergy	0.889	7	0.902	3
avg_gactivity	0.885	8	0.909	2
highest_bump_energy	0.663	9	0.861	6
sum_e2	0.410	10	0.742	15
sum_e3	0.345	11	0.874	5
total_number_of_bumps	0.310	12	0.793	8
sum_e4	0.192	13	0.766	13
count_e2	0.175	14	0.646	16
count_e3	0.140	15	0.744	14
count_e4	0.116	16	0.811	7
sum_e5	0.040	17		
count_e5	0.022	18		
sum_e6plus	0.017	19		
count_e6plus	0.013	20		
number_of_distressing_blasts	0.004	21		
number_of_rock_bursts	0.000	22		

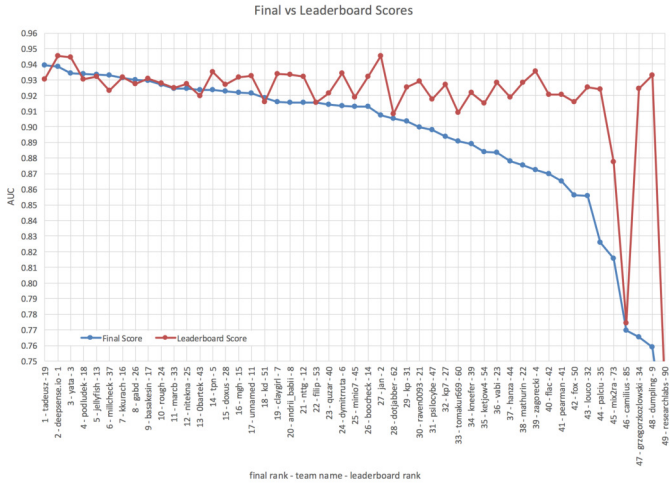


Fig. 2. Final vs. Leaderboard scores of the competitors.

As described earlier, the cross-validation scores were significantly different the leaderboard scores. On the other hand, different feature subsets performed similarly to each other with cross-validation and on the leaderboard. Because of those two reasons and the limited time we had for spending on this competition we decided to use only the features of the original time series and disable generation of new time series.

B. Oversampling in datasets

The fact that the size of the training set was significantly larger than the test set, pointed out that there must be some over-sampling in the training instances. To investigate this, we have analyzed the training dataset comparing consecutive rows that are for the same working site. We have discovered that in most cases the consecutive rows are shifted by 1 hour, thus overlapping for 23 hours. However, the test instances do not have any overlapping (or at least it can not be discovered). This mean that the training and test dataset are not identically and independently distributed. That combined with the over-sampling posed significant risk of over-fitting. In order to alleviate that, we have split the training set into 24 folds of instances that do not overlap. In particular, if we have consecutive overlapping instances (1 shift, 23 hours overlap) numbered with $1..n$ from the same working site, in the i-th fold the following instances would be assigned: $i, i+s, i+2\times s, ...$, where $s = 24$ is the number of folds. Likewise, we have tested with 12, 8, 6, 4 and 2 folds. Then, having s folds we train a separate classifier on each fold and then average the predictions of all s classifiers. Regardless of the feature subset and used classifier, when using 24 folds gave best results, about 1-3% more than training one classifier on the complete training set (i.e. only one fold). This was very peculiar discovery that should be considered when creating batches from continuous streams of data. A similar effect of over-sampling in datasets was discovered and explained in [11].

C. Final submission

The predictions based on different feature subsets and different training data subsets scored similarly on the public leaderboard. In order to diversify the predictions and aiming

to achieve more robust prediction models, we aggregated the results from 5 different predictions: 4 ERT models trained with various feature subsets that had 200-900 features (various basic statistics, histograms, percentiles, linear and quadratic fit coefficients, correlations, etc. obtained from the 16 retained time series listed in Table I) and 1 logistic regression model trained with only 21 features (reduced from the larger feature set with Correlation-based Feature Selection). This improved our leaderboard score to 0.9276, about 1.5% better than any the individual classifiers that were used in the ensemble. Some classifiers (e.g. logistic regression) made predictions close to 0 while all others were predicting exactly 0 (the probability of warning). To leverage these properties, we decided to determine whether the prediction is 0 or not with simple majority vote. If not then the individual predictions are averaged. Instead of averaging the predictions in such cases, predicting 0 improved the leaderboard score by 0.5%.

As it can be seen from Figure 2, many of the teams had dramatically over-fitted their models to the training and leaderboard datasets, thus their performance of the final dataset dramatically dropped. Only few teams were able to produce the same or better score on the final dataset as on the leaderboard.

V. Conclusion

This competition provided an interesting and interactive opportunity to tackle various challenges encountered in real-world data analysis. Our initial goal was to test if automatic feature extraction and selection from time series data would be feasible for such problem. We were able to discover and understand interesting patterns in the data, such as the ranking of importance of time series or other important features like expert seismic assessments. Additionally, we discovered the importance of overlapping in the training dataset in relation to over-fitting. Another interesting realization was that even simple classifiers like logistic regression with very simple features can give leaderboard score of about 0.92. Our best leaderboard score was average of predictions of several different classifiers and/or different feature sets and it was about 0.928.

It is important to point out that the process of automatically generating features did not require manual tuning. Furthermore, the framework was able to generate and recognize informative features and time series, as well as to retain only features that are robust to drift in the data. The only manual work that we performed was related to the cross-validation experiments and for submitting solutions to the competition system. We consider that the obtained performance is significant due to the automatically generated features. We acknowledge that the framework still requires work to be done in relation to more efficient searching of redundant features and more optimal iterative generation of new features. Our experiments with other datasets confirms that it is able to match the best published performance, or in some cases to even improve them.

References

[1] A. Janusz, M. Sikora, Ł. Wróbel, and D. Ślezak, "Predicting Dangerous Seismic Events: AAIA16 Data Mining Challenge," in *Proceedings of FedCSIS 2016*. IEEE, 2016, in print September 2016.

[2] M. Meina, A. Janusz, K. Rykaczewski, D. Slezak, B. Celmer, and A. Krasuski, "Tagging firefighter activities at the emergency scene: Summary of aaia'15 data mining competition at knowledge pit," in *Computer Science and Information Systems (FedCSIS), 2015 Federated Conference on*, Sept 2015. doi: 10.15439/2015F426 pp. 367–373.

[3] J. Lasek and M. Gagolewski, "The winning solution to the aaia'15 data mining competition: Tagging firefighter activities at a fire scene," in *Proceedings of the 2015 Federated Conference on Computer Science and Information Systems*, ser. Annals of Computer Science and Information Systems, M. P. M. Ganzha, L. Maciaszek, Ed., vol. 5. IEEE, 2015. doi: 10.15439/2015F418 pp. 375–380.

[4] A. Zagorecki, "A versatile approach to classification of multivariate time series data," in *Proceedings of the 2015 Federated Conference on Computer Science and Information Systems*, ser. Annals of Computer Science and Information Systems, M. P. M. Ganzha, L. Maciaszek, Ed., vol. 5. IEEE, 2015. doi: 10.15439/2015F419 pp. 407–410.

[5] E. Zdravevski, P. Lameski, R. Mingov, A. Kulakov, and D. Gjorgjevikj, "Robust histogram-based feature engineering of time series data," in *Computer Science and Information Systems (FedCSIS), 2015 Federated Conference on*, ser. Annals of Computer Science and Information Systems, M. P. M. Ganzha, L. Maciaszek, Ed., vol. 5. IEEE, Sept 2015. doi: 10.15439/2015F420 pp. 381–388.

[6] M. Grzegorowski and S. Stawicki, "Window-based feature engineering for prediction of methane threats in coal mines," in *Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing*, ser. Lecture Notes in Computer Science, Y. Yao, Q. Hu, H. Yu, and J. W. Grzymala-Busse, Eds. Springer International Publishing, 2015, vol. 9437, pp. 452–463. ISBN 978-3-319-25782-2

[7] A. Liaw and M. Wiener, "Classification and regression by randomforest," *R news*, vol. 2, no. 3, pp. 18–22, 2002.

[8] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," *Machine Learning*, vol. 63, no. 1, pp. 3–42, 2006. doi: 10.1007/s10994-006-6226-1

[9] M. Boullé, "Tagging fireworkers activities from body sensors under distribution drift," in *Proceedings of the 2015 Federated Conference on Computer Science and Information Systems*, ser. Annals of Computer Science and Information Systems, M. Ganzha, L. Maciaszek, and M. Paprzycki, Eds., vol. 5. IEEE, 2015. doi: 10.15439/2015F423 pp. 389–396.

[10] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[11] M. Boullé, *Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing: 15th International Conference, RSFDGrC 2015, Tianjin, China, November 20-23, 2015, Proceedings*. Cham: Springer International Publishing, 2015, ch. Prediction of Methane Outbreak in Coal Mines from Historical Sensor Data under Distribution Drift, pp. 439–451. ISBN 978-3-319-25783-9

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 249–252
DOI: 10.15439/2016F188
ACSIS, Vol. 8. ISSN 2300-5963

Application of RapidMiner and R Environments to Dangerous Seismic Events Prediction

Katarzyna Dusza*, Dominik Korda*, Krzysztof Kozłowski*, Bartłomiej Szwej*,
Michał Kozielski†, Marcin Michalak*‡, Marek Sikora*‡, Łukasz Wróbel‡
*Institute of Informatics, Silesian University of Technology
ul. Akademicka 16, 44-100 Gliwice, Poland
Email: {katadus637,domikor355,bartszw537}@student.polsl.pl
Email: krzysztofkozlowski1993@gmail.com
Email: {Marcin.Michalak, Marek.Sikora}@polsl.pl
†Institute of Electronics, Silesian University of Technology
ul. Akademicka 16, 44-100 Gliwice, Poland
Email: Michal.Kozielski@polsl.pl
‡Institute of Innovative Technologies EMAG
ul. Leopolda 31, 40-189 Katowice, Poland
Email: {Marek.Sikora, Lukasz.Wrobel, Marcin.Michalak}@ibemag.pl

***Abstract*—Underground coal mining is a branch of an industry which safety of operation is very dependent on the natural hazards. A proper seismic event prediction is a significant aspect of building classification models from the real data, which can affect the coal mining safety increase. In this paper four models, built in a well known data mining environments, are presented. The obtained models, depending on a given implementation of popular methods, occurred comparable to the best results from the competition.**

I. Introduction

THOUGH the production of energy from renewable resources has been increasing recently, the mining is still an important part of the industry. There are many countries — even in Europe — which produce most (e.g., 83% in 2012, Poland) or almost half (e.g., 45% in 2012, Germany) of its energy from coal. This means that problems of coal mining — especially the underground coal mining — still have a global meaning.

One of the aspects of safe and efficient coal mining is prediction of seismic hazards. Safety refers to saving workers from the accidents and injuries while efficiency refers to unplanned stops of longwall systems. Analysis and proper prognosis of potentially dangerous methane concentration [1, 2, 3] and seismic events [4, 5, 6] should improve the safety and reduce the costs of underground coal mining.

This paper presents several solutions of a problem of predicting dangerous seismic events in hard coal mines. This classification problem was a goal of a AAIA'16 competition for data scientists. The paper is organized as follows: it starts from a short description of a competition, data provided to the competitors and the evaluation method that was applied to submitted models. Then, the four models and the way of their development are described. The paper ends with a comparison of the result quality delivered by both the presented approaches and the contest winner followed by a final conclusions.

This work was partially supported by Polish National Centre for Research and Development (NCBiR) grant PBS2/B9/20/2013 within Applied Research Programmes. The infrastructure was supported by "PL-LAB2020" project, contract POIG.02.03.01-00-104/13-00.

II. Competition Task

This paper describes solutions submitted to the AAIA'16 data mining competition which summary is presented in [7]. Data for this edition of the competition came from the hard coal mining industry and was provided by Research and Development Centre EMAG. The main goal of the analysis was a prediction of dangerous seismic events in coal mines. The following part of the paper presents more detailed description of the data provided for the contest and a method of model evaluation. More detailed information about the competition can be found in [8].

The objective of each competitor was to devise a reliable prediction model able to detect periods of increased seismic activity that endangers miners working underground in coal mines.

A. Data

The competition training file contained 79,893 records, each corresponding to 24 hours of measurements. Values stored in a single record could be divided into two parts. The first one consisted of an identifier of the main working site and 12 other characteristics related to the whole period of 24 hours described by the record. The second part consisted of hourly aggregated measurements that count the number of seismic bumps perceived at longwalls and measure their total energy, thus, for each characteristic it included 24 consecutive values.

There was a total number of 541 columns in the data (including the main working site id). There was also available a separate file with additional information about all main working sites included in the data (in the training and test parts).

Labels in the data indicated whether a total seismic energy perceived within 8 hours after the period covered by a data record exceeded the warning threshold (i.e. $5 \cdot 10^4$ Joules). The labels of the test series were hidden from participants. It is important to note that time periods in the test data did not overlap and they were given in a random order.

An additional impediment for competitors and their models was the fact that the data was unbalanced. 78,722 records belonged to a "normal" class while the rest of them (only 1,171) was labelled as "warning".

The goal for the competition participants was to predict likelihood of the label "warning" for the records from the test set. For the consecutive objects exactly one real number corresponding to the predicted likelihood should be placed. The values did not have to be in a particular range, however, higher numerical values should indicate a higher chance of the label "warning".

B. Evaluation

The submitted solutions were evaluated on-line and the preliminary results were published on the competition leaderboard. The preliminary score was computed on a subset of the test set, fixed for all participants. It corresponded to approximately 25% of the test data. The final evaluation was done after completion of the competition using the remaining part of the test data. Those results were also published on-line. The assessment of solutions was done using the Area Under the ROC Curve (AUC) measure.

III. Overview of Solutions

The presented solutions were developed by students at the Institute of Informatics, Silesian University of Technology. Participation in the competition was an additional and optional activity for the students of the Computational Intelligence and Data Analysis course. The best achievements in the competition was promoted by the exemption from the final exam.

During the university course the students learn, among others, R [9] and RapidMiner [10] environments. Therefore, these two environments were applied by them to solve the competition task. Among the solutions presented in this paper one was developed in RapidMiner and the other three were developed in R environment. Besides, two solutions implemented the Artificial Neural Network (ANN) model and the other two implemented Boosted Trees model. The details of the data preprocessing and the model parameters are presented in the following paragraphs.

A. Solutions based on the Artificial Neural Network model

The presented solution was defined in the RapidMiner environment, where the process was based on the *Neural Net* operator. The whole process is presented in the Fig. 1. The usage of a *Nominal to Numerical* operator in the process was planned as a constant mapping of the consecutive (increasing) levels of threats a, b, c, d to the increasing integer values $1, 2, 3, 4$. In this approach the set of independent variables was reduced and therefore, the following variables were taken into consideration:

- latest_seismic_assessment,
- latest_comprehensive_assessment,
- max_gactivity.24,
- max_genergy.24,
- total_number_of_bumps.24.

These variables were determined by trial and error, starting from the attributes that are highly correlated with the predicted variable.

The final model of the network had 8 neurons in a hidden layer and the following set of initial parameters of the *Neural Net* operator was chosen:

- training cycles: 700,
- learning rate 0.05,
- momentum: 0.2,
- decay: False,
- shuffle: True,
- normalize: True,
- error epsilon: 1.0E-5,
- use local random seed: True,
- local random seed: 1337

The final prediction quality of this model — submitted by Krzysztof Kozłowski as *unnamed* to the Knowledge Pit platform — expressed by means of AUC criterion was calculated as 0.9215.

The second solution based on the ANN model was developed in R environment. The H2O platform [11] which can be used in R environment was chosen as an implementation of neural network engine. One of the reasons of this implementation selection was the fact that it is very well documented and many helpful remarks on neural network parameter tuning are available [12].

Due to the data structure where 24 hour measurements were contained in each record, it was required to aggregate information from 24 columns representing consecutive hours of a day into a single one. The other attributes were selected on the basis of their correlation. From the results of experiments it occurred that due to normalization of the data the quality of results decreased. Thus, this processing method was abandoned.

The following set of initial parameters of artificial neural network was chosen:

- neuron activation function: $\tanh$ with dropout,
- number of neurons in a hidden layer: 5,
- input neurons droput: 0.1,
- hidden neurons droput: 0.3,
- classes balancing: turned off,
- maximal number of epochs: 300,

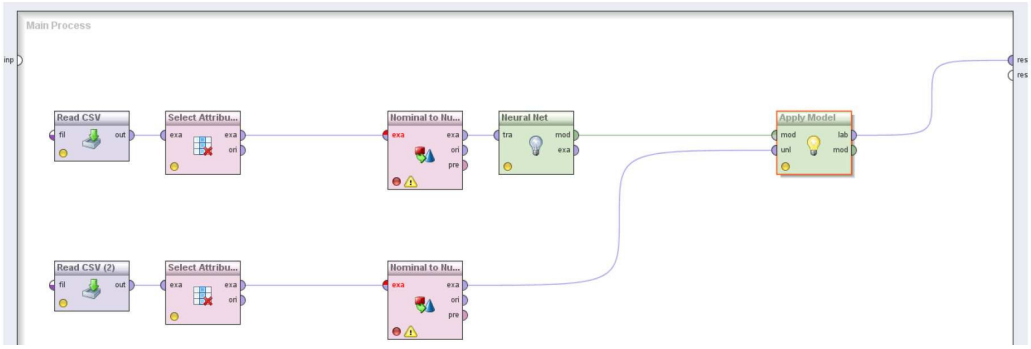


Fig. 1. Process of building the prediction model in RapidMiner.

The rest of the parameters was set to default values. The obtained neural network gave the following result expressed by means of AUC criterion: 0.9101.

Next, further tuning of parameters was performed what resulted in the following parameter values:

- neuron activation function: rectifier with dropout,
- number of neurons in a hidden layer: 4,
- input neurons droput: 0.2,
- hidden neurons droput: 0.3,
- classes balancing: turned off,
- maximal number of epochs: 250,
- $L1$ regularization: 10^{-5},
- adaptive speed of learning: turned off,
- number of training objects per iteration: 1000,
- number of testing objects: 80,
- maximum duty cycle fraction for scoring: 0.

This solution submitted by Dominik Korda and identified as *doxus* at the Knowledge Pit platform achieved the final prediction quality (AUC) value equal to 0.9225.

B. Solutions based on the Boosted Trees model

There are two approaches based on the Boosted Trees method presented in this section. Both of them were developed in R environment and both of them utilised a *caret* package [13]. The approaches differ due to the data processing stage.

Within the first approach the Kibana tool [14] was applied to visual analysis of the attributes. It enabled to notice an important association between the attribute *latest_seismic_assessment* and the decision attribute. Therefore, this attribute was included into the model in the first order.

Another important observation was connected with the *total_destressing_blasts_energy* attribute: for all objects that have a *warning* value of a decision, *total_destressing_blasts_energy* equals 0. Therefore, it was decided to introduce a new derived variable named *tdbeGTzero* (total destressing blasts energy greater than zero) defined in the following way:

$$\text{IF } total\_destressing\_blasts\_energy > 0$$

$$\text{THEN } tdbeGTzero = true$$

$$\text{ELSE } tdbeGTzero = false$$

The final set of the selected independent variables was as follows:

- sum_e3,
- sum_e4,
- sum_e5,
- sum_e6plus,
- highest_bump_energy,
- max_genergy,
- avg_genergy,
- tdbeGTzero,
- latest_seismic_assesment.

This solution, submitted by Bartłomiej Szwej and identified as *0bartek* at the competition platform, achieved the final prediction quality (AUC) value equal to 0.9238.

The set of independent attributes of the second approach was selected arbitrarily and it contained:

- latest_seismic_assessment,
- latest_seismoacoustic_assessment,
- latest_comprehensive_assessment,
- latest_hazards_assessment.

These attributes categorize the seismic activity into four levels (a, b, c, d), and a proper value is set by a domain expert working in a coal mine.

This solution, submitted by Katarzyna Dusza and identified as *kd* at the competition platform, achieved the final prediction quality (AUC) value equal to 0.9185.

IV. Results and Conclusions

Over one hundred (106) competitors accessed the challenge and 49 of them submitted their results. The quality of the top ten approaches and the solutions presented above are listed in Table I.

If we take into consideration all 49 results it can be stated that students' models placed higher than the median of all of them (25^{th} result was 0.91304342). This enables a positive assessment of these students' involvement to the contest. It is also worth to be noticed that some of them did not limit themselves only to tuning the method parameters but also tried to select and derive explainable attributes for the model.

TABLE I
SELECTED RESULTS FROM THE FINAL BOARD OF AAIA'16 DATA MINING CHALLENGE.

rank	participant	AUC
1	tadeusz	0.9393
2	deepsense.io	0.9384
3	yata	0.9342
4	podludek	0.9336
5	jellyfish	0.9336
6	millcheck	0.9329
7	kkurach	0.9312
8	gabd	0.9300
9	basakesin	0.9297
10	rough	0.9269
13	**0bartek**	0.9238
15	**doxus**	0.9225
17	**unnamed**	0.9215
18	**kd**	0.9185
49	researchlabs	0.6998

Additionally, it can be noticed that all the presented solutions were developed in a well known data mining environments. Therefore, these approaches are more general at the level of model creation, where only parameter tuning was performed. Besides, the presented solutions focused on a proper data pre-processing in order to select and derive the right independent variables.

Tuning of the parameters was performed in case of ANN-based approaches and the results presented in Tab. I show its positive impact. However, in case of the approaches based on the Boosted Trees model, where the parameters were identical and the results (see Tab. I) were significantly different, it is visible how important is the data processing phase of analysis.

Finally, from the university course leader perspective the involvement of the students into such data analysis competition is very promising. The students have the opportunity to operate on a real-life data and to compare the quality of their results with the other competitors. Therefore, it can be twofold interesting for them and hopefully it will increase their motivation to further studies.

REFERENCES

[1] M. Sikora and B. Sikora, "Improving prediction models applied in systems monitoring natural hazards and machinery," *International Journal of Applied Mathematics and Computer Science*, vol. 22, no. 2, pp. 477–491, 2012. doi: 10.2478/v10006-012-0036-3. [Online]. Available: http://dx.doi.org/10.2478/v10006-012-0036-3

[2] ——, "Rough natural hazards monitoring," in *Rough Sets: Selected Methods and Applications in Management and Engineering*. Springer, 2012, pp. 163–179. [Online]. Available: http://dx.doi.org/10.1007/978-1-4471-2760-4-10

[3] A. Zagorecki, "Application of sensor fusion and data mining for prediction of methane concentration in coal mines," *Mining — Informatics, Automation and Electrical Engineering*, vol. 524, no. 4, pp. 33–38, 2015.

[4] J. Kabiesz, B. Sikora, M. Sikora, and Ł. Wróbel, "Application of rule-based models for seismic hazard prediction in coal mines," *Acta Montanistica Slovaca*, vol. 18, no. 3, 2013.

[5] J. Kabiesz, "Effect of the form of data on the quality of mine tremors hazard forecasting using neural networks," *Geotechnical & Geological Engineering*, vol. 24, no. 5, pp. 1131–1147, 2006. doi: 10.1007/s10706-005-1136-8. [Online]. Available: http://dx.doi.org/10.1007/s10706-005-1136-8

[6] A. Leśniak and Z. Isakow, "Space-time clustering of seismic events and hazard assessment in the Zabrze-Bielszowice coal mine, Poland," *International Journal of Rock Mechanics and Mining Sciences*, vol. 46, no. 5, pp. 918–928, 2009. doi: 10.1016/j.ijrmms.2008.12.003. [Online]. Available: http://dx.doi.org/10.1016/j.ijrmms.2008.12.003

[7] A. Janusz and et al., "Predicting dangerous seismic events in active coal mines: Summary of AAIA'16 data mining competition at knowledge pit," *Proc of FedCSIS 2016*, vol. 00, no. 00, pp. 00–00, 2016.

[8] AAIA'16 data mining challenge: Predicting dangerous seismic events in active coal mines. [Online]. Available: https://knowledgepit.fedcsis.org/contest/view.php?id=112

[9] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2014. [Online]. Available: http://www.R-project.org

[10] RapidMiner. Rapidminer. [Online]. Available: http://rapidminer.com

[11] H2O platform. [Online]. Available: www.h2o.ai

[12] The definitive performance tuning guide for h2o deep learning. [Online]. Available: http://blog.h2o.ai/2015/02/deep-learning-performance/

[13] The caret package. [Online]. Available: http://topepo.github.io/caret/index.html

[14] Kibana software. [Online]. Available: www.elastic.co/products/kibana

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 253–259
DOI: 10.15439/2016F508
ACSIS, Vol. 8. ISSN 2300-5963

Face Occlusion Detection Using Skin Color Ratio and LBP Features for Intelligent Video Surveillance Systems

Pengfei Ji
Chongquing Key Lab. of Computational Intelligence, Chongquing University of Posts and Telecommunications, China
Email: jipfchn@163.com

Yonghwa Kim, Yong Yang
Dept. of Information & Communication Engineering, Inha University, Incheon 402-751, Korea
Email: yonghwa@inha.edu

Yoo-Sung Kim
Dept. of Information & Communication Engineering, Inha University, Incheon 402-751, Korea
Email: yskim@inha.ac.kr

***Abstract*—A face occlusion detection scheme which is based on both skin color ratio (SCR) and Local Binary Pattern (LBP) feature, is proposed. The proposed method mainly consists of four steps: foreground extraction, head detection, feature extraction, and occlusion detection. First, foreground is extracted by codebook background subtraction algorithm. Then, the head region is located using HOG head detector. After that, the skin-color ratio and LBP feature are extracted. Finally, SVM is trained based on LBP feature. The recognition result of SVM and the result of skin-color ratio feature are merged by weighted voting strategy, and then occluded faces are classified as three categories: concealed, partially concealed, and visible. Experimental results show that the proposed detection system can achieve desirable results in intelligent video surveillance systems.**

I. Introduction

Nowadays, intelligent video surveillance system has been a very active research field, and has many practical applications, such as face identification, behavior recognition, abnormal event detection, and so on. It greatly reduced the manual inspection and verification to improve the effectiveness. Face occlusion detection is an essential component of intelligent video surveillance system. This is necessary for the prevention of suspicious behavior in security zones, such as bank, government, and other sensitive areas. With the increasing requirements for security, face occlusion detection has become a hot research topic in pattern recognition and image processing. If the surveillance system can automatically detect a person with concealed face, it would be helpful for making more secure society. For example, the system can be widely used for automated teller machine (ATM) security.

Concealed face detection is a very challenging task in image understanding field due to the complexity of environment condition, such as low resolution, illumination changes, and object movement, etc. Under complex conditions, there is no guarantee of image quality, so face occlusion detection becomes extraordinary difficult. Furthermore, the designed system requests reduced computation and high reliability, so that it can be able to realize the real-time processing requirement. Although various methods have been proposed to detect occluded face, but most of them are for ATM, this fact leads to a problem that these methods cannot be used directly for video surveillance systems.

In order to solve the above-mentioned problems, we proposed an efficient approach that combines skin color ratio feature and texture feature for concealed face detection. The method consists of the main four steps: (1) Extraction of moving foreground by using codebook model which is a real-time segmentation algorithm, (2) Head detection based on HOG head detector, which only focus on the foreground moving region, (3) Skin color feature and LBP feature are extracted, (4) The two assessment methods are fused by weighted voting and the final recognition result is obtained.

The main contributions of this work can be stated as follows: (1) A novel framework of face occlusion detection in the application of the real-world video surveillance environment is proposed, (2) The skin color feature and texture feature with LBP operator are merged by weighted voting strategy in order to improve detection performance.

The rest of this paper is organized as follows: Section 2 introduces the related works on face occlusion detection. Section 3 describes the details of the proposed method. Section 4 presents the experimental results and analysis. Section 5 draws a conclusion with the future plan.

II. Related Works

Generally, SCR-based methods [1]–[4] adopt various skin models, extract color information, and then calculate skin ratio. Lin *et al.* proposed a method based on the combination of ellipse fitting and skin area ratio to determine whether the face is concealed or not [1]. Kim *et al.* presented a face occlusion verification method that combines a shape-based face detection and skin color [2]. Zhang *et al.* used Adaboost to combine skin color detection and face templates matching to produce the strong classifier for occlusion detection [3]. Charoenpong *et al.* proposed a face occlusion detection method from a viewpoint of face by skin color ratio of two parts of head re-

gion [4]. These methods present an effective detection system for bank ATM application. However, it can only handle the detection problem in which the target is close to the camera. Therefore, these methods cannot be directly used in intelligent video surveillance environments.

On the other hand, learning-based methods [5]–[9] require feature extraction to train the classification model, and then achieve classifying recognition. Yoon *et al.* proposed a method based on principal component analysis (PCA) and support vector machine (SVM). By applying PCA and SVM to extract the feature points and classify that the feature points are near the normal or concealed face [5]. Min *et al.* presented a novel learning based method for scarf detection. The method exploits Gabor wavelet feature and SVM classifier to distinguish the normal and scarf faces [6]. Priya *et al.* proposed a Mean Based Weight Matrix (MBWM) algorithm, which is an improved form of LBP, and the SVM classifier to detect face occlusion [7]. Bianco *et al.* designed a recognizability of concealed face system, in which Adaboost algorithm is taken to build the facial feature detectors [8]. Kim *et al.* also proposed a face occlusion detection method based on gradient map and SVM, used especially for the partially concealed face [9]. Although good results are reported in these methods, the robustness cannot be guaranteed under complex conditions. In particular, surveillance systems cannot always produce high quality images. Therefore, the proposed method should handle both high quality and low quality images to meet the result of robustness.

In addition, component-based methods [10], [11] use facial component regions to detect face occlusion. Suhr *et al.* proposed a face occlusion detection system, which combine the eye-mouth combinations and geometric constraints [10]. Eum *et al.* presented a face recognizability evaluation to use facial components and the exceptional occlusion handling (EOH) [11]. This method is insensitive to illumination condition and achieves the robustness against facial postures, but it needs a high computational cost.

Nowadays, deep learning is a hot topic in machine learning, and CNN is one of deep learning methods, which can learn hierarchical features from low-level data [12]. Xia *et al.* proposed a robust and effective facial occlusion detection method based on CNN and multi-task learning [13]. The trained CNN model is applied in extracting facial features and predicting face occlusion. However, the disadvantages are complex structure, and it requires large-scale and complete training data.

III. FACE OCCLUSION DETECTION SCHEME USING SKIN COLOR RATIO AND LOCAL BINARY PATTERN

This section mainly introduces the details of our proposed method. The processing procedure includes the following steps: foreground extraction, head detection, feature extraction, and occlusion detection. The proposed system architecture is shown in Fig. 1.

A. Foreground extraction

Extracting foreground accurately is the first step for face occlusion analysis. By foreground extraction, the moving foreground is segmented in the video sequence. There are many ways for foreground segmentation, such as Gaussian Mixture Model (GMM) and background subtraction. Compared with the above methods, the codebook model can detect foreground more accurately and efficiently, and handle scenes containing moving backgrounds or illumination variations [14]. It is a kind of adaptive background subtraction algorithm by building a codebook for each pixel. So the information of the foreground areas in current frame is obtained with codebook model. Some examples of foreground segmentation using codebook model are shown in Fig. 2.

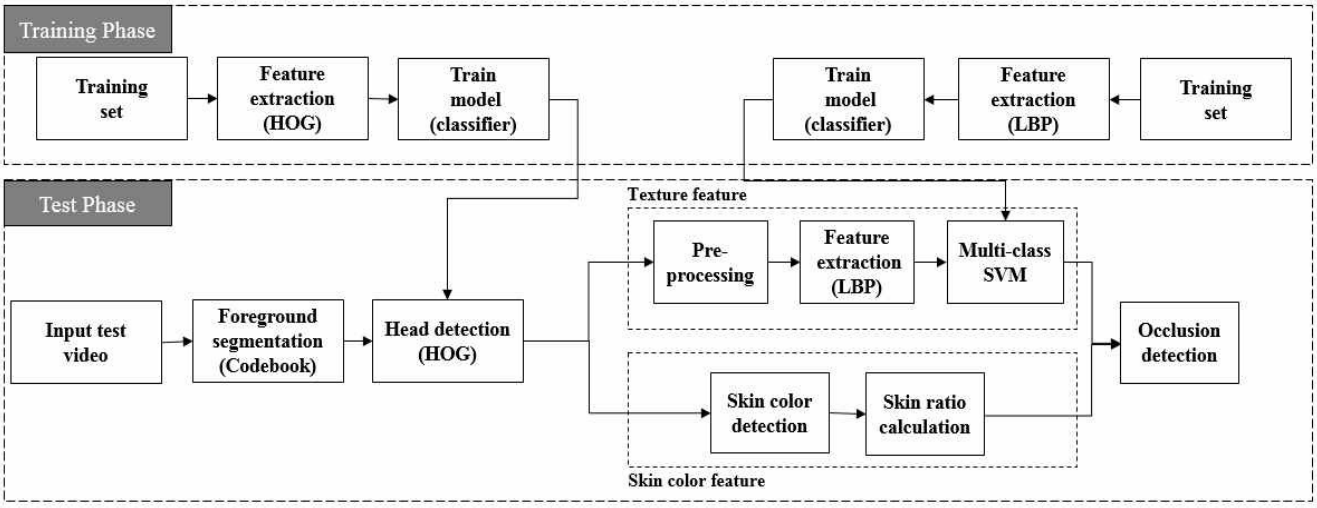


Fig 1. The framework of face occlusion detection system

(a) (b)

Fig 2. Foreground extraction results: (a) original (b) Codebook

B. Head detection

In our method, head detection is accomplished by combining HOG and SVM method proposed by Dalal and Triggs [15] and the part-based models proposed by Felzenszwalb *et al.* [16]. HOG can well describe the edge or local shape information, so it is suitable to depict characteristics of head. Meanwhile, this method is capable of handling head detection with challenging illumination conditions and low resolution.

While the HOG is one of the popular descriptors in image processing, it needs a high computational cost. To address the problem, GPU technique is used to accelerate process by using a parallel implementation of the HOG algorithm. Comparing CPU implementation, GPU based on parallel architecture has a better computational performance [17], [18]. The result of head detection demonstrates that our implementation using GPU can achieve a speedup of over 5 times, which allows for real-time detection. Moreover, the scanning area only focuses on the moving foreground region, and the probability of false and failure detections is significantly reduced. Examples of head detection are shown in Fig. 3.

Fig 3. Some results of head detection

C. Feature extraction

The analysis of skin color is adopted to discriminate skin and non-skin pixels. An unconcealed face includes a large number of skin color pixels. The whole process of skin color detection consists of two steps: color space transform and skin ratio calculation.

1) Color Space Transform: RGB is a basic color model. Because of the sensitivity to illumination variations, it is hardly separated from chrominance and luminance information [19]. Compared with RGB, YCbCr has a clear distinction between chrominance and luminance components. Thus the first step is to convert color space from RGB to YCbCr representation. Using (1), we transform RGB to YCbCr color space.

$$\begin{bmatrix} Y \\ C_b \\ C_r \end{bmatrix} = \begin{bmatrix} 65.481 & 128.533 & 24.966 \\ -37.797 & -74.203 & 112.0 \\ 112.0 & -93.786 & -18.241 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} + \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} \quad (1)$$

2) Skin ratio calculation: In chromatic space, the skin color mainly focuses on a narrow range. Obviously, we can determine the skin pixel by judging whether the YCbCr value is in a certain range. As shown in (2), the skin ratio is defined as the skin pixels of the head region.

$$S = \frac{number\ of\ skin\ pixels}{total\ \ number\ of\ head\ region\ \ pixels} \quad (2)$$

Where, S is the skin ratio. Because it is hard to ensure the detection result with only one frame, we can utilize multiple frames to ensure high accuracy. Consequently, the calculation of the ratio comes out from the combination of multi-frames.

For the reason that the camera will be a certain distance from the people in real scenes, it makes the face appear smaller and vaguer. However, with the movement of the body, the size of head region is also varying constantly. To assess the impact, scoring method is used for the evaluation process. Large size can obtain a higher score than head image of small size, but the performance is going to reach a specific limitation [20]. Given a specific value μ, it is suitable for the width and height. Using (3), we represent the score related to the size.

$$Q = min\left\{1, \frac{width}{\mu} \times \frac{height}{\mu}\right\} \quad (3)$$

Therefore, this method, which applies the head size as criterion, is used to realize multi-frame selection. When the head size is larger than a preset threshold, these frames are selected in video sequence. Let N be the number of frames, and we obtain the final ratio to integrate the scores. It can be expressed as

$$R = \frac{\sum_{i=0}^{N} S_i Q_i}{\sum_{i=0}^{N} Q_i} \quad (4)$$

where, S_i is the skin ratio of i-th frame, and Q_i is the score of i-th frame.

Based on fusion strategy, we need to know the classification result for the detected N frames. Thus we introduce two ratio threshold parameters τ_1 and τ_2, which will be discussed in the following section. For every frame, the

occlusion result can be determined by comparison of two parameters. Let X_i be the number of detection result for each class in all frames, *i= {1,2,3}*, which denote concealed, partially concealed, and visible, respectively. Assume p_i is the likelihood for every class, it is calculated by using (5).

$$p_i = \frac{X_i}{N} \tag{5}$$

In practice, the above skin color method is sensitive to light condition. And when face covers by objects of colors similar to the skin color, its effect is not ideal. To overcome this problem, we extract texture feature with LBP operator [21]. As a popular texture descriptor, it describes well the image texture information. Therefore, we can effectively distinguish them by the texture. The operator labels the pixels of an image by thresholding the neighborhood of each pixel with the center value and forming the result of binary pattern. Then, a pattern code is computed as follow

$$LBP_{(P,R)} = \sum_{n=0}^{P-1} s(g_n - g_c)2^n \tag{6}$$

$$\mathrm{s(x)} = \begin{cases} 1, & x \geq 0 \\ 0, otherwise & \end{cases} \tag{7}$$

where, g_c corresponds to the gray level of the center pixel, g_n to the grey values of the neighbor pixels. The notation (*P*, *R*) indicates *P* sampling points on a circle of *R* radius. This paper uses the uniform LBP, and the selection of *P* and *R* will be discussed in the next section.

The feature extraction process includes the following steps. The first step is pre-procession, such as normalization and histogram equalization. Second, we divide the image into several sub-blocks, and the feature histogram is extracted from each sub-block. Then, all the local histograms are concatenated into a feature vector. The more the block number is, the higher the computational complexity is. Therefore, a suitable block number should be selected. And this issue is explored future in the next section. Fig. 4 shows the process of feature extraction.

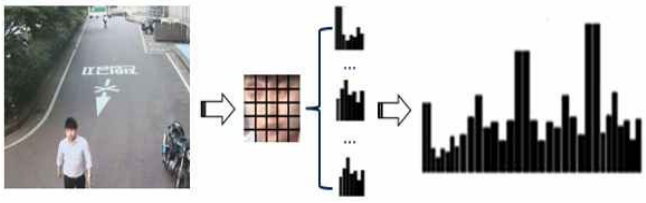


Fig 4. LBP feature extraction

In this paper, the detection of concealed faces can be considered as a multi-class problem: concealed, partially concealed, visible. Therefore, we use a multi-class SVM to fulfill the concealed face classification. Furthermore, assume X_i is the number of detection result for each class in all frames, which represent concealed, partially concealed, visible respectively, *i={1,2,3}*. Let p_i' is the likelihood for each class and it is calculated by using (8).

$$p_i' = \frac{X_i}{N} \tag{8}$$

D. Occlusion detection

The system uses the idea of decision level fusion to improve performance. On the phase of decision fusion, the weighted voting strategy is used to fuse the result for all levels to get target detection output. First, we determine whether it is the concealed face by skin color ratio in the skin color method, and whether it is the concealed face based on LBP. Then we assign different weight according to the contribution of various assessment methods. Using (9), we calculate the weighted vote result.

$$W_i = p_i w_1 + p_i' w_2 \tag{9}$$

where W_i is the weighted results, *i={1,2,3}*, W_i denote the weighted result in three classes: concealed, partially concealed, visible, respectively. p_i and p_i' can be obtained by using equation (5) and (8). w_1 and w_2 are the weights of *i*-th method, respectively, and they are measured by

$$w_x = \frac{n_x}{\sum_{x=1}^{k} n_x} \tag{10}$$

where k is the number of methods. Because the system has two methods, namely, k=2. n_x is the recognition rate based on various assessments on test video set, $x = \{1,2\}$. The idea of weight distribution comes from [22]. Eventually, the largest W_i will be chosen to determine the concealed face classification.

IV. EXPERIMENTS

To evaluate the system performance, experiment inputs are taken on a dataset with 138 video samples, which are collected from indoor or outdoor surveillance cameras. The video set includes three classes: concealed, partially concealed, visible. We chose 60 video samples as the training set, and the rest is used as test set. The video sequence has a resolution of 720×576 at 15 frames per second. The face covers by some objects such as hat, mask, sunglasses, book, etc. The sample set contains the different times of the day. The experiment is carried out with an Intel(R) Core(TM) i5-3230M CPU @2.6GHz and NVIDIA GeForce 710M. Fig. 5 illustrates some examples of face occlusion in surveillance cameras.

Fig 5. Examples from surveillance camera for the concealed face

A. Head detection

In order to evaluate the head detection accuracy and acceleration effect based on GPU implementation, experiments are taken. In experiments, head samples are collected from surveillance cameras, and then the head detector is trained on manually annotated data. The data set has approximately 5,000 head images as positives training samples and randomly chosen windows that do not contain any head area as negative training samples from surveillance videos. The size will vary depending on the distance between the people and camera, and is normalized to 64×64 pixels.

The algorithm is tested on video samples with different resolutions. Table I shows the processing time with the GPU and CPU implementations. It can be concluded that GPU implementation speeds up more than 5 times for the processing time over CPU implementation. And the GPU implementation shows more significant effect of acceleration for higher resolution image. On the other hand, to evaluate the accuracy of head detection, we carried out the experiments on several video clips. The experimental results are summarized in Table II. Hit rate can be calculated as the number of detected heads divided by total number of heads, while the false alarm rate is the number of false heads divided by total number of heads [23]. The results show that our approach has high detection rate and can reduce the false alarm rate effectively.

TABLE I.
THE COMPARISON OF PROCESSING TIMES FOR DIFFERENT RESOLUTION

Resolution	CPU Time(ms)	GPU Time(ms)	Overall Speedup
320*240	126.10	24.63	5.12
720*576	653.59	97.08	6.73
1280*1024	2083.33	299.40	6.94

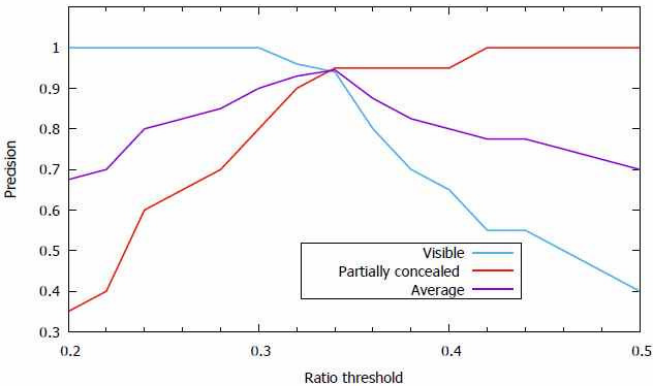
TABLE II.
ANALYSIS OF HIT RATE AND FALSE ALARM RATE

Total frames	1462
Total number of heads	1512
Number of detected heads	1373
Number of false heads	16
Hit rate(HR)	90.8%
False alarm rate(FAR)	1.0%

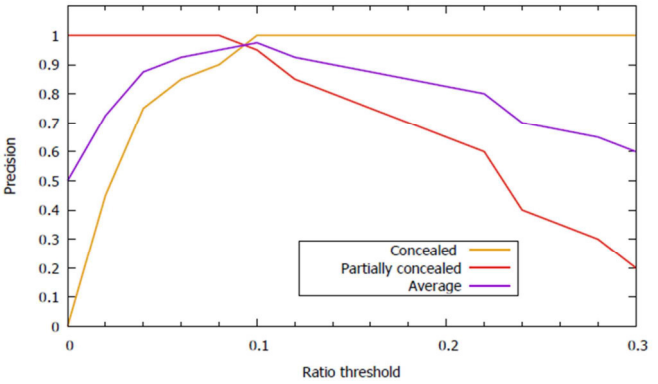
B. Threshold determination in skin color ratio

The skin ratio threshold is important factor that directly affects the accuracy of face occlusion detection. In order to achieve higher results, the optimal threshold should be determined. In our experiment, we use training video set to decide the optimal threshold. The result is illustrated in Fig 6. As seen in the Fig 6(a), it shows that along with the increasement of skin color ratio, the recognition rate of visible face is declining gradually, and partially concealed face is increasing at the same time. There is a peak value of average recognition rate at 0.34. On the other hand, a conclusion can be drawn from Fig. 6(b), which is similar to the above conclusion, and the peak point appears at 0.10. Then, let τ_1 and τ_2 denote two thresholds respectively, that is, the ratio thresholds are set to be τ_1= 0.34 and τ_2 = 0.10.

If the ratio is more than τ_1, this case is called visible. If the ratio is less than τ_2, it's called concealed. In addition to this, all other cases belong to partially concealed. Finally, we have tested its performance on the test video set. The average accuracy can reach to 87 %.



(a)



(b)

Fig 6. Precision of various ratio threshold on training video set.

C. Parameter determination for LBP

In order to analyzing the influences of some parameters, a series of experiments is carried out. In our experiment, we manually crop concealed & unconcealed face images from the training video set. All these concealed faces are normalized to 90×90 pixels. Then multi-class SVM is trained to detect the concealed face. In testing phase, we use the trained classifier to detect whether the face is concealed or not in every frame.

Taking into account the feature dimension, we choose uniform patterns of LBP operator. We adopt cross validation to determine some parameters. Let *N*N* and *R* denote the number of blocks and sample radius. In the experiment, the parameters are set to *N={1,2,3,5,6,9}*, *R={1,2}*. Fig. 7 gives the comparison with the recognition rate of different parameters. From this figure, it indicated that the recognition rate increased with the incensement of block number. But the feature dimension has doubled and redoubled when the number of blocks increased. Therefore, we select *N=5, R=1* as the best choice, which can both maintain a relatively high precision and lower dimension. The trained classifier can be used to classify the unknown face samples. Similarly, we have tested it on test video set. The average accuracy can reach to 70%.

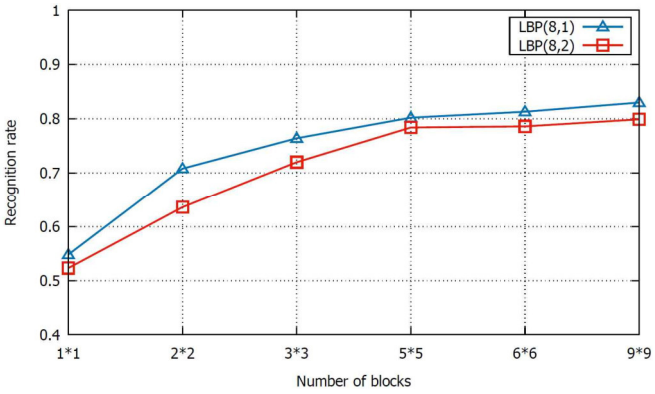


Fig 7. Comparison with the recognition rate of different parameter

D. Fusion strategy

For the purpose of comparing these two methods, the fusion experiment is carried out. In the fusion strategy, the average accuracy is above 95% by combining the above two methods in the test video set. Fig. 8 summarizes the confusion matrix. As can be seen from the chart below, our fusion strategy can improve the recognition rate obviously. In SCR approach, the concealed face was often classified as partially concealed face, primarily because of the individual difference of human beings, such as short and thin hair, which largely affects the skin color ratio. In LBP method, the partially concealed face has the poor accuracy of recognition, and this is due to the diversity of partially concealed face. Therefore, the fusion strategy verifies its effectiveness.

V. CONCLUSIONS

This paper proposed a novel scheme to detect concealed face for intelligent video surveillance systems. It combines texture feature & skin color feature to detect concealed face. To ensure the accuracy, we have studied a multi-frame detection algorithm. There are two parts in this method. First, we introduced and analyzed the codebook foreground segmentation model, and the head region is located by using HOG with GPU implementation. Then the decision fusion is used to validate concealed face by combining LBP and skin color. In the future work, we are planning to improve overall system performance through robust method for feature extraction, and solve head pose estimation problem.

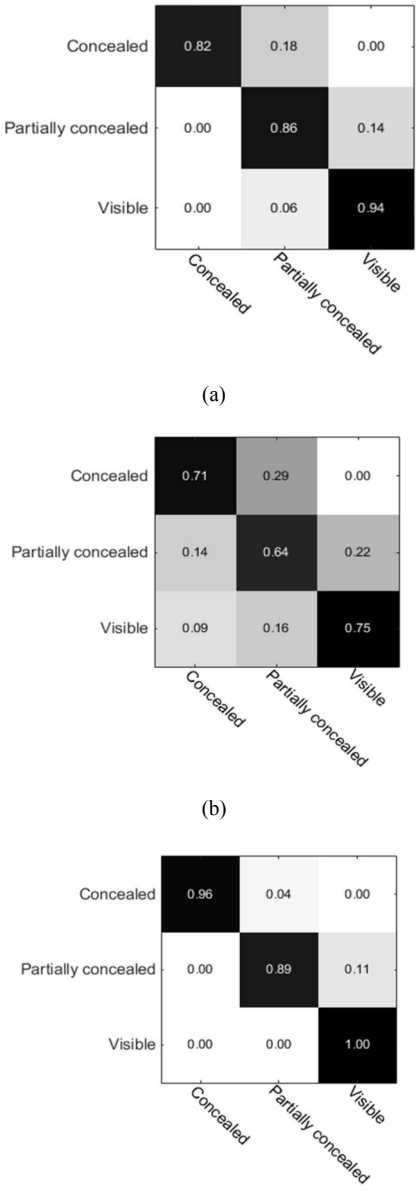


Fig 8. Confusion matrices for three occluded cases on test video set. (a) SCR (b) LBP (c) Fusion

ACKNOWLEDGMENT

This work was supported by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIP) (B0101-15-1282-00010002, Suspicious pedestrian tracking using multiple fixed cameras).

REFERENCES

[1] D. T. Lin and M. J. Liu, "Face occlusion detection for automated teller machine surveillance," in Advances in Image and Video Technology. Springer, pp. 641–651, 2006.

[2] G. Kim, J. K. Suhr, H. G. Jung, and J. Kim, "Face occlusion detection by using b-spline active contour and skin color information," in 11th International Conference on IEEE Control Automation Robotics & Vision (ICARCV), pp. 627–632, 2010.

[3] X. Zhang, L. Zhou ,T. Zhang, and J. Yang, "A novel efficient method for abnormal face detection in ATM," in International Conference on. IEEE Audio, Language and Image Processing (ICALIP, pp. 695–700), 2014.

[4] T. Charoenpong, C. Nuthong, and U. Watchareeruetai, "A new method for occluded face detection from single viewpoint of head," in 11th International Conference on IEEE Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), pp. 1-5, 2014.

[5] S. M. Yoon and S. C. Kee, "Detection of Partially Occluded Face using Support Vector Machines," in IAPR Workshop on Machine Vision Applications, pp. 546-549, 2002.

[6] R. Min, A. D'Angelo, and J. L. Dugelay, "Efficient scarf detection prior to face recognition," in Proceedings of the 18th European Signal Processing Conference, pp.259-263, 2010.

[7] G. N. Priya, and R. S. D. W. Banu, "Detection of occluded face image using mean based weight matrix and support vector machine," in Journal of Computer Science, Vol. 8, no. 7, pp.1184-1190, 2012.

[8] S. Bianco, G. Ciocca, G. C. Guarnera, A. Scaggiante, and R. Schettini, "Scoring recognizability of faces for security applications," in Proc. SPIE 9024, Image Processing: Machine Vision Applications VII, 90240L, 2014.

[9] J. Kim, Y. Sung, S. M. Yoon, and B. G. Park, "A new video surveillance system employing occluded face detection," in Lecture Notes in Computer Science, vol. 3533, pp. 65-68, 2005.

[10] J. K. Suhr, S. Eum, H. G. Jung, G. Li, G. Kim and J. Kim. "Recognizability assessment of facial images for automated teller machine applications." in Pattern Recognition, vol. 45, pp. 1899-1914, 2012.

[11] S. Eum, J. K. Suhr, and J. Kim. "Face Recognizability Evaluation for ATM Applications with Exceptional Occlusion Handling," in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp.82-89, 2011.

[12] C. B. Jin, S. Z. Li, T. D. Do, and H. Kim, "Real-Time Human Action Recognition Using CNN Over Temporal Images for Static Video Surveillance Cameras," in Advances in Multimedia Information Processing, Vol. 9315, pp. 330-339, 2015.

[13] Y. Xia, and F. Coenen, "Face Occlusion Detection Based on Multi-task Convolution Neural Network," in Proceedings of 12th International Conference on IEEE Fuzzy Systems and Knowledge Discovery (FSKD), pp.375-379, 2015.

[14] K. Kim, T. H. Chalidabhongse, D. Harwood, and L. Davis, "Real-time foreground-background segmentation using codebook model," in Real-Time Imaging, Vol. 11, no. 3 pp. 172–185, 2005.

[15] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), vol. 1, pp. 886 –893, 2005.

[16] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 9, pp. 1627–1645, 2010.

[17] M. Hirabayashi, S. Kato, M. Edahiro, K. Takeda, T. Kawano, and S. Mita, "GPU Implementations of Object Detection using HOG Features and Deformable Models," in Proceedings of the IEEE International Conference on Cyber-Physical Systems, Networks, and Applications, pages 106-111, 2013.

[18] M. Szkudlarek and M. Pietruszka, "Fast GPU and CPU computing for Head Position Estimation," in Proceedings of the Federated Conference on Computer Science and Information System, pp. 231-240, 2015. DOI: 10.15439/2015F410.

[19] D. Lee, J. Wang, K.N. Plataniotis, "Contribution of skin color cue in face detection applications," in: M.C. Emre, B. Smolka (Eds.), in Advances in Low-Level Color Image Processing, Springer, Netherlands, pp. 367–407, 2014.

[20] K. Nasrollahi and T. B. Moeslund, "Complete face logs for video sequences using quality face measures," in IET International Journal of Signal Processing, vol.3, no. 4, pp. 289–300, 2009.

[21] T. Ojala and M. Pietikainen, "Unsupervised Texture Segmentation Using Feature Distributions," in Pattern Recognition, vol. 32, pp. 477-486, 1999

[22] Z. Zhang, J. Iria, C. Brewster, and F. Ciravegna, "A comparative evaluation of term recognition algorithms," in Proceedings of the sixth international conference of Language Resources and Evaluation (LREC), pp. 2108-2113, 2008.

[23] Z. Zhang, H. Gunes and M. Piccardi, "Head detection for video surveillance based on categorical hair and skin colour models," in Proceedings of IEEE International Conference on Image Processing (ICIP), pp. 1137–1140, 2009.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 261–270
DOI: 10.15439/2016F541
ACSIS, Vol. 8. ISSN 2300-5963

Caption-guided patent image segmentation

Jerzy Sas, Urszula Markowska-Kaczmar
Wroclaw University of Technology
Faculty of Computer Science and Management
Wyb. Wyspianskiego 27, 50-370 Wroclaw, Poland
Email: jerzy.sas, urszula.markowska-kaczmar@pwr.edu.pl

Anastasia Moumtzidou
Information Technologies Institute,
Centre for Research and Technologies - Hellas,
6th km Charilaou - Thermi, 57001, Thessaloniki, Greece,
Email: moumtzid@iti.gr

***Abstract*—The paper presents a method of splitting patent drawings into subimages. For the image based patent retrieval and automatic document understanding it is required to use the individual subimages that are referenced in the text of a patent document. Our method utilizes the fact that subimages have their individual captions inscribed into the compound image. To find the approximate positions of subimages, first the specific captions are localized. Then subimages are found using the empirical rules concerning the relative positions of connected components to the subimage captions. These rules are based on the common sense observation that distances between connected components belonging to the same subimage are smaller than distances between connected components belonging to various subimages and that captions are located close to the corresponding subimages. Alternatively, the image segmentation can be defined as a specific optimization problem, that is aimed on maximizing the gaps between hypothetical subimages while preserving their relations to corresponding captions. The proposed segmentation method can be treated as the approximate solution of this problem.**

I. Introduction

BEFORE inventors prepare a patent application they should spend some hours doing a good search for patents that are related to their idea. Usually, searching for patents, they prepare phrases that in the best way describe the core concept. However, the list of results often contains hundreds or even thousands of patents depending on the popularity of the term or phrase selected. It also happens that one thing is described with different names and labels. Therefore, the results obtained are not always relevant.

Many specific tools exist that support patent databases searching ([4]), but in most cases they are mainly based on textual analysis. It is worth noticing however, that patent drawing is almost always required to illustrate the invention. Frequently, patents include numerous drawings showing a variety of views. Therefore, it seems that patent search results would be much more relevant, and it would be much easier to access the patent if a query considered illustrations included in patent documents. This observation leads the concept of content-based patent image search ([12], [13]). When applying content-based search paradigm, patent searchers are browsing thousands of patents looking only on images contained in drawings section. This task can be accelerated by using patent image search engines, which can retrieve images, based on their visual content. The importance of images in patent search can be further emphasized by the fact that images are both language independent and independent of the scientific terminology that may evolve over the years. They are also important in attempts to understand a patent.

Usually, a patent includes several figures or drawings showing how an invention looks. Drawings in patents can contain reference numerals that are used in the detailed description to identify parts of the drawings and to draw reader's attention, but they introduce difficulties in the case of image segmentation.

It happens that the invention is compound, and the patent document shows drawings of one or more parts. The drawings in patents are specific and differ much from other illustration in other types of documents.

A variety of styles can be encountered in patent images, e.g., surface shading, plots, pattern area fills, broken lines and varying line thickness. However, in most cases, patents include some views prepared as black-and-white line art. The drawings can be made using different techniques. Sometimes, CAD tools are used, but there are also numerous old patents with drawings prepared manually with ink. In many cases, component subimages are not separated by distinct wide areas of background, so naive approach based merely on the segmentation by wide background bands fails. The example of a compound patent image consisting of many subimages is presented in Fig.1.

Thus, to develop an image content based patent search engine, it is necessary to employ techniques to identify the number and the position of the figures on the page to isolate them. To this end, an efficient segmentation of the page in its figures is required. Its accuracy will determine to a large degree the performance of image patent search process. Our research is focused on drawings segmentation from patent documents.

All these mentioned above features cause that images and subimages segmentation in patents is a challenging task and methods developed in other areas, e.g. for segmentation of color or gray shade images included in journals are not appropriate in the application are being considered here.

The paper is organized as follows. Section II contains a short review of other works related to compound image segmentation into parts. In the next two sections, the segmentation

This work was supported by the statutory funds of the Department of Computational Intelligence, Faculty of Computer Science and Managament, Wroclaw University of Science and Technology and partially by EC under FP7, Coordination and Support Action, Grant Agreement Number 316097, ENGINE European Research Centre of Network Intelligence for Innovation Enhancement (http://engine.pwr.edu.pl/). All computer experiments were carried out using computer equipment sponsored by ENGINE project.

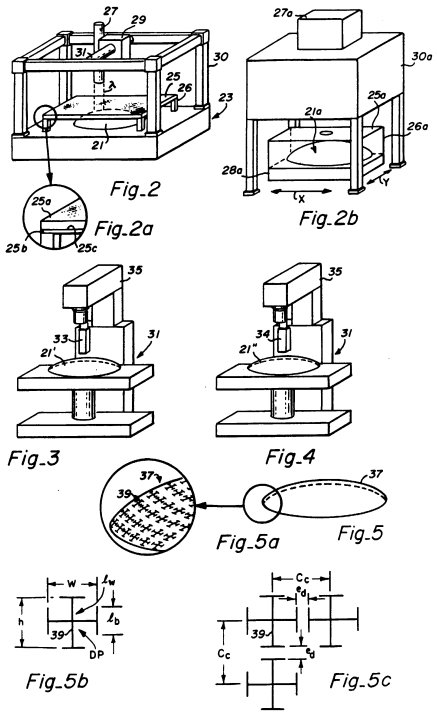


Fig. 1. Example of patent image consisting of many captioned subimages

problem is defined, first intuitively and then more formally. Sections V-A and V-B describe two possible algorithms of segmentation that we tested. In Section VI the specific method that can be applied to simple grid-layout images segmentation is presented. The method of caption detection used here, based on text extraction with OCR support is shortly described in Section VII. The method of automatic evaluation of human-defined and automatic segmentation consistency is explained in Section VIII. Experimental results obtained with proposed segmentation methods on a patent images benchmark database are presented in Section IX. Finally, some conclusions and suggestions for further works are formulated in Section X.

II. Related Work

The idea of searching documents on the basis of figures included in documents is not new. For instance, Lu et al. in [7] proposes to utilize the content of figures in searching scientific literature in digital libraries. The work described in this paper is only focused on the categorization of figures included in scientific documents. Authors developed a machine learning based method for document image categorization. A figure is categorized as a photograph or a non-photograph. A non-photograph is then further classified as plot, diagram, or another graph. The aim of a method presented in [6] is similar - image classification using machine learning methods. The authors also propose numerical data extraction from pictures.

Although the idea of figure content based document retrieval exists since many years, there are still many challenging tasks to implement. The problems attract the attention of researchers. It is important to detect figures in documents and separate them from a text. The exemplary approaches can be found in [9], [3], [1], [10]. Some of the figures are composed of subfigures. Their separation is also a challenge [2]. The next problem is a classification of figures to the defined groups. Research devoted to image content understanding is a rapidly developing area.

Considering the area of our research in this short survey, we have particularly focused on image and subimage extraction (segmentation). The paper [2] presents a technique of compound image separation. It is based on systematic detection and analysis of uniform space gaps. The method assumes the separation of subimages by thin horizontal or vertical uniform space that separate compound figures from a major part of compound figure images.

The paper [5] describes a solution to the similar problem but located in biomedical literature. Articles in this area are often composed of multiple subfigures and may illustrate diverse methodologies or results. This solution is similar to our approach in that it also is based on capture recognition. Their method first analyzes figure captions to identify the label style used to mark panels, then determines the panel layout, and finally, each figure partition into panels is performed. To identify the number of panels, authors developed three stage procedure. First, a simple lexical analysis is made to determine the potential panel labels in the captions. Then, the list of potential panel labels is analyzed in order to identify and remove false positive ones. Finally, the segmentation of captions according to the set of identified panels is carried out. In the next step, the image processing is executed. First, the optimum threshold value for segmenting the figure is computed. The text embedded in images is also detected in order to find panel labels. Next, the number of panels in the figure is determined. Then panels are partitioned it into a set of panel-subcaption pairs, on the basis of the set of subcaptions, panel labels, and connected components. To partition the figure, they first create a node for each recovered panel label and then create an edge connecting each node to its closest horizontal and vertical neighbors. It means that the method assumes the specific horizontal or vertical relative position of subimages.

Another method of figure classification and subimage extraction is that described in [14]. It also refers to the biomedical area. The method of subimage extraction retrieves subimages using reconstruction from Hough peaks.

Comparing to the presented methods, in the case of patent documents, subdrawings rarely have easy to detect vertical or horizontal separating spaces. Therefore it is a much challenging task.

III. Problem formulation

Let us consider a binary black and white image. Black pixels are assumed to represent drawn lines, captions and inscribed text (foreground), while white pixels constitute background. The image consists of subimages, where each subimage is associated with its individual text caption. We assume that

captions can be reliably detected in earlier stages of the image segmentation procedure. The applied method of caption detection is described shortly in section VII. So, at the current stage, we know the number of subimages, which is equal to the number of detected captions. If no caption is detected, then it can be assumed that the whole image constitutes the single component, unless there are sufficiently wide background areas separating clouds of foreground pixels. In the later case, one of segmentation methods based merely on separation by background (described in the previous section) can be applied. We will not deal with such cases in this paper.

We are considering here only such images where the set of detected captions is not empty. Our aim is to subdivide the set of all foreground pixels into disjoint subsets constituting subimages associated with captions in such a way, that it corresponds to the image author intent and to the intuition of a human observing and interpreting the image.

Unfortunately, in practice, there are no strict rules followed when drawing compound images, so depending on the degree of the image complexity, shape of components and the logical (semantic) relation between them, the subimages may be arranged on the image plane quite freely. For this reason, it is hardly possible to define the formal segmentation principle, so that it always corresponds to the intent of the image creator. Nevertheless, usually, some basic principles are applied when arranging the layout of the compound image. The intuitive rules typically used are as follows:

- subimages are separated by relatively wide areas of background,
- in most cases, subimages are not connected by foreground elements; if it is not the case, only few simple lines connect subimages (the example can be a single line connection between subimages captioned *"Fig.2"* and *"Fig.2a"* or *"Fig.5"* and *"Fig.5a"* in the drawing in Fig.1,
- each subimage contains at least one "dominant" element consisting of foreground pixel which size is comparable or greater than the size of the caption,
- captions are close to elements of the subimage associated with them,
- in certain cases, the subimages are regularly arranged in a grid-like manner, where subimages are clearly separated by horizontal and vertical bands of foreground pixels, which often include subimage captions (an example of such layout is presented in Fig.2.

Images complying the latter principle can be segmented very easily and also such a regular layout is easily distinguishable. The method proposed here first tries to detect whether or not the image being analyzed a case of a regular layout. Regular images are segmented with a specific (easy) method and excluded from further considerations. For other images, we apply the segmentation procedures based on remaining intuitive rules.

In the approach described in this paper, we follow these informal rules to build the clustering method, which groups foreground elements of the image into subsets associated with individual captions. The method of image segmentation

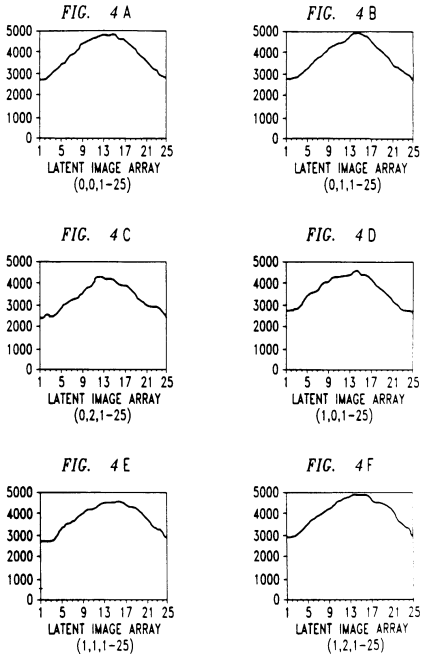


Fig. 2. Example of the regular grid-like layout of subimages

described here; clusters connected components of foreground pixels into sets representing captioned subimages. It means that, in most cases, each subimage consists of entire connected components. The only exception from this rule is where there exists a substantial evidence that some subimages share commonly connected components. In such situation, some connected components are split. We will start with defining the notion of connected components more formally, and then some proposals of connected components clustering will be described.

IV. Finding Subimages by Connected Components Clustering

Let us define the connected component (*CC*) of the image as the set of foreground pixels that are linked by other foreground pixels belonging to the same CC. A pixel p will be here represented by the pair of its coordinates:

$$p_i = (x_i, y_i), x_i \in \{0, ..., x_{res}-1\}, y_i \in \{0, ..., y_{res}-1\}, \quad (1)$$

where x_{res}, y_{res} denote the image horizontal and vertical resolution. Let $\mathcal{F}$ denotes here the set of all foreground pixels in the image being segmented. We divide the set $\mathcal{F}$ into the set of disjoint CCs:

$$C = \{c_1, ..., c_M\}, c_i \subseteq \mathcal{F}, c_i \cap c_j = \emptyset, \bigcup_{i=1}^{M} c_i = \mathcal{F}. \quad (2)$$

The connected component c is the set of pixels such that if two pixels $p_a, p_b \in c$ then there exists the sequence of pixels $(p_a = p_1, p_2, ..., p_k = p_b)$, all of them belonging to c that

for each pair of pixels $p_i, p_{i+1}, i = 1, ..., k-1$, pixels p_i and p_{i+1} are direct neighbors. We consider two pixels to be direct neighbors if they are *8-connected*, i.e. they have common edge or corner. A CC is maximal if no more foreground pixels can be attached to it. Further on, we will be considering *maximal CCs* only. We define the distance $d(c_i, c_j)$between two CCs c_i and c_j as:

$$d(c_i, c_j) = \underset{p_l \in c_i, p_k \in c_j}{min}(\quad e(p_l, p_k) \quad), \tag{3}$$

where $e(p_l, p_k)$ denotes Euclidean distance between pixels in 2D. Similarly, we can define the distance between two sets (clusters) s_i, s_j of CCs:

$$d(s_i, s_j) = \underset{c_l \in s_i, c_k \in s_j}{min}(\quad d(c_l, c_k) \quad). \tag{4}$$

Our aim is to split the whole binary image into a set of subimages S_A, where each subimage $s_i \in S_A$ is the set of CCs (we will call it *cluster*), $s_i \subseteq C$. We also assume that clusters are disjoint and they all sum up to the whole image, i.e. each foreground pixel from $\mathcal{F}$ belongs to exactly one cluster (or in other words - subimage). The partitioning of the image into clusters should be as close as possible to the intent of the image creator. Initially, we make the assumption that each CC belongs entirely to a single subimage, which may be not true in some cases (as shown in Fig.1). We will deal with such cases later and will propose a method of connected component splitting. By taking into account the set of intuitive rules given in the previous subsection we assume that: a) subimages consist of graphical elements (corresponding to CCs) that are located close each to other within the single subimage and b) disjoint subimages are separated with relatively large bands of background (white) pixels. Additionally, we assume that each subimage is associated with its individual caption. We assume that captions are reliably detected in earlier stages of the image segmentation procedure. So, at the current stage we know the number of subimages, which is equal to the number of detected captions. It seems also reasonable to make assumption, that caption areas are located close to the corresponding subimage. The problem is therefore how to split connected components set into disjoint subsets such that the obtained partitioning corresponds to subimages, as a human perceives it.

Formally, the presented intuitive assumptions correspond to finding such clustering of connected components, where the number of clusters is fixed (and is equal to the number of detected captions k) and the minimal distance between clusters of CCs is maximized. The graphical elements constituting captions (characters and their graphical elements) are being considered here as ordinary foreground image elements. We assume that each caption (as a graphical element) is entirely included in a single cluster and each cluster includes exactly one caption. The elements of a caption are being considered as a single indivisible CC (even though actually a caption consists of many "true" CCs). Let Γ denotes the set of all possible partitionings of the set C into k clusters $\{s_1, s_2, ..., s_k\}, s_i \subseteq C$, which satisfy the above restriction. By s we denote here the cluster of CCs. We need to find such "optimal" clustering $\gamma^* \in \Gamma$ that:

$$\gamma^* = \arg\max_{\gamma \in \Gamma}(\underset{s_i, s_j \in \gamma}{min} d(s_i, s_j)), \tag{5}$$

where $d(s_i, s_j)$ is the distance between clusters defined in equation 4. Because the number of possible partitioning in Γ is very big if the number of CCs is high (equal to the Stirling number of the second kind that determines the number of possible partitioning of n-element set into k nonempty subsets) the problem is computationally hard and a suboptimal solution must be applied.

A. Connected component splitting

In certain subimages, the single large CC may span many subimages as shown in Fig. 1, where subimages 2 and 2a or 5 and 5a share a common CC. Any method based on complete CCs clustering cannot retrieve the correct segmentation in cases like this. The selected CCs need to be split into smaller ones, so that the principle of composing subimages from complete CCs can be still used. The method applied there consist of: first, carying out the segmentation procedure using the original CCs set, detecting CCs that possibly need to be split, splitting them into smaller CCs and executing the clustering procedure again. In the second clustering, the new set of CCs is used, where original large ones are replaced by their parts.

1) Detection of CCs that need to be split: The CCs that are "suspect" of spanning between adjacent subimages are detected by examining the position of sufficiently large CCs, with relation to the nearest captions. The detection is performed after initial clustering with original CCs. In this way we can consider only these captions that are close enough to obtained clusters. If two captions are close to a large CC then, instead of having only one cluster containing this large CC, we should split the large CC and allocate its parts to two smaller clusters labeled by captions close to "suspect" large CC. CC splitting seems especially adequate if two candidate captions are closer to this CC than to other clusters.

Let $l(C)$ denotes the caption l assigned to the cluster C. By L_C we denote the set of captions being candidates for captions assigned to subimages obtained by possible split of the cluster C into smaller ones. The set L_C consists of: a) the caption $l(C)$ assigned to C by the primary clustering carried out using the original CCs found in the image and b) other captions, which distance to C is comparable to the distance between $C \backslash \{l(C)\}$ and $l(C)$. Only these captions are included in L_C which are not "strongly bound" to other clusters. We assume here that $l(C')$ is strongly bound to C' if its distance to C' is much lower than the distance of any other caption to C'. The threshold of distances used for classifying the caption as "strongly bound" can be determined experimentally using the set of validation images.

The captions collected in L_C are then used to detect sufficiently large CCs $c \in C$ that possibly span many actual subimages. For simplicity, let us consider the pair of captions

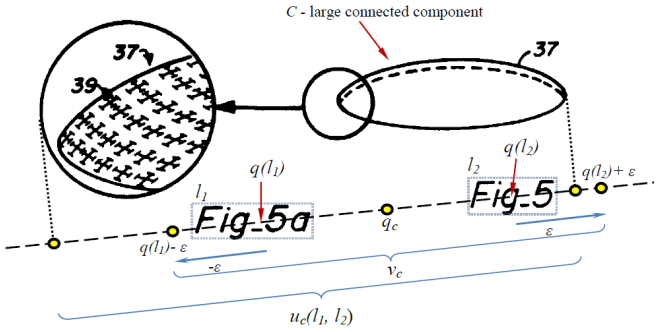


Fig. 3. Evaluation of a CC as a candidate for split with respect to two captions

$l_1, l_2 \in L(C)$. They define the line containing the centroids of captions $q(l_1), q(l_2)$ and the line segment with end points at $q(l_1)$ and $q(l_2)$. Each CC $c \in C$ can be projected onto the line $(q(l_1), q(l_2))$. Because, by definition, each CC is compact then its projection onto the line also constitutes the compact line segment $u_c(l_1, l_2)$. We consider the CC $c \in C$ to be a candidate for split if $u_c(l_1, l_2)$ covers the center of the line segment $q_c = (q(l_1), q(l_2))$ and it is sufficiently long. In our experiment we assumed the minimal length of $u_c(l_1, l_2)$ to be at least $0.75 * \|q(l_1) - q(l_2)\|$. In this way, for a pair of captions in L_C we can selected some CCs in C that are candidates for splitting. If there are more than two elements in L_C then they can be ordered into the sequence $(l_1, l_2, ..., l_m)$ so that the path $(q(l_1), q(l_2), ..., q(l_m))$ is the shortest. Hence we obtain pairs of captions that are close each to another and possibly are assigned to adjacent subimages. Then for each pair of captions $(l_i, l_{i+1}), i = 1, ..., m-1$ adjacent in the ordered sequence, the described above procedure is carried out resulting is some CCs as candidates for splitting. The idea of finding CCs for splitting in presented in Fig. 3.

2) Finding CC split point: The next element of the proposed procedure is finding the split point for a CC that has been selected as a candidate for split with respect to two captions l_1 and l_2. We want to split it into two parts which are close to captions l_1, l_2 and so that the minimal number of connections between resultant components exists. The later assumption follows from the observation that in cases where the large CC needs to be split, it consists of fragments connected by a few lines, most often - by just a single one. We applied the method where the boundary between components of a CC is the straight line which satisfies the following conditions:

- it is perpendicular to the line defined by $(q(l_1), q(l_2))$,
- it crosses the extended line segment $v_c = (q(l_1) - \epsilon, q(l_2) - \epsilon)$, where $\epsilon = (q(l_2) - q(l_1))/4$,
- it minimizes the number of split lines of the skeletonized image of the CC being considered; if there is more than one position of the split line where this minimum is reached then the position closest to the center of the line segment $(q(l_1), q(l_2))$ is selected.

If the described above CC splitting procedure creates at least one subdivided CC then clustering is being carried out again using the set of modified CCs.

V. Image Segmentation by Connected Components Clustering

We propose two methods to find suboptimal solution of the problem defined in eq. 5. The first one is based on human intuition that binds subimage elements to the close captions and initially focuses attention on big graphical elements. The second method is typical k-means algorithm application to the set of CCs. The only introduced constraint is that the components corresponding to captions cannot be assigned to the same cluster.

A. CCs clustering by human perception imitation

In this method we follow the human way of reasoning when dividing the image into subimages. Later on, we will call it *intuitive segmentation*. Although precise way of human reasoning is unknown, it seems that when splitting an image into subimages, humans proceed as follows. Initially, we seek for big graphical components that are located close to captions. They become cores of further activities. If a smaller element is completely surrounded by closed shapes of big components already associated with a caption, a human tends to assign it to the same caption. All remaining smaller elements are assigned to these already constituted subimages to which the distance is smallest. The detailed procedure that can be applied is defined as Algorithm 1.

B. CCs clustering by k-means

k-means algorithm is a widely known and successfully applied method of clustering, so we will not describe it here in details. Its primary description can be found in [8]. In order to apply it to our problem, we need to specify how the initial clustering is obtained and how centers of clusters are determined. Initial clustering is obtained creating k clusters that initially contain CCs representing captions. Remaining CCs are assigned to initial clusters so that the including cluster corresponds to the caption closest to a CC. We mean here the distance computed as in equation 3 where in place of c_j the single pixel is used that is the center of the caption bounding box. The core of k-means algorithm is the computation of cluster means and computing the distance of elements being clustered to cluster means. Here we are clustering CCs (not individual foreground pixels). Therefore two methods of cluster center positions computing can be proposed:

- a cluster center is the centroid of all equally weighted foreground pixels belonging to CCs in the cluster,
- a cluster center is the weighted mean of bounding box centers that enclose CCs belonging to the cluster, the component weights can be based either on the size (maximal length along x and y, or the bounding box area) or on the number of pixels belonging to a CC.

Both variants were evaluated experimentally and the better one is finally recommended. Details are presented in Sec.IX.

Algorithm 1 "Intuitive" CC clustering

Require: C - set of all connected components; L - set of captions; ($L \subseteq C$)

sort CCs by size in decreased order;
create the family G of initially empty sets of CCs assigned to captions;
while empty sets in G exist **and** there exist unassigned CCs **do**
  get the first biggest CC that is not assigned to any caption;
  assign it to the set in G corresponding to the closest caption;
end while
if there exist empty sets in G **then**
  for all captions with no CCs assigned **do**
    find CC c closest to unassigned caption which is not the only element of the family in G that contains it;
    remove c from the set in G that currently contains it;
    assign c to the current caption;
  end for
end if

/* Now we have big CCs assigned to captions */

make initial clusters of CCs assigned to labels so far;
while not all CCs are assigned to clusters **do**
  get the next unassigned CC;
  find the cluster closest to it;
  assign CC to the closest cluster;
end while

return (G) - the family of CC sets assigned to individual labels

Algorithm 2 "CC clustering with modified k-means"

Require: C - set of all connected components; L - set of captions; ($L \subseteq C$)

$G \leftarrow$ the family k initial clusters containing individual captions from L;
compute the centroid position for each cluster in G;
assign all CCs in $(C \setminus L)$ to clusters in G using minimal distance to the centroid as a criterion;
$G' \leftarrow G$;
repeat
  $G \leftarrow G'$;
  compute the centroid position for complete clusters in G;
  $G' \leftarrow \{g_i : i = 1, ..., k; g_i = \emptyset\}$;
  for all $c \in C \setminus L$ **do**
    $g^* \leftarrow \underset{g \in G}{argmin}(X(g), c)$;
    $I(g^*|G, G') \leftarrow I(g^*|G, G') \cup \{c\}$;
  end for

/* Now we have all non-caption components initially clustered */

sort clusters in G' in descending order of their sizes;
$L' = L$;
for all $g \in G$ in decreasing order of size **do**
  $l^* \leftarrow \underset{l \in L'}{argmin}(d(g, l))$;
  $g \leftarrow g \cup l^*$;
  $L' \leftarrow L' \setminus \{l^*\}$;
end for
until $G' \leftarrow G$ - no change in clustering
return (G') - the family of CC sets assigned to individual labels

In order to preserve the restriction that each cluster contains exactly one CC representing a caption, the typical k-means algorithm must be modified. In the initial partitioning, this restriction is obviously satisfied, provided that it is established according to the recipe described above. In the k-means algorithm phase, where cluster centroids are computed, all CCs assigned to a cluster in the previous iteration are taken into account. However, when the new clusters are being build in the next iteration only CCs not being elements of captions are a subject to be moved to the cluster with nearest centroid. In this way, new clusters are constituted, which do not contain CCs representing clusters. The caption CCs are then uniquely assigned to these clusters. The assignment is achieved by considering created clusters in decreasing order of its size (biggest first) and assigning such already unassigned cluster to the current cluster for which the distance to a cluster is minimal. Finally, centroids of new clusters that include captions are computed and the next k-means iteration starts. The modified algorithm is presented in Algorithm 2. In the algorithm the following symbols are used: $d(\alpha, \beta)$ is the distance between elements α and β, L is the set of CCs that are graphical elements constituting captions (each caption is treated as a single CC), $X(g)$ is the centroid of the cluster g. Let G and G' denote two families of CC sets, such that in both G and G, each caption $l \in L$ belongs exactly to a single set in G and G'. $I(g|G, G'), g \in G$ denotes the element from G' that contains the same caption $l \in L$ as g.

VI. Segmentation of grid-like composed images

In certain cases subimages are located in a compound image, so that they create grid-like layout as presented in Fig. 2. Correct segmentation is quite easy in such cases and it can be carried out using a simplified method. Sometimes, general methods described in preceding sections fail in grid-like layouts, so it seems reasonable to apply specific method to this class of images which very often leads to the segmentation consistent with an image creator intent.

By grid-layout of subimages we mean here such a placement of subimages where

- captions can be enclosed by narrow horizontal bands spanning from left to right edge of the image, which do not contain significant elements of subimages and all CCs located in caption strips can be separated by the path consisting only of background pixels that connects left and right edge of the image;
- the subimages are consistently located either under or above their captions - it means that there are no significant CCs either below the lowest caption band or above the highest caption band. In result, horizontal areas between caption bands can be uniquely assigned to their caption bands;
- if there are more than one captions in a caption band then all CCs in the corresponding image band can be separated into as many clusters as the number of captions in the caption band. The separation areas are vertical areas consisting only of background pixels.

The above conditions can be easily tested programmatically. The testing procedure immediately provides the segmentation of the image, which appears to be very reliable. In our approach, each image being segmented is subject to grid layout test before other segmentation methods are tried. If the test passes then final segmentation is a byproduct of the grid layout testing procedure. If the grid layout test fails then the image is passed to general segmentation procedure based on concepts presented in preceding sections.

VII. Caption localization

Text extraction methods are used in order to find subimage captions. In this work we used the method described in [11]. The procedure consists of two phases. In the first phase, some rectangular areas are detected which are likely to contain (any) text inscribed to the image. In the second phase, areas recognized as containing text are tested for occurrence of specific text patterns being the actual captions.

1) Detection of text areas in the image: The first phase consists in turn of three stages. In the first stage, the procedure finds candidate areas that possibly contain texts, by grouping small CCs into rectangular areas of shapes and sizes typical for areas including individual words of text inscribed into the image. In the next stage, candidate areas are passed through the classification procedure which classifies them as "text" or "non-text" using the set of features related to distribution of foreground pixels and line segments within the candidate area. Areas recognized as "text" are passed to OCR module running in no-dictionary mode. Finally, only such candidate areas are considered to be text areas, for which the results of OCR recognition satisfy an empirical criterion related to the ratio of untypical characters occurrence in the recognized textual string.

Initially, connected components (CCs) that possibly enclose individual characters or their sequences are found. The series of criteria were experimentally elaborated that must be satisfied in order to reject CCs that are not likely to contain text characters. All used criteria are discussed in [11]. One of the most important criteria appeared to be the ratio $f_i^{(hv)} = h_i/v_i$ of the height of the shape contained in CC h_i to the average line width v_i in the i-th CC. If the ratio value is out of the interval covering the range typical for text areas in images, the test fails and the area is not further considered as a candidate area. The acceptance interval is determined by considering text areas appearing in the validation set of patent images, for which "true" text areas were manually annotated. By analyzing the set of manually confirmed text areas in the validation set, the histogram of $f_i^{(hv)}$ has been created. As the acceptable range of $f_i^{(hv)}$ we assumed the interval covering 98% of values encountered in the validation set, leaving aside 1% of very small and 1%of very high text areas as outliers. Other parameters of criteria used in candidate text areas selection were established in the similar way.

The candidate areas that passed two first stages of text extraction are finally subject to OCR that runs in no-dictionary mode. The result of OCR recognition is first used as the data for the last stage of text area detection. It has been observed that in the case of "false" candidates that contain no text, the string created by OCR includes many punctuation characters (like . , ; : -). The final criterion of a candidate area acceptance as the text area (no necessarily the area of caption) is the ratio of punctuation characters occurrence in the whole recognized string. Details are described in [11].

2) Selection of text areas containing caption patterns: If the OCR module were absolutely accurate then caption detection would be a trivial problem. It would be possible to simply compare the OCR results to one of predefined text strings that can be a caption. However, in patent images, inscribed textual elements are hard to recognize. In most cases, the main reason is that images are hand-drawn, so is the included text. While OCR works well with machine printed text, its performance seriously deteriorates when it is presented with handwriting or hand-printed text. Therefore, it seems reasonable to consider as captions not only areas where OCR produced exact caption patterns, but also to accept such ones, where the recognized sequence is in some sense similar to one of acceptable caption patterns.

The idea of recognizing the text area as a caption is based on finding in the sequence recognized by the OCR module, the subsequence that is similar to one of *caption patterns*: "FIG" "fig" "Fig". The visual similarity of characters in the recognized subsequence to the corresponding characters in these patters is taken into account. It may happen that the recognized sequence is a correct word containing the pattern as a subsequence (e.g. "conFIGuration", "FIGhter"). In order to exclude such texts from considerations, first the result of recognition is compared with the list of correct words containing the pattern as a subsequence. Only words longer that 4 characters are used. If the length of the sequence recognized by OCR module is longer than 4 characters, then the minimal edit distance to words in the dictionary is found with Levenshtein algorithm. If the edit distance is less than one fourth of the number of characters in the closest word, then the label is rejected as rather being the part of ordinary

word appearing in the image.

If the OCR recognition result was not rejected by the dictionary test then the similarity to caption patterns is computed. The procedure finds a three-character substring in the recognized sequence that is most similar to patterns. When computing the similarity, we multiply similarity factors for the individual characters in a pattern. The similarity measures should be specific to properties of used OCR module and used fonts. The similarity factors can be computed as a relative frequency of errors consisting in replacing the actual character c_a by another erroneous character c_r: $n(c_r|c_a)/n(c_a)$. $n(c_r|c_a)$ is the count of errors consisting in replacing c_a by c_r and $n(c_a)$ is the number of c_a occurrences. In our experiments we however used similarity factors based on visual intuitive similarity of various character shapes. We assumed than nonzero similarities are given only to the following sets of characters recognized by OCR:

- for F : F E f P
- for I : I l i 1 J L T
- for G : G C 6 c
- for f : f t
- for i : i j 1 l
- for g : g 9

For remaining characters we assume that the similarity is equal 0.0.

Let us consider the character sequence $(c_1, c_2, ..., c_m)$ returned by OCR module running in no-dictionary mode. For each subsequence consisting of three consecutive characters $s^{(3)}(j) = (c_j, c_{j+1}, c_{j+2})$, the similarity to the three-characters pattern sequence f is defined as:

$$\Delta(s, f) = \prod_{i=1}^{3} \delta(s_i, f_i), \tag{6}$$

where s_i, f_i is the i-th character in the sequence, $\delta(a, b)$ is the similarity between characters a and b. Here we assume that $0.0 \leq \delta(a, b) \leq 1.0$ and $\delta(a, a) > \delta(a, b)$ for all $a, b; a \neq b$.

The final likelihood that the sequence s represents a caption is computed as:

$$Q(s) = \max_{j=1,...,m-2} \max_{f \in \Phi} \left(\Delta(s^{(3)}(j), f)\right), \tag{7}$$

where $\Phi = \{"fig", "Fig", "FIG"\}$ is the set of caption patterns. In order to make the final decision whether or not the text area passed to OCR is the caption, the computed similarity is compared to a threshold. The lower is the threshold value, the more actual captions are recognized but also the likelihood to accept "false" captions increases. On the other hand - the higher it is, the higher is the likelihood that an actual caption will be omitted. The threshold value is determined as a metaparameter. The threshold should be set depending on he losses following form two types of caption recognition errors (i.e. rejection of true caption and false acceptance of non-caption text as a caption). The threshold value was fixed using the validation set in images containing captions and non-caption texts. Details are described in Section IX.

VIII. Automatic segmentation evaluation

In order to evaluate the accuracy of the automatic segmentation, the results obtained automatically need to be compared with "ground truth". By ground truth we mean here the results of manual segmentation of images in the test set, done by humans We assume that each manually defined segment has at most one automatically created counterpart that most closely matches it. Each automatic segment can be then used at most once as a counterpart of a certain manual segment. Some automatic segments may remain unassigned to any manual segment, as well as some manual segments may have no corresponding automatic segments. More formally, let us denote the set of manual and automatic segments by S_M and S_A. At the first stage, we define the function $f(s) : S_M \rightarrow S_A + \{\phi\}$, where ϕ denotes here "no automatic segment matches the segment s". The function $f(s)$ can be constructed in such way that maximizes the matching measure between elements of S_M and S_A. In the case of small number of elements in these sets, the exhaustive search that tries all possible mappings can be applied. In other cases, suboptimal procedures of $f(s)$ construction must be applied.

Having f function defined, we can evaluate the matching between manual and automatic segmentation by evaluating the matching defined by $f(s)$ for each segment $s \in S_M$. Popular $F1$ measure is used to evaluate the matching. Let for some $s \in S_M$ we have $a \in S_A \cup \{\phi\}$. We treat here s and a as sets of foreground pixels. If $a = \phi$ then a is assumed to be the empty set. The pixels in s are "relevant" elements, while the pixels in a are "retrieved" elements. The precision and recall can be then computed as

$$prec = \frac{\mid s \cap a \mid}{\mid a \mid} \tag{8}$$

$$rec = \frac{\mid s \cap a \mid}{\mid s \mid} \tag{9}$$

$$F1 = 2\frac{prec * rec}{prec + rec} \tag{10}$$

In the case when $\mid a \mid = 0$ (no automatic subimage is assigned to the manual one) we assume that $F1 = 0$. The $F1(s)$ score is computed for each $s \in S_M$. The overall automatic segmentation accuracy for the whole image is calculated as:

$$F1_T = \frac{\sum_{s \in S_M} \mid s \mid * F1(s)}{\sum_{s \in S_M} \mid s \mid}. \tag{11}$$

Thus, bigger subimages are assigned higher weights than smaller ones in the overall evaluation. In order to compute the assessment for the collection of images, the measures $F1_T$ computed for individual images are averaged.

IX. Experimental results

For the sake of drawings segmentation testing we used the subset of images collected in *PATExpert* project conducted by ITI CERTH institute and available at http://mklab.iti.gr/project/patentbase web page. The image set is described in [13]. Images in the database are scanned images of manually

created drawings. Only a few of them seem to be created with drawing software. Also in these cases they were printed and the image was finally acquired by scanning. Near-duplicates (i.e. identical or almost identical images) included in the original database were excluded.

For experiment purposes, 1461 images containing at least one caption were selected from *PATExpert* database. Each selected image was manually segmented into subimages and caption areas were marked. In such a way, we obtained "ground truth" information about the actual segmentation and localization of captions. The set of images was divided into two parts: a) the subset of images containing exactly one caption (1135 images) and b) the set of images containing multiple captions (296 images). 600 images from the part a) were assigned to the validation set. It was used for caption detection metaparatemers tuning. Remaining images from the set a) and images from the set b) were used as the testing set.

The experiment carried out consists of two phases. In the first phase we evaluated the accuracy of caption area recognition. The second phase was aimed on the assessment of the automatic segmentation consistency with the manual segmentation.

A. Evaluation of caption recognition accuracy

The aim of this experiment was to set the metaparameters used in the caption detection algorithm and then to estimate the performance of the proposed method. Tesseract OCR module with its default character set for English was used for extracted text recognition.

The validation set consisting of 600 images from the part a) was used for tuning the text area detection algorithm as described in [11]. It was also used to set the likelihood threshold $Q(s)$ defined by equation 7. The threshold value was set as the maximal value that accepts 98% of all true captions appearing in text areas in the validation set.

The test set for caption detection consisted of 535 remaining images from the part a) and all 296 images from the part b). They included total 1322 subimages with captions.

The caption recognition errors occurred in 168 of 1322 caption occurrences, so the caption detection error rate was 12.71%. The error caused by insertion of false captions occurred only in 5 images. In two cases, captions were erroneously recognized where actually they were parts of longer words inscribed in the image. Only three false captions were recognized in text candidate field that in fact contained graphic elements. 163 errors out of the whole number 168 were caption omissions. The reasons of all observed caption errors are summarized in the Table I.

It is evident that the most frequent reason of caption omissions is related to handwriting and unusual fonts. In future, some of this errors can be avoided by providing OCR module with samples of character shapes used in patent images. Validation set can be used for this purpose. Another possibility is to use OCR recognizer aimed more on handwriting. Tesseract program used in our experiments is trained on machine fonts, so it performs poorly in case of handprinted texts.

B. Automatic segmentation assessment

For the sake of automatic segmentation, only these elements of the test set were used, for which all captions were recognized correctly. 267 images from the part b) passed caption detection test successfully and they were used as the test set in the next experiment.

First, the images of grid-like structure were selected using the procedure described in Section VI. The procedure is purely technical and very accurately selects grid-like layout. The method detected 152 grid-like subimages. The segmentation defined by caption bands corresponded exactly to the actual segmentation in 150 images from this group. Only in two images, small and probably well-tolerated inaccuracies occurred. They were caused by existence of subcaptions appearing on the opposite side of the caption than the side including the related subimage. The proposed procedure for grid-like layouts detection and segmentation based on grids can be therefore assessed as very accurate and useful in many cases.

Remaining 115 images from the test set were segmented by intuitive segmentation described and by the modified k-means algorithm described in Sections V-A and V-B. For methods evaluation we used F1-score as explained in Section VIII. We observed that the F1-measure computed in this way is related to the degree of inaccuracy of "true" and automatic subimage matching. The intervals of F1 values roughly correspond to the following inaccuracy levels:

- $F1 \in (0.99, 1.0 >$ - subimages match almost perfectly, the differences are not meaningful and concern only individual pixels and dot-like graphical elements;
- $F1 \in (0.97, 0.99 >$ - differences in matching images concern usually misplaced very small graphical elements, in most cases of minor importance for image understanding;
- $F1 \in (0.95, 0.97 >$ - small elements like individual characters or digits appearing in descriptions, pointing arrows, distant and not graphically connected elements are displaced; usually it does not impair the concept presented in a image;
- $F1 \in (0.85, 0.95 >$ - relatively big elements of subimages are misplaced, but the presented concept is usually still readable;
- $F1 \leq 0.85$ - major elements of subimages are misplaced or missing (i.e. - they remain unassigned to subimage), the presented concept is not readable.

TABLE I
CAPTION DETECTION ERROR REASONS

Error reason	Number of occurrences
Strange but repeatable font	35
Handwritten captions	66
Oversegmentation of text fields	43
Caption patterns in ordinary words	2
Untypically rotated image	7
Falsely inserted captions	3
Unexplained	12
TOTAL	**168**

TABLE II
$F1$ SCORE DISTRIBUTION FOR TWO PROPOSED SEGMENTATION PROCEDURES

$F1$ **interval**	**Fraction of subimages**	
	intuitive segmentation	k-means segmentation
$F1 \in (0.99, 1.0>$	72.65	66.85
$F1 \in (0.97, 0.99>$	11.87	10.22
$F1 \in (0.95, 0.97>$	3.59	4.14
$F1 \in (0.85, 0.95>$	5.52	6.91
$F1 \leq 0.85$	6.35	11.88

Taking the above observations into account, we can assume that the segmentation fails if for at least one of subimages in the image, its $F1$ score is less than 0.95.

In order to evaluate the accuracy of automatic segmentation procedures and to compare them, $F1$ scores were evaluated for all automatically created subimages assigned to their "true" counterparts. Table II shows fractions of all tested subimages falling into $F1$ score intervals described above.

Overall performance of k-means segmentation is lower than the performance of intuitive segmentation. In the test, intuitive segmentation provided usable results for 88.13% of images, while k-means segmentation gave good results only in 81.20%. The main weakness of k-means based segmentation lies in its tendency to create unbalanced subimages, where one automatically determined subimage contains numerous big CCs, while others consist only of small components. In many cases it leads to segments that contain CCs actually belonging to various subimages. One of reasons of such situations is frequent appearance of empty clusters created in the course of the algorithm execution. The correction consists of forced assignment of some CCs to the empty cluster. The implemented cure consists in selection of the CC which is closest to the caption not assigned to any nonempty cluster. Sometimes it is a small CC, what can lead to the phenomena of unbalanced clusters.

The results of k-means segmentation is usually close to the human intent in the case of images containing many small, almost equally sized CCs. In such images, k-means often outperforms the intuitive segmentation. However, the k-means segmentation accuracy decreases significantly in images consisting of a few large CCs located close to each other.

X. CONCLUSIONS AND FURTHER WORKS

In this paper we presented the complete procedure of patent image segmentation supported by caption detections. In both phases of the procedure (detecting captions and segmenting) we obtained the correctness at the level of about 90%. This result seems to be practically acceptable and makes it possible to recommend it for practical application in patent document processing.

In the future research we plan to introduce the third approach to clustering problem which is most closely related to the formal definition of the problem in equation 5. We are going to apply the simulated annealing method which seems well suited to our discrete optimization problem with a huge search space. . Caption detection method should be also improved to avoid some segmentation errors induced by true caption omission. Promising direction seems to apply 2D Fourier spectrum analysis in order to raise text area detection accuracy.

REFERENCES

[1] K. Suchet Chachra, Z. Xue, S. Antani, D. Demner-Fushman, and G. R. Thoma. Extraction and labeling high-resolution images from pdf documents. In *Proc. SPIE 9021, Document Recognition and Retrieval XXI, 90210Q (24 March 2014);*, 2014. doi:10.1117/12.2042336.

[2] A. Chhatkuli, A. Foncubierta-Rodriguez, D. Markonis, F. Meriaudeaub, and H. Mueller. Separating compound figures in journal articles to allow for subfigure classification. In *Proc. SPIE 8674, Medical Imaging 2013: Advanced PACS-based Imaging Informatics and Therapeutic Applications*, 2013. doi:10.1117/12.2007971.

[3] C. Clark and S. Divvala. Looking beyond text: Extracting figures, tables, and captions from computer science paper. In *Scholarly Big Data: AI Perspectives, Challenges, and Ideas: Papers from the 2015 AAAI Workshop*, pages 2–8, 2015.

[4] D. Hunt, L. Nguyen, and M. Rodgers. *Patent Searching: Tools & Techniques*. Wiley, 2007.

[5] L. D. Lopez, J. Yu, C. O. Tudor, C. N. Arighi, H. Huang, K. Vijay-Shanker, and C. H. Wu. Robust segmentation of biomedical figures for image-based document retrieval. In *2012 IEEE International Conference on Bioinformatics and Biomedicine*, 2012. doi: 10.1109/BIBM.2012.6392706.

[6] X. Lu, S. Kataria, W. J. Brouwer, J. Z.Wang, and M. Prasenjit ůand C. Lee Giles. Automated analysis of images in documents for intelligent document search. *IJDAR*, 2009. doi:10.1007/s10032-009-0081-0.

[7] X. Lu, P. Mitra, J. Z. Wang, and C. Lee Giles. Automatic categorization of figures in scientific documents. In *Joint Conference on Digital Library, JCDL 06, USA.*, 2006. doi:10.1145/1141753.1141778.

[8] J. Macqueen. Some methods for classification and analysis of multivariate observations. In *In 5-th Berkeley Symposium on Mathematical Statistics and Probability*, pages 281–297, 1967.

[9] P. A. Praczyk, J. Nogueras-Iso, and S. Mele. Automatic extraction of figures from scientific publications in high-energy physics. *Information Technology and Libraries*, pages 25–52, December 2013.

[10] M. Prasenjit S. R. Choudhury and G. Clyde Lee. Automatic extraction of figures from scholarly documents. In *DocEng '15 Proceedings of the 2015 ACM Symposium on Document Engineering*, pages 47–50, 2015. doi:10.1145/2682571.2797085.

[11] J. Sas and A. Zolnierek. Three-stage method of text region extraction from diagram raster images. In *Proceedings of the 8th International Conference on Computer Recognition Systems CORES 2013, Milkow, Poland, 27-29 May 2013*, pages 527–538, 2013. doi:10.1007/978-3-319-00969-8-52.

[12] A. W. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-based image retrieval at the end of the early years. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(12):1349–1380, December 2000. doi:10.1109/34.895972.

[13] S. Vrochidis, S. Papadopoulos, A. Moumtzidou, P. Sidiropoulos, E. Pianta, and I. Kompatsiaris. Towards content-based patent image retrieval; a framework perspective. *World Patent Information Journal*, 32(2):94–106, 2010. doi:10.1016/j.wpi.2009.05.010.

[14] X. Yuan and D. Ang. A novel figure panel classification and extraction. *Int. J. Data Min. Bioinformatics*, 9(1):22–36, November 2014. doi:10.1504/IJDMB.2014.057779.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 271–274
DOI: 10.15439/2016F133
ACSIS, Vol. 8. ISSN 2300-5963

Using Spatial Pooler of Hierarchical Temporal Memory for object classification in noisy video streams

Maciej Wielgosz, Marcin Pietroń, Kazimierz Wiatr
AGH University of Science and Technology
al. Mickiewicza 30, 30-059 Krakow, Poland
Academic Computer Centre 'Cyfronet' AGH
Nawojki 11, 30-950 Krakow, Poland
Email: wielgosz,pietron,wiatr@agh.edu.pl

***Abstract*—This paper focuses on analyzing a Spatial Pooler (SP) of Hierarchical Temporal Memory (HTM) ability for facilitating object classification in noisy video streams. In particular, we seek to determine whether employing SP as a component of the video system increases overall robustness to noise. We have implemented our own version of HTM and applied it to object recognition tasks under various testing conditions. The system is composed of a video preprocessing block, a dimensionality reduction section which contains SP, a histograms collecting module and SVM classifier.**

Our experiments involve assessing performance of two different system setups (i.e. a version featuring SP and one without it) under various noise conditions with 32–frame video files. In order to make tests fair and repeatable the videos of several 3–D geometric shapes were artificially generated. Subsequently, Gaussian noise of a different intensity was introduced to the videos making them more indistinct. Such an approach mimics real–life scenarios where the system is taught ideal objects and then faces in its normal working conditions the challenge of detecting noisy ones.

The results of the experiments reveal the superiority of the solution featuring Spatial Pooler over the one without it. Furthermore, the system with SP performed better also in the experiment without a noise component introduced and achieved a mean F1–score of 0.91 in ten trials.

I. Introduction

DESPITE the huge technological growth witnessed nowadays, there are still no autonomous machines available which would be capable of operating in the real world. Such machines would take over most of our tedious everyday duties and clear the way for a breakthrough in Artificial Intelligence. However, such robots need to be able to process inputs in real time, learn, generalize and react to events. This requires building an appropriate processing system which has human–like capabilities [1] [2].

A mammalian brain is an example of such a system which evolved over millions of years. Despite its apparent complexity there is only one algorithm [3] within the brain which governs the body functions. This allows for scalability of the solutions based on the algorithm since more complex systems may be built on a top of the simpler ones just by duplication of the basic structure.

The human brain as a whole has not been completely explored yet, making its artificial implementation and verification a very hard task. However, there are initiatives [4] which have taken up the challenge of simulating and modeling a brain as we know it today. Rather than model the brain, the authors of this paper have adopted a slightly different approach of gradually introducing selected components of HTM to the video processing system with the intention of enhancing its performance. By doing so we aim to develop a complete system working on the principles of the human brain as they were presented in [3][5] with our modifications making the algorithm suitable for hardware implementation.

Consequently, this paper attempts to take a step forward in examining the feasibility of using HTM for classifying objects in noisy video streams. The authors state the following hypothesis: employing Spatial Pooler in the video processing flow will improve the object classification ratio due to its beneficial reduction property of mapping to Sparse Distributed Representation (SDR). It is verified through a series of experiments.

The rest of the paper is organized as follows. Sections I-A and I-B provide the background and related work of object classification in video streams and Hierarchical Temporal Memory, respectively. The custom designed system used for the experiments is presented in Section II. Section III provides the results of the experiments. Finally, the conclusions of our research are presented in Section IV.

A. Object classification in video streams

Most state–of–the–art information extraction systems consist of the following sections: preprocessing, feature extraction, dimensionality reduction and classifier or ensemble of classifiers (Fig. 1). Their construction requires expertise knowledge as well as familiarity with the data that will be processed [6][7].

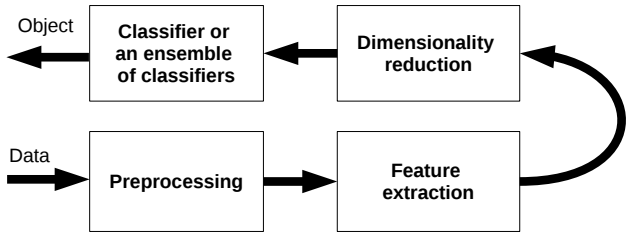


Fig. 1. Architecture of a video processing system

Usually, systems for object classification in video streams are also designed according this scheme. Consequently, the proper choice of the operations which constitute all the mentioned stages of the system is important and decides about the classification result [8]. One of the most challenging stages is feature extraction, which substantially affects the overall performance of the system.

There are also systems which take advantage of the spatial–temporal [5] profile of the data [1][9]. They are closer to the concept of the solution presented in this paper, which may be considered a hybrid approach since it features components of both schemes.

B. Hierarchical Temporal Memory

Hierarchical Temporal Memory (HTM) replicates the structural and algorithmic properties of the neocortex. It can be regarded as a memory system which is not programmed, but trained through exposing it to data flow. The process of training is similar to the way humans learn which, in its essence, is about finding latent causes in the acquired content. At the beginning, the HTM has no knowledge of the data stream causes it examines, but through a learning process it explores the causes and captures them in its structure. The training is considered completed when all the latent causes of data are captured and stable. The detailed presentation of HTM is provided in [5][10][11].

HTM is composed of two main parts, namely Spatial and Temporal Pooler. This paper focuses on Spatial Pooler (SP), aka Pattern Memory, which is employed in the processing flow of the system. Is contains columns with synapses connected to the input data [5]. The main role of SP in HTM is finding spatial patterns in the input data. It may be decomposed into three stages:

- Overlap calculation,
- Inhibition,
- Learning

The first two stages are very computationally demanding but can be parallelized, therefore the authors decided to implement them on GPU in OpenCL. The learning stage is implemented on CPU in Python. The detailed description of algorithms is provided in [5].

The overlap section computes *col.overlap* for every column in SP structure i.e. a number of active and connected synapses.

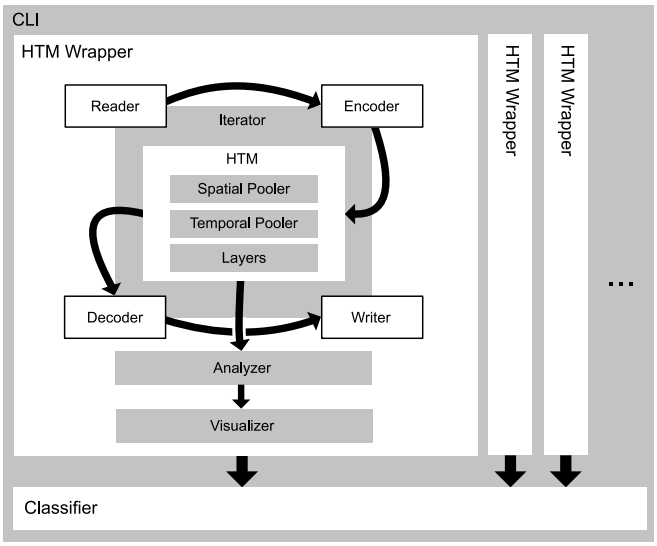


Fig. 2. Architecture of the implemented system

If the number is larger than *col.min_overlap* then it is boosted and passed on to the inhibition section.

The inhibition stage implements a winner–takes–all procedure where for each column a decision is made as to whether it belongs to a range of n columns of the highest values.

II. System description

The implemented HTM version [12] follows description from [5], with the exception of synapses bias implementation being replaced with random connections. Its main purpose however is to emphasize the algorithm parallelism and to allow progressively replace parts of it with GPU–accelerated (and in the future FPGA–accelerated) fragments written in OpenCL. The system (Fig. 2) is highly configurable, with numerous parameters responsible for the core HTM's structure, the encoder behavior, statistics rendering, etc. The configuration is stored in a file written in JSON format, which allows it to maintain its readability while providing clear structure. In addition to the core module, a set of supporting modules has been developed. Most of them are used for feeding video data to the core module, and receiving and analyzing the results.

The complete processing flow of the system is presented in Fig. 3. The data is fed into the system in a frame–by–frame manner. In the first step the original frame is reduced and binarized using OpenCV procedures. During the encoding process, the original video frame is converted to a smaller binary image. Preliminary tests showed that reducing a 960x540 image by a factor of 16 (producing 60x33 images) has a low impact on the end results while significantly shortening the processing time (Fig. 4). After reduction, the color image is converted to grayscale, which later is turned into a binary one using adaptive thresholding (using a potentially different threshold value for each small image region).

Those operations constitute the encoding which allows the generation of input data for the SP processing stage. There-

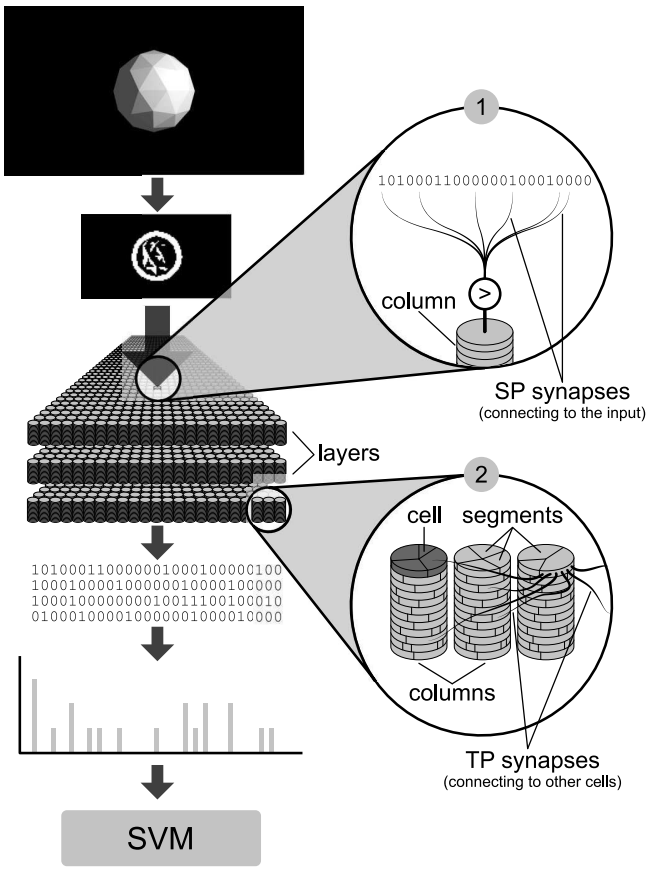


Fig. 3. Block diagram of the proposed approach

after, the data is mapped to SDR with SP and may be passed on to TP. Since our current work focuses on finding the optimal SP parameters values we do not use TP in this particular application, but the system in general has such capability. In the next step, histograms of consecutive processed frames are built on a per–video basis. The histograms are used as the input data for the SVM classifier which comes next.

There are two different working modes of the system, namely learning and testing modes. The system is trained in the learning mode with 80% of available data and then it is tested with the remaining 20% of the data in the testing mode.

III. Experiments and the discussion

A series of experiments (details of which are provided in Tab. I and Tab. II) was conducted to validate the hypothesis stated in the introduction of the paper. They allow a comparison of the performance of the system featuring Spatial Pooler in the processing flow with the one lacking it.

The challenging part involved generation of sample videos for testing (available from [13]. The videos had to meet a series of requirements such as object location, camera location and object–camera distance. Consequently, a dedicated application was used to generate the videos (i.e. Blender [14]). Blender provides Python API, which was used to automate and randomize the video generation. Each video contains a single, centered shape and a randomly positioned light source which brightness is picked from predefined range. During the course of a video the camera randomly changes position. Since noise addition at the runtime proved to be a very time consuming process, a separate script introducing noise to a large set of generated videos was created. Embedding noise also ensures equal conditions for all the experiments and test setups.

TABLE I
VIDEO PARAMETERS

Parameter		Value
Size of a single video frame		960x540
Reduction level		16
Frame size after reduction		60x33
No. of frames in a single video		32
Object classes		cone, cube, cylinder, monkey, sphere, torus
No. of classes		6
Total no. of videos	all	6000
	training	4800
	testing	1200
Videos per class	all	1000
	training	800
	testing	200
Videos per trial	all	100
	training	80
	testing	20

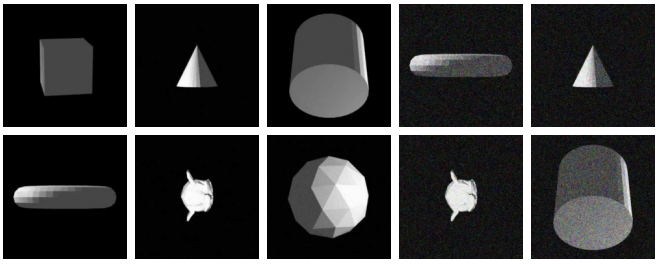


Fig. 4. Sample frames of different shapes and noise levels

The F1–score is used as a quality evaluation of the experiments' results presented in this paper.

Experiments conducted for higher noise levels than presented in Tab. III resulted in an unacceptably low F1–score (below 0.75). Therefore, the authors decided not to include the results of those experiments in the table.

It is worth noting there there was a huge difference in the calculation time of a single experiment across setups, see Tab. IV. This is the result of an HTM algorithm complexity,

TABLE II
SP PARAMETERS

Parameter	Value
No. of columns	2048
No. of synapses per column	64
Perm value increment	0.1
Perm value decrement	0.1
Min overlap	8
Winners set size	40
Initial perm value	0.21
Initial inhibition radius	80

TABLE III
EXPERIMENTS RESULTS FOR DIFFERENT NOISE LEVELS

Noise level (σ)	F1–score : mean and variance (10 repetitions)	
	SVM	SVM + SP
0	0.82 (0.001)	0.91 (0.001)
4.25	0.81 (0.001)	0.89 (0.001)
8.5	0.78 (0.001)	0.88 (0.003)

TABLE IV
EXECUTION TIME OF A SINGLE EXPERIMENT FOR DIFFERENT SETUPS

Setup	Time [hours]
SVM	0.23
SVM + SP	28.5

object–oriented programming approach and high level scripting language used in implementation. Code optimization and introducing more hardware–accelerated fragments should improve execution time. The tests were run on Intel(R) Core(TM) i5–4210M CPU @ 2.60GHz, and Nvidia GeForce GT 730M (selected sections of SP). Each experiment consisted of 10 trials.

According to the authors knowledge it is hard to find papers which directly correspond to the research conducted in this work (i.e. video classification in noisy video streams). Nevertheless, we examined the following papers : [15], [16], [17] which presents results of video classification using UCF-101 dataset. The best systems presented in those papers are based on various architectures of Convolutional Neural Networks (CNNs) and achieve accuracy of 80% and more. It is worth emphasising that despite similar performance in terms of the quality results our test setup is different mostly in a process of the video generation.

IV. CONCLUSIONS AND FUTURE WORK

This paper presents the preliminary experimental results of using an HTM–based system for object classification in video streams. The authors showed that using SP in the video processing flow improves the object classification ratio by approx. 10%. In future work, the authors are going to modify the preprocessing stage of the video processing flow and introduce TP. The authors are going to implement the most computationally–exhaustive routines in OpenCL and deploy the system on platforms equipped with GPU– or FPGA–based acceleration. This will enable conduction of experiments with video of a lower image reduction ratio.

TABLE V
EXAMPLE CONFUSION MATRIX FOR SVM–ONLY SETUP AND GAUSSIAN NOISE WITH $\sigma = 8.5$

	Predicted classes					
	cone	cube	cylinder	monkey	sphere	torus
cone	19	0	0	0	0	1
cube	0	10	8	0	2	0
cylinder	0	3	16	0	1	0
monkey	0	1	2	14	2	1
sphere	0	3	3	0	13	1
torus	0	1	0	0	0	19

ACKNOWLEDGMENT

I would like to thank my wife Urszula Wielgosz for her huge contribution to the preparation of the paper.

REFERENCES

[1] S. Sengupta, H. Wang, W. Blackburn, and P. Ojha, "Spatial information in classification of activity videos," in *Proceedings of the 2015 Federated Conference on Computer Science and Information Systems*, ser. Annals of Computer Science and Information Systems, M. Ganzha, L. Maciaszek, and M. Paprzycki, Eds., vol. 5. IEEE, oct 2015. doi: 10.15439/2015F382 pp. 145–153.

[2] J. F. Sowa, "The Cognitive Cycle," in *Proceedings of the 2015 Federated Conference on Computer Science and Information Systems*, ser. Annals of Computer Science and Information Systems, M. Ganzha, L. Maciaszek, and M. Paprzycki, Eds., vol. 5. IEEE, oct 2015. doi: 10.15439/2015F003 pp. 11–16.

[3] V. Mountcastle, "The columnar organization of the neocortex," *Brain*, vol. 120, no. 4, pp. 701–722, apr 1997. doi: 10.1093/brain/120.4.701

[4] "The Human Brain Project - Human Brain Project," (Accessed on 10.04.2016). [Online]. Available: https://www.humanbrainproject.eu

[5] J. Hawkins, S. Ahmad, and D. Dubinsky, "Hierarchical temporal memory including HTM cortical learning algorithms," Numenta, Inc, Tech. Rep., sep 2011. [Online]. Available: http://numenta.org/resources/HTM_CorticalLearningAlgorithms.pdf

[6] H. He and E. A. Garcia, "Learning from imbalanced data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 9, pp. 1263–1284, sep 2009. doi: 10.1109/TKDE.2008.239

[7] P. Zhang, X. Zhu, and L. Guo, "Mining Data Streams with Labeled and Unlabeled Training Examples," in *2009 Ninth IEEE International Conference on Data Mining*, IEEE. Miami, USA: IEEE, dec 2009. doi: 10.1109/ICDM.2009.76 pp. 627–636.

[8] R. N. Hota, V. Venkoparao, and A. Rajagopal, "Shape Based Object Classification for Automated Video Surveillance with Feature Selection," in *10th International Conference on Information Technology (ICIT 2007)*, IEEE. Rourkela, India: IEEE, dec 2007. doi: 10.1109/ICIT.2007.57 pp. 97–99.

[9] Y. Bengio, A. Courville, and P. Vincent, "Representation Learning: A Review and New Perspectives," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1798–1828, aug 2013. doi: 10.1109/TPAMI.2013.50

[10] X. Chen, W. Wang, and W. Li, "An overview of Hierarchical Temporal Memory: A new neocortex algorithm," in *Modelling, Identification & Control (ICMIC), 2012 Proceedings of International Conference on*. Wuhan, China: IEEE, 2012, pp. 1004–1010.

[11] D. Rachkovskij, "Representation and processing of structures with binary sparse distributed codes," *IEEE Transactions on Knowledge and Data Engineering*, vol. 13, no. 2, pp. 261–276, 2001. doi: 10.1109/69.917565

[12] "Hierarchical temporal memory implementation," (Accessed on 12.04.2016). [Online]. Available: https://bitbucket.org/maciekwielgosz/htm-hardware-architecture

[13] "HTM Test Datasets," (Accessed on 02.07.2016). [Online]. Available: http://data.wielgosz.info

[14] "Blender project - Free and Open 3D Creation Software," (Accessed on 12.04.2016). [Online]. Available: https://www.blender.org/

[15] Joe Yue-Hei Ng, M. Hausknecht, S. Vijayanarasimhan, O. Vinyals, R. Monga, and G. Toderici, "Beyond short snippets: Deep networks for video classification," *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4694–4702, jun 2015. doi: 10.1109/CVPR.2015.7299101

[16] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei, "Large-Scale Video Classification with Convolutional Neural Networks," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, jun 2014. doi: 10.1109/CVPR.2014.223 pp. 1725–1732.

[17] S. Zha, F. Luisier, W. Andrews, N. Srivastava, and R. Salakhutdinov, "Exploiting Image-trained CNN Architectures for Unconstrained Video Classification," *ArXiv e-prints*, mar 2015. [Online]. Available: http://arxiv.org/abs/1503.04144

6th International Workshop on Artificial Intelligence in Medical Applications

THE workshop on Artificial Intelligence in Medical Applications – AIMA'2016—provides an interdisciplinary forum for researchers and developers to present and discuss latest advances in research work as well as prototyped or fielded systems of applications of Artificial Intelligence in the wide and heterogenious field of medicine, health care and surgery. The workshop covers the whole range of theoretical and practical aspects, technologies and systems based on Artificial Intelligence in the medical domain and aims to bring together specialists for exchanging ideas and promote fruitful discussions.

TOPICS

- Artificial Intelligence Techniques in Health Sciences
- Knowledge Management of Medical Data
- Data Mining and Knowledge Discovery in Medicine
- Health Care Information Systems
- Clinical Information Systems
- Agent Oriented Techniques in Medicine
- Medical Image Processing and Techniques
- Medical Expert Systems
- Diagnoses and Therapy Support Systems
- Biomedical Applications
- Applications of AI in Health Care and Surgery Systems
- Machine Learning-based Medical Systems
- Medical Data- and Knowledge Bases
- Neural Networks in Medicine
- Ontology and Medical Information
- Social Aspects of AI in Medicine
- Medical Signal and Image Processing and Techniques
- Ambient Intelligence and Pervasive Computing in Medicine and Health Care

EVENT CHAIRS

- **Lasek, Piotr,** University of Rzeszow, Poland
- **Paja, Wiesław,** University of Rzeszów, Poland
- **Pancerz, Krzysztof,** University of Rzeszów, Poland

PROGRAM COMMITTEE

- **Azar, Ahmad Taher,** Benha University, Egypt
- **Iantovics, Barna,** Petru Maior University, Romania
- **Kountchev, Roumen,** Technical Univerity of Sofia, Bulgaria
- **Leniowska, Lucyna,** University of Rzeszow, Poland
- **Majernik, Jaroslav,** Pavol Jozef Safarik University in Kosice, Slovakia
- **Ngan, Ben C.K.,** The Pennsylvania State University, United States
- **Olszewska, Joanna Isabelle,** University of Gloucestershire, United Kingdom
- **Sawada, Hideyuki,** Kagawa University, Japan
- **Schwarz, Daniel,** Masaryk University, IBA, Czech Republic
- **Subbotin, Sergey,** Zaporizhzhya National Technical University, Ukraine
- **Wojciechowski, Konrad,** Silesian University of Technology, Poland
- **Zaitseva, Elena,** University of Zilina, Slovakia

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 277–280
DOI: 10.15439/2016F165
ACSIS, Vol. 8. ISSN 2300-5963

Customized Web-based System for Elderly People Using Elements of Artificial Intelligence

František Babič
Department of Cybernetics and Artificial Intelligence, Faculty of Electrical Engineering and Informatics, Technical university of Košice, Slovakia
frantisek.babic@tuke.sk

Adrián Jančuš
Department of Cybernetics and Artificial Intelligence, Faculty of Electrical Engineering and Informatics, Technical university of Košice, Slovakia
adrian.jancus@student.tuke.sk

Katarína Melišová
Department of Cybernetics and Artificial Intelligence, Faculty of Electrical Engineering and Informatics, Technical university of Košice, Slovakia
katarina.melisova@student.tuke.sk

***Abstract*—Making life easier for the elderly represents a new challenge for the ICT sector. This paper presents a new web-based system designed and implemented with the aim to support the social inclusion and to improve the daily routine of the elderly people within basic information and communication features. The system provides some advanced functionalities to utilise the information value of the data collected within the presented system, e.g. the recommendations based on similar hobbies or health problems; a simple medical diagnostics; a creation of a knowledge base containing experiences and best practices, etc. We designed the system in accordance with local conditions in Slovakia, so its full functioning relies on the progress in e-Health legislation. Presented version is a preliminary result that will be further improved and tested within a real practice.**

I. Introduction

PROGRESSIVE ageing of the population represents a big challenge for various European politics and societies. The European Union has created some initiatives such as Digital Competence or Lifelong Learning Programme (LLP) to support these activities on the national and European level too [7]. Improving the lives of elderly people within suitable ICT solutions in combination with existing social or medical services represents a big challenge for not only public but also the private sector. Successful adaptation of these solutions requires cooperation between all involved organisations, governments, stakeholders and providers. Of course, various local conditions affect the whole implementation process in line with the common EU principles and create the barriers that need to be passed.

Acquired experiences and knowledge from two international EU funded projects OLDES and SPES motivated us to propose a new solution that will meet the identified requirements, will respect existing standards and will be customised for the Slovak local conditions. The project OLDES (Older People's e-services at home) aimed to develop a very low cost and easy to use entertainment and health care platform, designed to ease the life of the elderly in their homes. The implemented platform was tested at two different locations and countries: Prague in the Czech Republic and Bologna in Italy. The SPES project (Support Patients through E-services Solutions) aims at transferring the original approach and results achieved in implementing the OLDES focusing on new target problem domains: dementia, mobility-challenged persons, respiratory problems, and social exclusion [14]. The new tele-health and entertainment platform was tested in four different conditions: Ferrara (Italy), Vienna (Austria), Brno (Czech Republic) and Košice (Slovakia.)

The paper is organised as follows: the introduction with a brief overview of the selected similar approaches or initiatives to identify possible gap for a new solution. The second one describes the proposed solution with emphasis to meet the identified requirements and existing best practices in this domain. Next section presents the current prototype with the results of the testing. The last one concludes the paper and outlines some directions for our future work.

A. State of the Art

This section presents some selected initiatives and research works with the aim to identify possible gaps or research problems. As we mentioned before, the OLDES project resulted in the solution developed in accordance to the specific needs of the two pilots [13], [15]. The core of this solution is a low-cost PC creating a hub for other connected devices as TV set and the medical devices communicated via Bluetooth. The TV display information provided by the platform and users can access it through a simple remote control. In addition, users can actively participate in discussion groups through adapted handset connected the PC. In the case of medical devices, users tested e.g. a glucometer, scale, adapted version of a sphygmomanometer, etc. After testing partners identified following directions for future improvement: access through the remote controller, prefer medical devices commonly available on the market, the overall architecture of the platform.

The follow-up project SPES aimed to solve these issues and to support four target groups suffered by respiratory problems, dementia, handicapped people and social exclusion. It resulted in an information and communication technology platform connected to the different medical devices and installed in a patients' home [14], [21]. The main advantages of this solution are continuous monitoring of the elderly with the respiratory problems; the presence of various entertainment oriented application that supports orientation and practising memory and a set of services to support the social inclusion of the elderly through virtual communication, information sharing and a personal social network creation. Despite some improvements over the OLDES platform, still, some open issues and gaps existed,

e.g. SPES platform was available only as a desktop application and this approach required a two-layer transfer of the collected measurements from the medical devices, to the local and main database. From the users point of view, a possibility to create an own personal network was missing, e.g. based on similar hobbies or health problems. In addition, the GUI needed to be refined based on existing trends and best practices. Whereas the currently valid legislation in Slovakia does not allow sharing the medical information between patients and the doctor in the electronic form, some type of decision support system for common health problems can be a very interesting part of the supporting ICT (Information and communication technologies) solution for the elderly.

Decision support systems (DSS) in medicine have a long history that started within a system called Mycin [3]. The actual trend includes a creation of the sufficient quality and broad knowledge base further used for diagnostics of various diseases within suitable analytical methods [16], [12], [11], [6]. These works specifically focus on selected diseases so this factor strongly limits their generality and usability in a wider range. A separate group contains applications called symptom checkers. Typically, they are available as web-based applications and patients have a possibility to start the diagnostics process with an initial description of their symptoms. According to a recent study realised by the Harvard Medical School, most of these sites and applications provide inconsistent and unclear information and patients should not focus on this diagnosis and results [19].

Early diagnosis of the typical civilisation diseases represents one of the main challenges for the cooperation between artificial intelligence represented by analytical methods and medical data obtained through various medical tests or devices [8], [10], [22]. Equally important is a selection of the appropriate attributes for an effective diagnostics [17], [4], [5]. Authors in [1] used data collected in a family practice from Croatia to extract the patient's characteristics necessary for the positive diagnosis of the Metabolic Syndrome. The data include patients in retirement. Authors generated a set of decision tree models within CART or C4.5 algorithms. In addition, authors investigated the optimal cut-off values for the identified key biomarkers.

In summary, we concluded that still some gaps and potential for a new system to support the daily life and health status exist. In close cooperation with a possible target group, based on our previous experiences and actual existing approaches, we identified some features creating the core of a new web-based system. For example a possibility to create an own network of friends without a need to store personal data such as existing social network, a possibility to find a new friend based on similar hobbies or health problems, a possibility to use a simple decision support system to evaluate the discovered symptoms.

II. Web-based System

A. Functional Requirements

We specified following requirements based on the provided state of the art analysis, experiences obtained from the SPES project and in communication with a participated group of elderly from the Košice City. Users can use the system on the various devices, i.e. the system will be responsible. Users have a possibility to create their own profile containing information such as a nickname, hobbies, health problems, personal general practitioners or specialist, etc. In additional, users can create their own social network including people recommended based on the similar hobbies or health problems. Through this network, users have a possibility to share various multimedia or news, to use simple on-line chat or to invite their friends to the public or private event. From a medical point of view, users can manage their planned medical examinations, import and visualise the collected measurements from the supported medical devices or use the continuously constructed knowledge base to support the basic medical diagnosis. Last, but not least, the system informs the users about relevant information within simple notifications, current weather forecast or published newspaper articles from selected RSS (Rich Site Summary) source.

B. GUI Proposal

We designed the graphical user interface (GUI) in accordance to the current relevant W3C standard [22] and verified recommendations [9], i.e. we use a sans serif typeface, 12pt or 14pt type size for body text, medium or bold face type. The optimal option for a text alignment is the left. We present the body text in upper and lowercase letters; we use the capital letters and italics in headlines only. We reserve underlining only for the web links. We avoid using a combination of yellow, blue and green colour in a close proximity, because some elderly have problems to discriminate these colours. We used the graphics against a light background. We ensure that users can resize the content without assistive technology up to 200 percent without loss of content or functionality. We provide the labels when content requires a user input. The headings and labels describe topic or purpose.

C. Architecture Proposal

Presented prototype is a typical 3-layer client-server model containing a database, back-end and front-end service; see Fig. 1. The analytical package (language R, RStudio) is an integrated module to meet requirements from the legislative and safety point of view. We tested this proof of concept in the conditions in which the elderly mainly verified the predefined scenarios. During the next testing in the larger group of participants, we will discuss relevant lessons learned and possible improvements, not only from the usability point of view but also on the technological side, e.g. to take an advantage of the cloud computing.

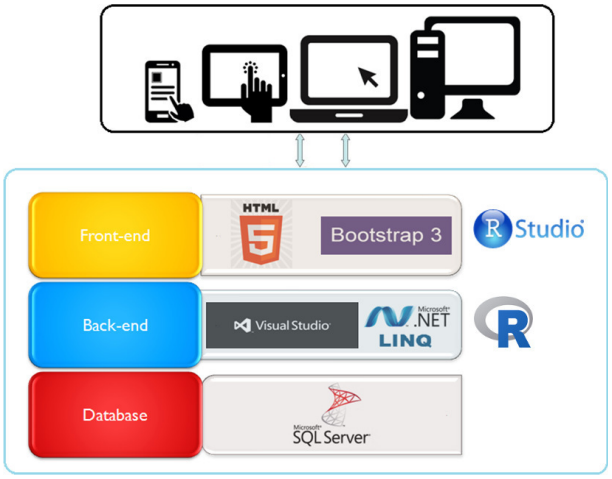


Fig. 1 Architecture with the relevant technologies

D. Analytical Services Proposal

Proposed analytical package includes the services necessary for storage, processing, analysis and knowledge base building. We generated initial decision rules and models through free available data samples for selected diseases to show a potential of this type of solution. In real operation, the system will replace this initial data within knowledge relevant to the current users and conditions, i.e. something like self-learning using the methods of artificial intelligence. We organised the whole analytical process in accordance with the CRISP-DM methodology which represents widely accepted approach on how to manage this process effectively [20]. We selected the decision trees based on their ability to process different types of input data, to process data with missing values and the most important factor was a simple understandable visualisation of the generated outputs for people with less knowledge from data mining or machine learning.

We used C4.5 [17] and CART [2] to extract possible interesting rules for diagnostics that were further added to the knowledge base. For example, IF a *female* is more than 70 years *old* AND *triceps skinfold thickness* is more than 28.5 mm AND *the level of insulin* is more than 15.15µIU/l THEN diagnosis of the Mild Cognitive Impairment is positive. Alternatively, IF *average level of blood glucose over the previous 3 months* was more than 4.4 AND *the level of insulin* is more than 27.1µIU/l THEN diagnosis of the Metabolic Syndrome is positive. It is important to say that all included rules need to be verified by cooperated medical experts and it is possible to use them only to support the decision, not as the final diagnosis.

E. Current Prototype

Tested prototype (available in Slovak language) contained all presented features that were available through a simple and intuitive user environment; see Fig. 2.

Users could read the articles from their favourite web newspapers (Noviny) and watch the weather forecast for the selected location (Počasie). They could create an own profile included hobbies or health problems (Profil) and based

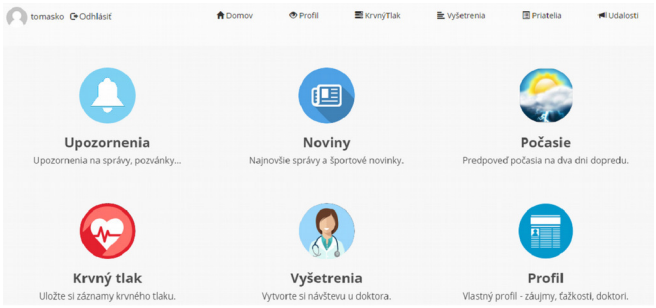


Fig. 2 Main page of the proposed system

recommendation generated from this data find a new friend. In order to improve the health status, they could import their measurements from medical devices (Krvný tlak) with a possibility to visualise (see Fig. 3) and export in the selected format; or create a database containing information about relevant doctors and planned examinations (Vyšetrenia). In addition, a possibility to analyse the collected medical data was at disposal, but the final diagnostics is the responsibility of the general practitioner or specialist.

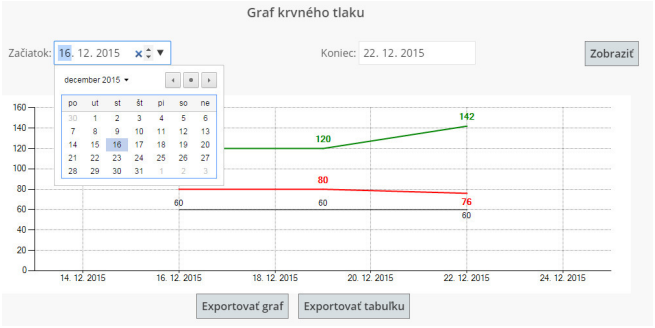


Fig. 3 Visualisation of imported blood pressure measurements

F. Testing

One of the main reasons why it is not possible to present the proposed system as a finished solution, but only as a proof of concept, is the current legislative in Slovakia covering protection of personal data and e-Health. However, this situation is changing every year under the influence of the various initiatives on the European Union level. We tested our prototype with a small group of elderly in order to evaluate their satisfaction and overall potential of this type of system in Slovak conditions. Elderly realised predefined test scenarios representing typical uses for such systems and answered a predefined questionnaire. This testing confirmed our initial hypotheses such as elderly enjoy the graphical design of the system and it meets relevant requirements and best practices. Also, elderly are satisfied with a set of offered features and with an approach how these features are provided to the users, i.e. the system is easy to use without necessary deep knowledge from the ICT domain. The system helps elderly to reach more active life; to make their daily routine more pleasant and to make their contact with an external environment more intensive. The participated medical expert evaluated the generated models and extracted decision rules based on her knowledge and existing

literature, e.g. accordance to the IDF (International Diabetes Federation) definition of the Metabolic Syndrome.

III. Conclusion

The article presents the design and development of the customised web-based system, which would help elderly to the living that is more active and better health status. We aimed to solve some gaps as a possibility to recommend a friend based on similar characteristics or to analyse the data about health status through the simple decision support system. The positive results of the testing confirm our expectations and motivate our future work devoted to the e.g. automatic creation of the user profile based on his behaviour and typical daily habits, or a self-learning and adapting mechanism for created knowledge base. In this case, we will continue to use the methods of artificial intelligence in combination with actual trends in domains as Internet of things, home health care and robotics. In addition, we can understand this version as a proof of concept that creates a good foundation for our future research activities oriented to the H2020 project proposal.

Acknowledgment

The work presented in this paper was partially supported by faculty internal research project no. FEI-2015-2 and by the Slovak Grant Agency of the Ministry of Education and Academy of Science of the Slovak Republic under grant No. 1/0493/16.

References

[1] F. Babič, L. Majnarič, A. Lukáčová, J. Paralič, A. Holzinger, "On Patient's Characteristics Extraction for Metabolic Syndrome Diagnosis: Predictive Modelling Based on Machine Learning", in Information Technology in Bio- and Medical Informatics, LNCS Vol. 8649, 2014, pp. 118-132, 10.1007/978-3-319-10265-8_11.

[2] L. Breiman, J. H. Friedman, R. A. Olshen, C. J. Stone, "Classification and regression trees", Monterey, CA: Wadsworth & Brooks/Cole Advanced Books & Software, 1984.

[3] B. G., Buchanan, E.H. Shortliffe," Rule Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project", Reading, MA: Addison-Wesley, 1984.

[4] P. Butka, J. Pócs, J. Pócsová, "On Equivalence of Conceptual Scaling and Generalized One-Sided Concept Lattices", in Information Sciences 259, 2017, pp. 57-70, http://dx.doi.org/10.1016/j.ins.2013.08.047.

[5] P. Butka, J. Pócs, J. Pócsová, "Distributed Computation of Generalized One-Sided Concept Lattices on Sparse Data Tables", in Computing and Informatics 34 (1), 2015, pp. 77-98.

[6] R. J. Conejar, H. K. Kim, "A Medical Decision Support System (DSS) for Ubiquitous Healthcare Diagnosis System", in International Journal of Software Engineering and Its Applications, 8 (10), 2014, pp. 234-244, http://dx.doi.org/10.14257/ijseia.20104.8.10.22.

[7] R. Edwards, "Changing Places?: Flexibility, Lifelong Learning and a Learning Society", Routledge, 1997.

[8] P. Han-Saem, C. Sung-Bae, "Evolutionary attribute ordering in Bayesian networks for predicting the metabolic syndrome", in Expert Systems with Applications, 39(4), 2012, pp. 4240-4249, 10.1016/j.eswa.2011.09.110.

[9] R. J. Hodes, D. A.B. Lindberg, "Making Your Web Site Senior Friendly", published by the National Institute on Aging and the National Library of Medicine, 2002.

[10] A. Holzinger, M. Dehmer, I. Jurisica, "Knowledge Discovery and Interactive Data Mining in Bioinformatics – State-of-the-Art, in Future challenges and Research Directions", BMC Bioinformatics 15(suppl. 6), I1, 2014, 10.1186/1471-2105-15-S6-I1.

[11] V. A. Kamaev, D. P. Panchenko, N.V. Le, O. A. Trushkina, „An Intelligent Medical Differential Diagnosis System Based on Expert Systems", in Knowledge-Based Software Engineering, Communications in Computer and Information Science, Volume 466, 2014, pp. 576-584.

[12] T. Matsumoto, Y. Ueda, S. Kawaji, "A software system for giving clues of medical diagnosis to clinician", in Proceedings of 15th IEEE Symposium on Computer-Based Medical Systems (CBMS 2002), IEEE, Maribor, Slovenia, 2002, pp. 65-70, 10.1109/CBMS.2002.1011356.

[13] D. Novák, O. Štepanková, M. Mráz, M. Haluzík, M. Bussoli, M. Uller, K. Maly, L. Nováková, P. Novák, "OLDES: new solution for long-term diabetes compensation management", in Proceedings of 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Vancouver, Canada, 2008, pp. 4346-4349, 10.1109/IEMBS.2008.4650172.

[14] D. Novák, O. Štepánková, S. Rousseaux, M. Busuoli, M. Carulli, G. D'Agosta, T. Gallelli, M. Uller, et al., "Does IT Bring Hope for Wellbeing?", in book Handbook of Research on ICTs for Human-Centered Healthcare and Social Care Services, IGI Global, Editors: Maria Manuela Cruz-Cunha, Maria Miranda, 2103, pp. 270-302.

[15] D. Novák, M. Uller, S. Rousseaux, M. Mráz, J. Smrž, O. Štepanková, M. Haluzík, M. Busuoli, "Diabetes management in OLDES project", in Proceedings of 31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Minneapolis, Minnesota, USA, 2009, pp. 7228-7231, doi: 10.1109/IEMBS.2009.5335256.

[16] W. Oude Nijeweme-d'Hollosy., L. S. van Velsen, R. Soer, H. J. Hermens, "Design of a web-based clinical decision support system for guiding patients with low back pain to the best next step in primary healthcare", in: Proceedings of the 9th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2016), Rome, Italy, 2016, pp. 229-239, 10.5220/0005662102290239.

[17] J. R. Quinlan, "C4.5: Programs for Machine Learning", San Mateo: Morgan Kaufmann, 1993.

[18] N. Pérez, M. A. Guevara, A. Silva, I. Ramos, J. Loureiro, "Improving the performance of machine learning classifiers for Breast Cancer diagnosis based on feature selection", in Proceedings of the 2014 Federated Conference on Computer Science and Information Systems, 2014, pp. 209-217, 10.15439/2014F249.

[19] H. L. Semigran, J. A. Linder, C. Gidengil, A. Mehrotra, "Evaluation of symptom checkers for self-diagnosis and triage: audit study", in BMJ, 351:h3480, 2015, 10.1136/bmj.h3480.

[20] C. Shearer, "The CRISP-DM Model: The New Blueprint for Data Mining", in Journal of Data Warehousing, 5 (4), 2000, pp. 13-22.

[21] G. Tutoky, F. Babic, Wagner J., "ICT-based solution for elderly people", in Proceedings of IEEE 11th International Conference on Emerging eLearning Technologies and Applications, Stará Lesná, Slovakia, 2013, pp. 366-404, 10.1109/ICETA.2013.6674466.

[22] E. Zaitseva, M. Kvassay, V. Levashenko, J. Kostolny, "Introduction to knowledge discovery in medical databases and use of reliability analysis in data mining", in Proceedings of the 2015 Federated Conference on Computer Science and Information Systems, 2015, pp. 311-320, 10.15439/2015F327.

[23] W3C: Web Accessibility Initiative: Developing Websites for Older People: How Web Content Accessibility Guidelines (WCAG) 2.0 Applies.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 281–285
DOI: 10.15439/2016F244
ACSIS, Vol. 8. ISSN 2300-5963

The new method of the selection of features for the k-NN classifier in the arteriovenous fistula state estimation

Marcin Grochowina
University of Rzeszów
al. Rejtana 16, 35-310 Rzeszów, Poland
Email: gromar@ur.edu.pl

Lucyna Leniowska
University of Rzeszów
al. Rejtana 16, 35-310 Rzeszów, Poland
Email: lleniow@ur.edu.pl

***Abstract*—In this paper the application of a new method of features selection was presented. Its effects were compared with several other methods of features selection. The study were performed using a data set containing samples of the sound signal emitted by the arteriovenous fistula. The aim was to create a solution with multiclass classification based on the k-NN classifier family allowing for effective and credible assessment of the state of arterial-venous fistula.**

I. Introduction

EACH classification process is based on the set of features delivered to the classifier on the basis of which a decision is taken and the result obtained. Proper selection of a set of features significantly improves the quality of the classification process.

The approach ensuring the best quality is to test all possible non-empty subsets of the input set. Unfortunately, the number of non-empty subsets of n-element is $2^n - 1$, which implies the possibility of a full review of the subsets only for the n that does not exceed a dozen elements. Full analysis for larger values of n is too time-consuming. It is therefore the use of quasi-optimal methods, which is determinative of a subset of the features of possible high efficiency of classification.

K-NN is the most popular minimum distance classifier. It assigns the unknown sample to the class most often representing its neighborhood[14][15]. There are many variants of this method. They differ among themselves, inter alia, by methods of calculating the distance and the method of voting that determines the result.

The most common variation of k-NN is weighted k-NN , in which weight of the neighbor of samples $\boldsymbol{x}$ depends on its distance from the $\boldsymbol{x}$[13]. An interesting solution is the Diplomatic Nearest Neighbors (k-DN) [12], which seeks k neighbors of each class separately, and then selects the class for which the average distance from the found neighbors to the tested sample is the smallest.

Due to its flexibility, simplicity and the possibility of use in tasks of classification and regression k-NN is popular despite its flaws: it requires storage in the memory the whole training set and high demand for computing power, especially for large training sets.

II. Data set

In the studies the data set consisting of sounds emitted by the arteriovenous fistula was used. The studies to date [1][2] show that the character of the sound emitted by the blood flowing within the fistula differs depending on the condition of the fistula.

The research data set was collected from 19 patients with radiocephalic fistula. Aquisition of the material consisted in recording the sound of the blood flowing through the arteriovenous fistula. Material was collected using a dedicated head equipped with an electret microphone CZ034 manufactured by Ringford, with a sensitivity of -42dB (0dB=1V/Pa, 1kHz), ie. 8mV/Pa and an interval signal/noise ratio greater than 60dB. To register a signal, an integrated sound card was used as part of the RV730 Radeon 4000 manufactured by AMD as well as dedicated software running under the Linux operating system. Sampling frequency was set at 8 kHz.

Numerical processing of data was performed using WEKA 3.7.13 package running with the JRE Oracle Java 1.8. The calculations were performed on a computer with Intel Core 2 T6570 2.1GHz under the Linux operating system. During the measurements the algorithms time requirement only a single core processor was used.

Fistulas were rated as effective, however, to differing degrees. Eight groups representing a fistula with varying degrees of stenosis were extracted. A total of 1190 samples was collected.

The groups were lettered with labels ***a-h*** , wherein the group ***a*** were fistulas in the best condition and in the group ***h*** in the worst condition. With the collected data set 23 features were extracted; 6 in the time and 17 in the frequency domain. Features in the time domain named t0, t4, y0, y4, p0 and p4 describe the timing, amplitude and shape of the signal envelope within a single period of the rhythm of the heart. Features in the frequency domain named f1-f17 describe the density of the frequency spectrum of the recorded signal at specific intervals from the scope of 20-600Hz.

978-83-60810-90-3/$25.00

III. METHODS

In this study five methods of feature selection were tested. Each of them belongs to a different category of methods (Figure 1).

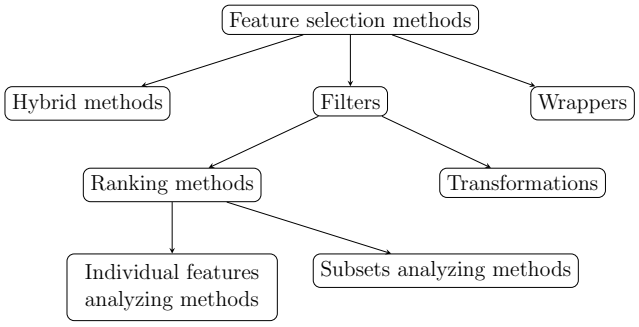


Fig. 1: Feature selection methods

The first four are commonly known and available in the WEKA package. The fifth is an own method developed proprietarily to the needs of this particular task.

The methods used are:

- *Correlation* - builds ranking of features evaluating the characteristics of each of them individually. The rate criterion is the absolute value of the correlation of coefficient feature with the class. The higher the correlation, the higher the position of feature in the ranking.
- *SVMeval* - evaluates the worth of an attribute by using an SVM classifier. Attributes are ranked by the square of the weight assigned by the linear SVM classifier.
- *PCA* - performs a linear transformation of the features into another space in which features included in the new set are mutually uncorrelated and sorted with respect to the amount of input information in the classification process.
- *Forward search* - wrapper method building the set of features starting from one and gradually adding these features that provide the best quality of classification. This method is based on a classifier to be used in the target solution - in this case the k-NN.
- *Joined pairs* - a method developed by the author. It creates a ranking of features based on the ability of pairs of features for classification. In the first stage, a collection of all possible two-element subsets of features is formed. Then, basing on each subset of features a classifier is constructed and evaluated. As a result, each two-element subset is assigned a numerical value that indicates the quality of classifier built on the basis of this subset. Finally, the ranking of features is created. Features are added into in order indicated by quality of classifiers built in the previous step. The principle of operation of the method is shown in algorithm1.

The method has been tested using four selected data sets available from the UCI Machine Learning Repository[16] – glass, vote, segment challenge and wine quality. In each of the cases a rapid convergence of the level of quality classifications

Algorithm 1: Joined pairs

```
input : tf: table of features
output: fr: features ranking
// variable: pair of features
def pof: structure:
    featureA
    featureB
    quality
for each possible pairs of features from tf do
    add new pair to pof
    pof.quality ← classifierQuality(pof.featureA, pof.featureB)
Sort(pof) by pof.quality, ascending
for each pof do
    if pof.featureA ∉ fr then
        add pof.featureA to fr
    if pof.featureB ∉ fr then
        add pof.featureB to fr
return fr
```

to the maximum value was obtained, indicating that the joined pairs method works properly Figure 2 shows the graphs indicating the level of quality of classification described by the F-measure as a function of features number taken into account during the classification process. Number of features included was increased by adding features one by one, in the order indicated by the ranking produced by joined pairs algorithm.

In the study, k-NN classifier with distance weighing was used. For the distance measure the Manhattan metric was used:

$$d(X,Y) = \sum_{i=1}^{N} |X_i - Y_i|, \tag{1}$$

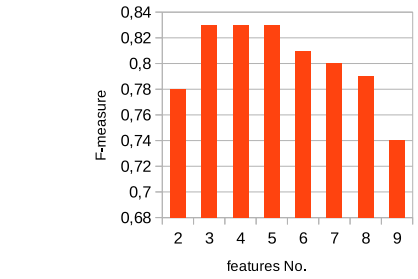
where X and Y are the points in N-dimensional space of features and d is a distance betwen these points. The tested element was assigned to a class on the basis of the vote. The weight of the vote of the $i-th$ neighbor was distance weighed according to the formula:

$$w(i) = \frac{1}{d(i) + 0.0001}. \tag{2}$$

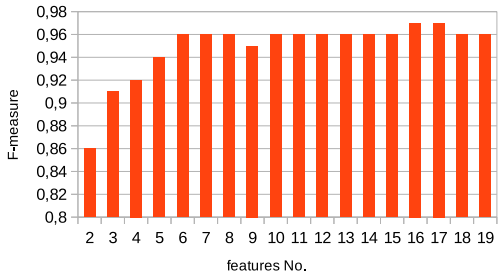
Value of 0.0001 in the denominator is added to the distance in order to avoid division by zero when the distance is equal to zero[6].

Quality rating of classification was based on the F-measure[1] indicator. The indicator can be betwen 0 and 1 and the quality of classification is the higher the F-measure value is closer to 1. The test method was 10-fold cross-validation.

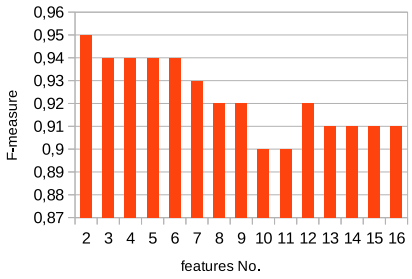
[1]F-score, F1-score



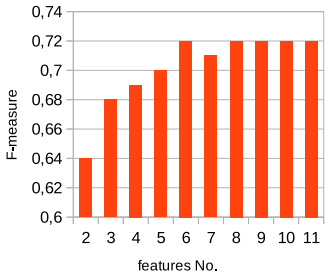
(a) glass



(b) segment challenge



(c) votes



(d) wine quality

Fig. 2: The quality of the classification as a function of the number of features for the selected data sets

IV. Results and discussion

The quality of feature sets obtained using each method was evaluated by construction of the k-NN classifier and assessment of its quality. For each set of the features, 23 subsets of features were generated, containing from 1 to 23 features. In the next subsets the features were included in the order indicated by the ranking. For each subset of features, 15 classifiers differ by a n parameter were generated. Parameter n was varied from 1 to 15. Summary of rankings of features for each method are schown in table.I.

TABLE I: Features ranking

no.	Correlation	SVMeval	PCA	Forward search	Joined pairs
1	f14	f11	v1	f3	-
2	f15	f5	v2	f13	f3,f13
3	f8	f14	v3	f11	f11
4	f16	f16	v4	y4	f1
5	f7	f13	v5	f10	f12
6	f6	f9	v6	f4	f4
7	f13	f15	v7	f1	f9
8	f9	f3	v8	f14	f2
9	f5	f8	v9	f15	f7
10	f4	f12	v10	f16	f5
11	f10	f6	v11	f9	f10
12	f12	f7	v12	f8	f8
13	f11	f10	v13	t4	y4
14	f3	f1	v14	f7	t4
15	f1	f2	v15	f12	f14
16	f2	f4	v16	fm	f15
17	fm	fm	v17	f2	fm
18	t1	y4	v18	f5	y0
19	y0	t4	v19	t1	t1
20	p4	y0	v20	p4	f16
21	t4	t1	v21	y0	f6
22	y4	p1	v22	f6	p1
23	p1	p4	v23	p1	p4

Graphical comparison of results of calculations for the classification was presented in figure 3.

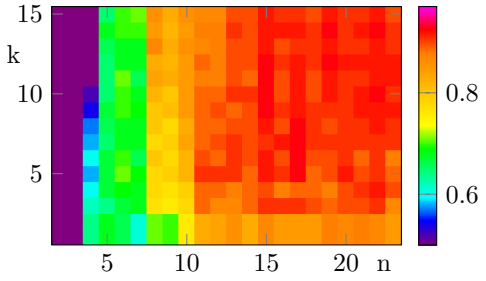
The worst result was achieved by the correlation method with its F-measure not exceeding 0.93. Not much better were SVMeval and PCA methods for which F-measure reached a value of 0.94. All the above methods have achieved the maximum quality for $n \geq 15$.

The best was the Forward search method, which reached a maximum value of F-measure equal to 0.97 for $n = 9$. In addition, a large area, stretching from $n \in \langle 8-18 \rangle$ and $k \in \langle 5-15 \rangle$, for which $F-measure \geq 0.95$ provides a good stability of the solution. Comparable in quality but far superior in the minimum amount of features was Joined pairs method. The maximum value specified by $F-measure = 0.96$ was achieved for $n = 6$ and $k = 12$.

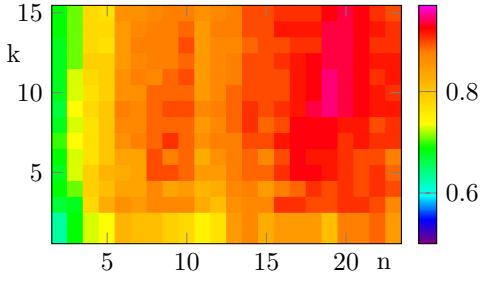
A tabular summary of the F-measure for selected values of k was presented in Table.II.

The chart shows that the *Joined Pairs* method attains the best F-measure using the smallest set of features. However, an increase in the feature count causes quality loss, which is regained only for $n = 15$ and $n = 16$. The Forward Search method achieved a stable maximum for n=9. Other schemes generated feature sets that were best for high values of n, yet none reached the quality level of *Joined Pairs* or *Forward Search*.

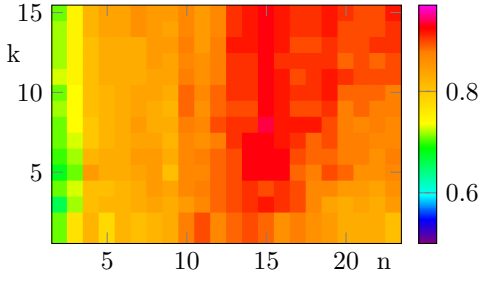
The *PCA* method allows the use of non-empirical methods for selecting the amount of features (eg. the igenvalues criterion), therefore evaluation time assumed zero. Evaluation time for *Forward search* method is zero because the evaluation of set is made up to date during the construction of the rankings.



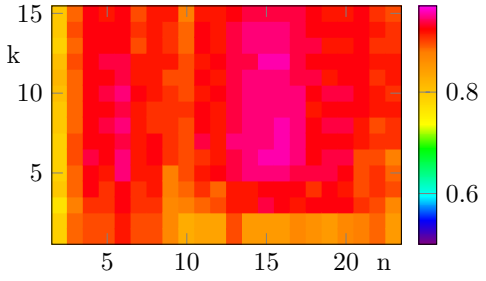
(a) Correlation



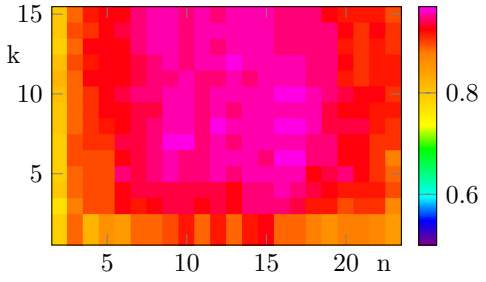
(b) SVMeval



(c) PCA



(d) Joined pairs



(e) Forward search

Fig. 3: F-measure as a function of k and n

TABLE II: F-measure

k=	12	10	6	7	12
n	Correlation	SVMeval	PCA	Forward search	Joined pairs
2	0,14	0,68	0,71	0,79	0,81
3	0,39	0,73	0,74	0,9	0,9
4	0,49	0,77	0,8	0,91	0,93
5	0,65	0,79	0,82	0,91	0,94
6	0,68	0,86	0,83	0,94	0,96
7	0,68	0,86	0,85	0,94	0,92
8	0,83	0,89	0,83	0,95	0,92
9	0,82	0,89	0,83	0,97	0,92
10	0,85	0,89	0,88	0,97	0,9
11	0,89	0,85	0,87	0,95	0,92
12	0,88	0,87	0,9	0,96	0,93
13	0,9	0,89	0,91	0,95	0,94
14	0,9	0,9	0,91	0,96	0,94
15	0,93	0,9	0,94	0,96	0,96
16	0,91	0,9	0,92	0,96	0,96
17	0,92	0,92	0,92	0,96	0,94
18	0,91	0,93	0,92	0,95	0,93
19	0,93	0,95	0,9	0,95	0,93
20	0,92	0,94	0,88	0,93	0,93
21	0,92	0,93	0,87	0,93	0,93
22	0,92	0,92	0,88	0,91	0,92
23	0,92	0,91	0,87	0,91	0,92

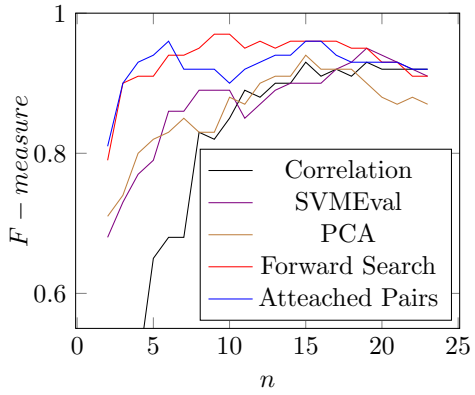


Fig. 4: F-measure as a function of n for optimal k

Table III summarizes the results for all the feature selection methods. Its first part presents the time requirements for each schema, including the time needed to generate the ranking and the time necessary to assess the quality of respective sets of features. The filtering methods (*Correlation* and *SVMeval*), unsurprisingly used the smallest amount of time. Similarly, the time requirements of the *PCA* method were negligible. The running time of both *Forward Search* and *Joined Pairs* was significantly slower. The *Forward Search* method used 276 classifiers: 23 one-feature classifiers, 22 two-feature classifiers, 21 classifiers that used three features, and so on. This was the main source of higher time requirements. The *Joined Pairs* method analyzed all the pairs of features and thus built in total 253 classifiers. Due to lower complexity of the classifiers, this approach needed less time that the *Forward Search* method. The quality of the ranking for *Correlation*, *SVMeval* and *Joined Pairs* methods was based on generation and quality assessment for all the feature sets for n from 2 to 23 and each k from 1 to 15. Since the only variable components of

TABLE III: Summary of results

	Correlation	SVMeval	PCA	Forward search	Joined pairs
rankings construction time	00:00:01	00:00:10	00:00:01	02:48:35	01:40:27
sets evaluation time	00:17:03	00:17:03	0	0	00:17:03
total time	00:17:04	00:17:13	00:00:01	02:48:35	01:57:30
optimal n	15	19	15	9	6
optimal k	12	10	8	7	12
F-measure	0,93	0,95	0,94	0,97	0,96

the ranking process were the fearure sets, the running time was the same for all of them. The *PCA* method allowed for non-empirical ways of choosing the size of feature sets (for example, the Kaiser criterion or scree plots).

As this algorithms are not computationally-heavy, their time-requirements were assumed to be zero. The running time of quality assessment for the *Forward Search* method was assumed to be zero as well, because the method does the necessary calculations online, while generating the ranking. The second part of Table 4 presents the optimal values for k and n with the respective F-measure.

All methods of feature selection achieved similar quality indicators of the constructed models.

V. Conclusion

It is possible to notice the general principle that computing power consumption feature selection algorithm translates into the quality of the obtained subsets of features. Undemanding methods of filter group indicated subsets of more features than other methods. The *Joined pairs* algorithm gives good results in the classification task.

With respect to the problem of evaluation of the arteriovenous fistula it can be concluded that the results are very good. Each of these methods has allowed to obtain a very high quality classification. It is suspected that, such optimistic results may be the effect of insufficient amount of analyzed data. Vectors describing individual patients form in the a feature space the easily separated clusters.

Verification of the results should be made on unrelated set of test data and having regard to a greater number of patients and samples.

Therefore, it would be appropriate to extend the scope of the study, increasing the set of input data.

References

[1] Marcin Grochowina, Lucyna Leniowska and Piotr Dulkiewicz, "Application of Artificial Neural Networks for the Diagnosis of the Condition of the Arterio-venous Fistula on the Basis of Acoustic Signals," *Brain Informatics and Health,* Springer, 2014, pp. 400–411.

[2] Marcin Grochowina and Lucyna Leniowska, "Comparison of SVM and k-NN classifiers in the estimation of the state of the arteriovenous fistula problem," *Computer Science and Information Systems (FedCSIS), 2015 Federated Conference on,* IEEE, 2015, pp. 249–254.

[3] Zbigniew Suraj, Neamat El Gayar and Pawel Delimata, "A rough set approach to multiple classifier systems," *Fundamenta Informaticae,* IOS Press, 2006, pp. 393–406.

[4] Mikkel Grama , Jens Tranholm Olesena , Hans Christian Riisa , Maiuri Selvaratnama and Michalina Urbaniaka, "Stenosis detection algorithm for screening of arteriovenous fistulae," *15th Nordic-Baltic Conference on Biomedical Engineering and Medical Physics (NBC 2011),* Springer, 2011, pp. 241–244.

[5] Fan, Rong-En and Chen, Pai-Hsuen and Lin, Chih-Jen, "Working set selection using second order information for training support vector machines," *The Journal of Machine Learning Research vol.6,* JMLR. org, 2005, pp. 1989–1918.

[6] "WEKA documentation," http://www.cs.waikato.ac.nz/ml/weka/documentation.html

[7] Remco R. Bouckaert, Eibe Frank, Mark Hall, Richard Kirkby, Peter Reutemann, Alex Seewald, David Scuse, "WEKA Manual," University of Waikato, 2013.

[8] Tadeusz Morzy, "Eksploracja danych - metody i algorytmy," PWN, 2013.

[9] Dymitr Ruta "Robust Method of Sparse Feature Selection for Multi-Label Classification with Naive Bayes," Computer Science and Information Systems (FedCSIS), 2014 Federated Conference on, IEEE, 2014, pp. 375–380. http://dx.doi.org/10.15439/2014F502

[10] Zdravevski, Eftim and Lameski, Petre and Kulakov, Andrea and Gjorgjevikj, Dejan "Feature selection and allocation to diverse subsets for multi-label learning problems with large datasets," Computer Science and Information Systems (FedCSIS), 2014 Federated Conference on, IEEE, 2014, pp. 387–394. http://dx.doi.org/10.15439/2014F500

[11] Daniel T. Larose, "Data mining methods and models," John Wiley & Sons, Inc, 2006.

[12] Sierra, B., Larrañaga, P., Inza, I. "K Diplomatic Nearest Neighbour: giving equal chance to all existing classes," Journal of Artificial Intelligence Research, 2000

[13] Dudani, S. A. "The distance-weighted k-nearest neighbor rule," IEEE Tran-sactions on Systems, Man, and Cybernetics, Vol. SMC-6, No. 4, 1976, pp. 325–327

[14] Fix, E., Hodges Jr., J. L. "Discriminatory analysis — nonparametric discrimination: Consistency properties," Project 21-49-004, Report No. 4, USAF School of Aviation Medicine, Randolph Field, TX, USA, 1951, pp. 261–279.

[15] Fix, E., Hodges Jr., J. L. "Discriminatory analysis — nonparametric discrimination: Small sample performance," Project 21-49-004, Report No. 11, USAF School of Aviation Medicine, Randolph Field, TX, USA, 1952, pp. 280–322.

[16] Lichman, M. "UCI Machine Learning Repository," http://archive.ics.uci.edu/ml Irvine, CA: University of California, School of Information and Computer Science 2013

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 287–290
DOI: 10.15439/2016F156
ACSIS, Vol. 8. ISSN 2300-5963

Automatic Keyword Extraction from Medical and Healthcare Curriculum

Martin Komenda[a], Matěj Karolyi[a], Andrea Pokorná[a,b]
[a] Institute of Biostatistics and Analyses,
[b] Department of Nursing,
Faculty of Medicine, Masaryk University,
Kamenice 126/3, 625 00
Brno, Czech Republic
Email: {komenda, karolyi, pokorná}@iba.muni.cz

Martin Víta
NLP Centre,
Faculty of Informatics,
Masaryk University,
Botanická 68a, 602 00
Brno, Czech Republic
Email: 333617@mail.muni.cz

Vincent Kríž
Faculty of Mathematics and Physics
Charles University
Malostranské nám. 25, 118 00
Prague, Czech Republic
Email: kriz@ufal.mff.cuni.cz

***Abstract*—Medical and healthcare study programmes are quite complicated in terms of branched structure and heterogeneous content. In logical sequence a lot of requirements and demands placed on students appear there. This paper focuses on an innovative way how to discover and understand complex curricula using modern information and communication technologies. We introduce an algorithm for curriculum metadata automatic processing -- automatic keyword extraction based on unsupervised approaches, and we demonstrate a real application during a process of innovation and optimization of medical education. The outputs of our pilot analysis represent systematic description of medical curriculum by three different approaches (centrality measures) used for relevant keywords extraction. Further evaluation by senior curriculum designers and guarantors is required to obtain an objective benchmark.**

I. Introduction

THE domain of medical and healthcare curriculum harmonization captures the systematic transmission of specialized required knowledge based on a suitable combination of theoretically focused courses and clinical teaching training [1]. It links traditional proven pedagogical approaches and medical expertise with computer sciences and technologies with a view to the discovery of an innovative way to understand the complex structure and content of medical and healthcare study programmes. The proper use of data extracted from curriculum management systems can significantly improve the global overview on entire curriculum including performing up-to-date information in real time. The attention of this paper is paid to the core technologies of automatic processing for documents classified as a supervised machine learning task called automatic keyword extraction, and its real application on curriculum metadata. Automatic Keyword Extraction (AKE) is a research topic focusing on the identification of a small set of words, key phrases, keywords, or key segments from a document that can describe the meaning of the document [2]. The aim of keyword assignment is to find a small set of terms that appropriately describes a specific document, a medical discipline in our particular case, independently of the domain it belongs to. Here are two fundamental issues, which are supposed to be answered by our research: (i) Are we able to automatically generate sets of keywords? Are the extracted keywords relevant? Do they express properly individual medical discipline? (ii) Which kind of centrality does identify the most accurate set of keywords and what is the proper value of determined threshold?

II. Methods

The research methods stems from our previously published and reviewed papers [3]–[5], wherefrom the well-known and proven step-by-step data mining guide CRISP-DM (CRoss-Industry Standard Process for Data Mining) [6]. We have used the CRISP-DM reference model, which was primarily developed by means of the effort of a consortium from the business sphere, as a powerful scientific technique to excavate the knowledge and patterns concealed in the diversely complex mountain of medical and healthcare education data [7]. This standardized methodology provides the complete process model for exploring data. Below, all the CRISP-DM steps are described in accordance with determined research questions.

A. Business understanding

In this stage, we focus on the understanding the research objectives from the perspective of curriculum mapping. In general, curriculum mapping is a procedure that creates visual representation of the curriculum based on real time information, as a way to increase collaboration and collegiality in higher education institutions. This phase also involves more detailed fact-finding about all of the resources, constraints, assumptions, published results and other factors that should be considered in determining the goals analysis [8].

Two main objectives were identified: (i) to make the curriculum more transparent to all stakeholders; (ii) to demonstrate the links between the various components of the curriculum (such as modules, disciplines, courses and learning units) [9]. In terms of new trends and reforms in medical education, we have proposed a general model for curriculum management and harmonization supported by our

original web-oriented platform called OPTIMED[1] [10]. This innovative and dynamic platform provides a clear way how to describe medical curriculum with the use of given text attributes including links to relevant study materials. We set up the structure of curriculum, which covers study programmes (e.g. General Medicine), specialized modules (e.g. Theoretic sciences), medical disciplines (e.g. Anatomy), courses (e.g. Anatomy I – lecture), and learning units (e.g. Central nervous system). This phase involves more detailed fact-finding about a systematic keyword generation of individual medical disciplines from various metadata sources stored in the OPTIMED database.

For the pilot experiment two disciplines were chosen (Nursing and Psychiatry), which are both used for the acquisition of knowledge and skills and also allow the development of so-called "soft skills". Both disciplines are taught at different periods of study. While Nursing has been lectured almost on the beginning of the medical study (in the second year) and is focused on basic knowledge and skills in caring (helping with basic needs of patients and also caring for patients with different types of illness). Psychiatry is taught as one of the last courses of undergraduate study of general medicine. Both courses include theoretical instruction in the classroom and subsequently clinical placement where students should use knowledge and skills obtained in the previous tuition and both of them are not accepted as the core medical disciplines.

B. Data understanding

Data understanding starts with initial data collection and proceeds with activities that enable to become familiar with the data, to evaluate the quality of data, to discover first insights into the data, and/or detect interesting subsets to form hypotheses regarding hidden information. The following metadata attributes describing medical curriculum was mined from PostgreSQL database and divided into the classes in accordance with fields of medicine. Names of the attributes, data types in squared brackets and samples are shown below in Table I.

TABLE I.
SELECTED ATTRIBUTES OF A CURRICULUM

Attribute (data type)	**Sample value**
Name of learning unit (varchar)	Biologic therapy in psychiatry
Importance of learning unit (text)	The aim of the study unit is to introduce types of biological therapy in psychiatry, including psychopharmacotherapy, electroconvulsive therapy, Recently, modern neurostimulative methods such as repetitive transcranial magnetic stimulation …
Description of learning unit (text)	Psychopharmaceuticals can be classified in several ways. Lehman s diversification is often used and is based on effects on three mental functions: vigilance, effectivity and mental integration (thinking). Vigilance is positively influenced by nootropics, cognitives and psychostimulants, negatively by hypnotics…
Group learning outcome (varchar)	Main indications, differences from adults (indications, efficacy), adverse effects, ethical aspects.
Index (varchar)	Psychopharmacological drugs in relation to children, Antipsychotics, antidepressants, stimulants, thymoprophylactics.
Learning outcome (varchar)	Student knows key indicators relating to psychopharmacological drugs.

C. Data preparation

Data preparation covers all activities needed to construct the final dataset from the initial raw data. First of all, table, record and attribute selection as well as data pre-processing covering transformation and data cleaning procedure were automatically done. The input data set consists of information-rich attributes (name, importance, description of learning units and all related learning outcomes including indexes) mined from OPTIMED. The data preparation phase appears again at the end of the modeling process described in the next section. An output of an algorithm for keyword extraction is very compact list of keywords including just the self-sorting information. For the purposes of plotting the graph the simplified table combined with products is created during the calculation. This final data connection is executed every time when a user starts to explore new discipline keyword dataset.

D. Modeling

In this section, we propose a novel algorithm for keyword extraction that forms the basis of the modeling and visualization stages, which are also introduced here.

The algorithm is based on two main components: word2vec model and network/graph centralities. Word2vec model proposed by Mikolov [11] is a word embedding model that belong to a wide class of distributed representations models. It arises from predicting the neighbors of words using a deep neural network – roughly said, the weights in the neural network between input and hidden layer constitute the vector representations of words. For our purposes, we have created a word2vec model over TC wikipedia[2] using the following main parameters: model = CBOW, dimension = 200. The concept of selected network/graph centrality measures was firstly developed in social network analysis. We successfully used these methods in our previous work [3], [4], namely we deal with the closeness, betweenness, and eigenvector centralities.

Input data for the algorithm are represented by trained word2vec model, given document (bag of words for individual medical disciplines), word similarity threshold `t`, number of keyword `n`. (1) Select all terms that are contained in the document at least two times. (2) For all pairs of terms, compute the cosine similarity of their word2vec representations. (3) Create the graph `G` in a following way: Vertices are the terms (obtained in the Step 1). Two vertices are connected with an edge if and only if their mutual cosine

[1] http://opti.med.muni.cz/en/

[2] http://nlp.cs.nyu.edu/wikipedia-data/

similarity is at least `t`. (4) Compute closeness, betweenness and eigenvector centrality of vertices in the graph `G`. Select `n` terms with the highest values of various centrality and set them up as keywords.

Finally, we have implemented a new online dynamic visualization as a feature of the OPTIMED reporting tools[3], which is based on outputs from described algorithm above. Innovative graphical interpretation allows users to create complex data overview, which cannot be achieved with basic data views. In our case, the simple network visualization was created with the D3.js JavaScript library. We decided to integrate a force-directed graph (see Fig. 1) displaying complex structures with expand-on-demand clusters and convex hulls around leaf nodes [12].

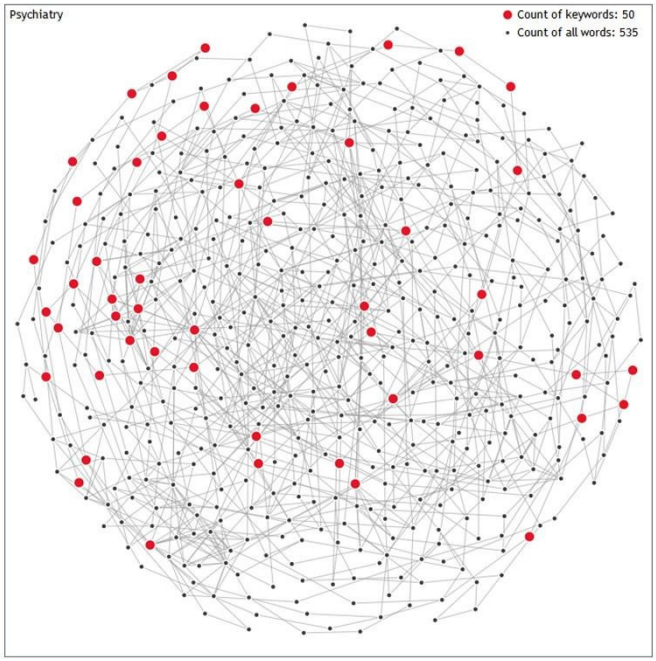


Fig. 1 The force-directed graph of Psychiatry (based on closeness centrality, cosine similarity is higher than 0.6)

The graphs include two types of keyword representation – red nodes (bigger) for more frequent keywords and black nodes (smaller) with less frequent keywords. Moreover, we wanted to determine what kind of centrality type (closeness, betweenness or eigenvector) and what value of the cosine similarity (> 0.6; > 0.5; or > 0.4) would be the most suitable and could the most accurately describe concrete medical discipline. Using advanced filters included in web reporting service, we are able to modify the final visual interpretation of available preprocessed data in accordance with selected parameters – medical discipline, centrality type, cosine similarity threshold (see Fig. 2).

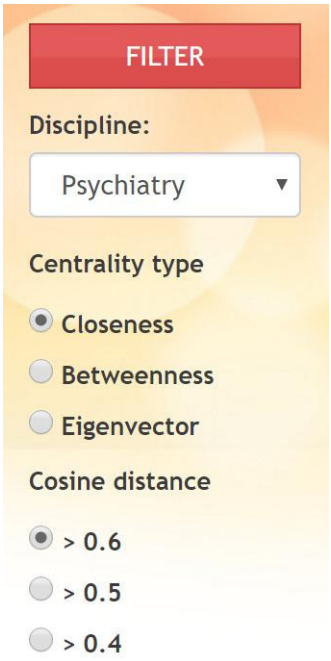


Fig. 2 The force-directed graph filter

E. Evaluation and Deployment

A checking procedure is performed in this stage in order to find the right meaning of analytical outputs. The obtained results were verified by representatives of the faculty management, in order to confirm the final interpretation [3]. Based on the assessment of the graphical presentation of the subject Psychiatry we decided that the best threshold for cosine similarity is 0.6 (see Fig. 1) and the most suitable centrality type is closeness. Table II shows keywords concerning different type of centrality. As you can see the first column (type of centrality – closeness) includes keywords especially in view of the clinical description of psychiatric illness, the second column (type of centrality – betweenness) cannot be held as summary of clear signs of typical keywords rather confusing mix of terms and third column (type of centrality – eigenvector) includes keywords focused on describing the pathophysiology and chemical substances important in Psychiatry, followed by keywords of clinically important symptoms of disorders as well as therapeutic modalities and drugs. So, different type of centrality enables to view the discipline from different perspectives through the different sample of keywords. What has to be highlighted that when we used higher cosine similarity we could identify less duplicate terms and terms which are not describing the key issues (as nouns or verbs) and we could identify them as „stop-words“ without any informative value.

[3] http://opti.med.muni.cz/en/reporting/web/analyticke-reporty/extrakce-klicovych-slov/dis-44

TABLE II.
TOP TWENTY KEYWORDS FOR PSYCHIATRY ACCORDING TO INDIVIDUAL CENTRALITIES.

Rank	Closeness	Betweenness	Eigenvector
1	disorders	neurological	dopamine
2	neurological	dysfunction	neurotransmitter
3	schizophrenia	stimulation	acetylcholine
4	disorder	inhibition	serotonin
5	anxiety	serotonin	neurotransmitters
6	psychological	anxiety	receptors
7	symptoms	antidepressant	receptor
8	dysfunction	brain	neurons
9	psychopathology	cortex	noradrenaline
10	brain	disorders	Presynaptic
11	illnesses	symptoms	inhibition
12	epilepsy	disorder	reuptake
13	mental	neurons	neuron
14	cognitive	cognitive	cells
15	stimulation	psychological	hippocampus
16	behavioral	schizophrenia	antidepressant
17	pathological	perception	amygdala
18	clinical	clinical	cortex
19	psychiatric	particular	prefrontal
20	emotional	emotional	benzodiazepines

III. Discussion

The possibility to graphically visualize and interpret medical curriculum and content of the individual courses through keywords from different scientific disciplines allow teachers to identify the main issues discussed in the concrete tuition. The identification could help them to explore whether the main keywords correspond to any other keywords as descriptors of the main themes or major topics representing individual similarly or differently oriented teaching units. We have used the dynamic visualization using force-directed graph with clusters of key points (point graphs) and tables while we have identified 50 keywords as a border (maximum evaluated number of keywords for one visualization).

We could say that generally description of Nursing disciplines is not so straightforward, clear and simply as for Psychiatry. This can be explained mainly by the fact that the content of the discipline is very inhomogeneous. From the description of the summary of keywords for nursing evidently arise higher amount of duplicated terms and "empty" or so-called "stop words", which do not bring additional value for improvement of orientation in the curriculum content not only for teachers neither for students. This fact should not affect the view of usability of curriculum content visualization but rather to encourage the prudent use and analyses of the information gathered and from data mining and also in feedback when assessing the homogeneity of data.

IV. Conclusion and Future Work

In this work we have proposed a novel method for identifying key terms from a free text covering medical curriculum. This method is based on computing node centralities in a similarity graph of terms contained in the given medical and healthcare discipline description, whereas the similarity is obtained by a popular word2vec model. First results seem to be promising in terms of face validity. Methodologically sound evaluation of this method and comparison with traditional methods of keyword extraction - even on different domains - is a current issue. Centrality measures in word2vec graphs can also serve as features in supervised machine learning algorithms for keyword extraction. In the near future this approach is also planned to be investigated.

Acknowledgment

The authors were supported from the following grant projects: (i) CROESUS – clinical reasoning skills enhancements with the use of simulations and algorithms – reg. no.: 2014-1-CZ01-KA203-002002, which is funded by the European Commission ERASMUS+ program; (ii) MEDCIN – medical curriculum Innovations – reg. no.: 2015-1-CZ01-KA203-013935, which is funded by the European Commission ERASMUS+ program; and (iii) SVV – specific academic research – reg. no.: 260 333, which is funded by the Charles University in Prague.

V. References

[1] M. Komenda, "Towards a Framework for Medical Curriculum Mapping," Doctoral thesis, Masaryk University, Faculty of Informatics, 2015.

[2] A. Hulth, "Improved automatic keyword extraction given more linguistic knowledge," in Proceedings of the 2003 conference on Empirical methods in natural language processing, 2003, pp. 216–223.

[3] M. Víta, M. Komenda, and A. Pokorná, "Exploring medical curricula using social network analysis methods," presented at the Federated conference on computer science and information systems, 5th International Workshop on Artificial Intelligence in Medical Applications, Lodz, 2015, doi:10.15439/2015F312.

[4] M. Komenda, M. Víta, C. Vaitsis, D. Schwarz, A. Pokorná, N. Zary, and L. Dušek, "Curriculum Mapping with Academic Analytics in Medical and Healthcare Education," PloS One, vol. 10, no. 12, 2015, doi:10.1371/journal.pone.0143748.

[5] M. Komenda, D. Schwarz, J. Švancara, C. Vaitsis, N. Zary, and L. Dušek, "Practical use of medical terminology in curriculum mapping," Comput. Biol. Med., 2015, doi: 10.1016/j.compbiomed.2015.05.006.

[6] A. I. R. L. Azevedo, "KDD, SEMMA and CRISP-DM: a parallel overview," 2008.

[7] S. C. Chen and M. Y. Huang, "Constructing credit auditing and control & management model with data mining technique," Expert Syst. Appl., vol. 38, no. 5, pp. 5359–5365, May 2011, doi: 10.1016/j.eswa.2010.10.020.

[8] P. Chapman, J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer, and R. Wirth, "CRISP-DM 1.0 Step-by-step data mining guide," 2000.

[9] R. M. Harden, et al, "Curriculum mapping: a tool for transparent and authentic teaching and learning". Association for Medical Education in Europe, 2001, doi: 10.1080/01421590120036547.

[10] M. Komenda, D. Schwarz, J. Hřebíček, J. Holčík, and L. Dušek, "A Framework for Curriculum Management - The Use of Outcome-based Approach in Practice," in Proceedings of the 6th International Conference on Computer Supported Education, Barcelona, 2014, doi:10.5220/0004948104730478.

[11] T. Mikolov, W. Yih, and G. Zweig, "Linguistic Regularities in Continuous Space Word Representations.," in HLT-NAACL, 2013, pp. 746–751.

[12] M. Bostock, V. Ogievetsky, and J. Heer, "D3 data-driven documents," Vis. Comput. Graph. IEEE Trans. On, vol. 17, no. 12, pp. 2301–2309, 2011, doi: 10.1109/TVCG.2011.18.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 291–297
DOI: 10.15439/2016F445
ACSIS, Vol. 8. ISSN 2300-5963

HD: Efficient Hand Detection and Tracking

C. Rouge, S. Shaikh, and J.I. Olszewska
School of Computing and Technology
University of Gloucestershire
Cheltenham, United Kingdom
Email: joanna.olszewska@ieee.org

***Abstract*— Automated hand detection is useful for applications requiring reliable hand posture and hand gesture processing. Such applications include human-computer/human-robot interfaces for rehabilitation, serious games, or non-invasive medical diagnosis. Hence, in this paper, we focus on the design and development of a robust and fast hand detection and tracking (HD) system. The design of our HD system involved the study of the human skin colour and people's foreground properties, in order to merge these information for an efficient hand detection and tracking. Experiments have been carried out in real-world environment and have demonstrated the excellent performance of our HD system.**

I. Introduction

Hand detection and tracking are the actions of automatically locating and following human hands to extract information useful for various applications, which involve interactions of humans with computers (HCI) [1] through interfaces such as mouses [2], gloves [3], interactive tables [4], or natural gestures [5].

Automated hand detection could be applied for multimodal interactions of humans within augmented-reality frameworks, e.g. game control [6], character animation [7], etc., or it could be used for secure and safe interactions of humans within real-world environment. Indeed, hand detection could be used in biometrics [8], since palms of the human hands contain unique patterns of ridges and valleys, in the same way that fingerprints, whereas presenting much larger areas and thus being even more distinctive than the fingerprints [9]; in surveillance for abnormal behaviour detection [10] or in crime prevention [11].

Medical applications involving the monitoring of human interaction with robots (HRI) [12] can embed hand detection, e.g. for functional rehabilitation [13] such as exercises of patient's interaction with objects s/he grasps [14] to improve stroke recovery [15].

Moreover, hand detection and tracking is promising in diagnosis of neurologic disorders such as Parkinson's disease. Indeed, this progressively degenerative movement disorder [16] implies quantitative evaluations of hand movement as well as hand tremor for neurological examination [17]. These measurements could be based on the visual tracking of patient's hand motion [18] or on the paced finger tapping test [19] assessed with visual rating scales such as the Unified Parkinson's Disease Rating Scale (UPDRS), rather than on the traditional invasive methods such as electroencephalograms (EEGs) [20].

Other applications of hand detection and gesture recognition could improve the communication and interaction of people presenting difficulties [21]. For example, hand detection in context of sign language could help deaf people naturally interacting with machines or non-deaf people interacting with deaf ones [22].

The main challenge of vision-based hand detection is to cope with the large variability of human hand's appearance due to a huge number of degrees of freedom (DoFs) of the hand's movements, to different skin-colour possibilities as well as to the variations in view points, scales, and speed of the camera capturing the scene [21].

Hence, hands have been represented as 2D [23] or 3D models [24], and various methods have been proposed, involving eigen-based hand model [25], hand/not-hand SVM classifier [26] or AdaBoost-based hand detector [27], to automatically recognize human hands in images or videos. Despite their effectiveness, these methods lack of detection accuracy in some real situations.

In this work, the adopted approach for hand detection is focused on robustness, and it combines threshold-based colour detection with background subtraction. Moreover, enhanced Adaboost face detection [28] is used to differentiate hands from the human face region, while hand tracking is achieved by adding the foreground extraction obtained from one frame to another.

Hands could be visually detected or tracked by a range of sensors [29], such as ego-centric/self-mounted cameras [30], stereo camera [31], or Kinect [32]. In this paper, we focus on hand detection and tracking solely based on a single, static camera feed, which is a non-invasive and low-cost solution, leading to a convenient, computer-vision-based system.

The product developed in this work is capable of both effectively detecting human hands without building a hand model, but instead applying pixel thresholding based on a skin-colour model, and efficiently tracking an individual's hand movement based on the foreground blob and background subtraction information. This HD system could be used for numerous real-world purposes, and in particular for medical applications.

The original contribution of our work is twofold and consists in the design and development of a new hand detection and tracking system based on features such as colour and motion as well as in a study of the colour skin appearance within YCbCr colour space.

978-83-60810-90-3/$25.00

The paper is structured as follows. In Section II, we present our hand detection and tracking (HD) system, while in Section III we report and discuss the carried out experiments which results show the developed HD system has excellent performance on real-world video datasets. Conclusions are drawn up in Section IV.

II. Our Method

The developed HD system as depicted in Fig. 1 uses skin-colour thresholding as explained in Section II-A. Foreground detection and background subtraction are described in Section II-B. Head detection is computed in order to discard that region to improve hand detection accuracy in case of full-body tracking as mentioned in Section II-C. Moreover, the fusion of all the computed information to detect the hands as well as the tracking process of the hand are presented in Section II-D.

A. Skin-Colour Detection

Let us consider a colour image or video frame $I(x, y)$ with M and N, its width and height, respectively.

At first, our system performs skin thresholding [33]. Among major colour spaces such as RGB [34], HSV [35], or YIQ [36], our system uses YCbCr colour space for thresholding to distinguish between skin and non-skin colours. Indeed, the YCbCr colour space is the most popular choice in skin-colour detection methods, because its luminance component Y as well as chrominance components Cb and Cr are separated and could be easily computed from RGB values [37]. This RGB-YCbCr transformation possibility as well as the explicit separation of luminance and chrominance components, which furthermore brings some degree of robustness to illumination variations [38], make this colour space attractive for skin-colour modeling [39].

Moreover, YCbCr colour space is used in our work due to the fact it has a good skin feature clustering as different human skin colours from different races fall in a compact region in the YCbCr colour space [40]. On the opposite, the RGB colour space, which is the basic colour space of image processing, cannot be segmented by simple thresholding [41], because RGB colour space has perceived non-uniformity [42]. Although it may be possible to detect one skin colour using other colour spaces, it would be not possible to detect with the RGB colour space the perceived colour of human skin which varies greatly across human races or even between individuals of the same race. So, to increase invariance against illumination variability, our HD system operates in the YCbCr colour space in order to approximate the chromaticity of skin rather than its apparent colour value.

Indeed, in our system, the distribution of skin pixels values is highlighted based on Cb and Cr components [40] as an additional threshold conditions reported in Fig. 2, whereas the luminance component is eliminated to remove the effect of shadows, illumination changes, and modulations of orientation of the skin surface relative to the light source(s) [38].

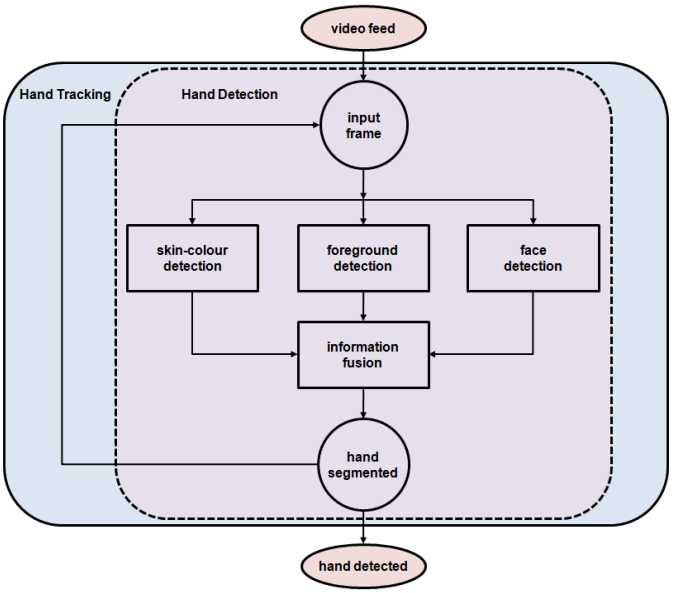


Fig. 1. Overview of our hand detection and tracking (HD) system.

So, the adopted YCbCr colour threshold detection values are as follows:

$$Cb = 0.148 \times R - 0.291 \times G + 0.439 \times B + 128, \quad (1)$$

$$Cr = 0.439 \times R - 0.368 \times G - 0.071 \times B + 128, \quad (2)$$

where the skin region is segmented based on the thresholding values [43] defined as $140 \leqslant Cb_{skin} \leqslant 195$ and $140 \leqslant Cr_{skin} \leqslant 165$.

B. Foreground Detection

To extract the foreground blob, we combine frame difference and background subtraction techniques [44]. This consists in computing in parallel, on one hand, the difference between a current frame $I_k(x, y)$ and the previous one $I_{k-1}(x, y)$, and on the other hand, the difference between the current frame $I_k(x, y)$ and a background model of the scene, and afterwards, to combine both results in order to extract the foreground in the corresponding view [44]. It is worth noting that background subtraction is based on the assumption that the camera does not move with respect to the static background.

To model the background, we adopt the running Gaussian average (RGA) [45], which is suitable for real-time tracking, and which is characterized by the mean μ_b and the variance $(\sigma_b)^2$.

Hence, the foreground is determined by

$$F(x, y) = \begin{cases} 1 & \text{if } |F_f(x, y) \cup F_b(x, y)| = 1, \\ 0 & \text{otherwise,} \end{cases} \quad (3)$$

with

$$F_f(x, y) = \begin{cases} 1 & \text{if } |I_k(x, y) - I_{k-1}(x, y)| > tf, \\ 0 & \text{otherwise,} \end{cases} \quad (4)$$

and

$$F_b(x,y) = \begin{cases} 1 & \text{if } |I_k(x,y) - \mu_b| > n \cdot \sigma_b, \\ 0 & \text{otherwise,} \end{cases} \quad (5)$$

where tf and $n \in \mathbb{N}_0$ are the thresholds for frame difference and background subtraction processes, respectively [44].

Finally, to compute a blob defined by labeled connected regions, morphological operations [4], [46], such as dilation, filling, etc. are applied to the extracted foreground F, in order to exploit the existing information on the neighboring pixels, in the frame:

$$f(x,y) = Morph(F(x,y)). \quad (6)$$

C. Face Detection

Face detection is performed using the lighting-variable Adaboost algorithm [28], relying on global image intensity [28] and Haar-like features [47] rather than face skin-colour as in [48], in order to robustify the HD system against illumination changes.

Hence, based on the gray image $I_g(x,y) = 0.299\,R(x,y) + 0.587\,G(x,y) + 0.114\,B(x,y)$,, the average value $I_{g_{AVG}}$ of the global image intensity is defined as

$$I_{g_{AVG}} = \frac{1}{M\,N} \sum_{x=1}^{M} \sum_{y=1}^{N} I_g(x,y), \quad (7)$$

while local Haar-like features f [47] encode the existence of oriented contrasts between regions in the processed image, and are computed by subtracting the sum of all the pixels of a subregion of the local feature from the sum of the remaining region of the local feature using the integral image representation $II(i,j)$ which is defined as follows:

$$II(i,j) = II(i-1,j) + II(i,j-1) - II(i-1,j-1) + I(i,j), \quad (8)$$

where $I(i,j)$ is the pixel value of the original image at the position (i,j).

This detection system requires a training phase during which it builds strong classifiers based on cascades of weak classifiers [28]. In particular, to form a T-stage cascade, T weak classifiers are selected using the AdaBoost algorithm [47]. In fact at a t stage of this cascade, a sub-window u of an image from the training set is computed by Eq. (8) and it is passed to the corresponding t^{th} weak classifier. If the region is classified as a non-face, the sub-window is rejected. If not, it is passed to the $t+1$ stage, and so forth. Consequently, more stages the cascade owns, more selective it is, i.e. less false positive detections occur. However, that could lead to the increase in the number of false negative detection.

In order to select at each t level (with $1 < t < T$) the best weak classifier, an optimum threshold θ_t is computed by minimizing the classification error due to the selection of a particular Haar-like feature value $f_t(u)$. The resulting weak classifier k_t is thus obtained as follows:

$$k_t(u, f_t, p_t, \theta_t) = \begin{cases} 1 & \text{if } p_t f_t(u) < p_t \theta_t, \\ 0 & \text{otherwise,} \end{cases} \quad (9)$$

where p_t is the polarity indicating the direction of the inequality.

Then, a strong classifier $K_T(u)$ is constructed by taking a weighted combination of the selected weak classifiers k_t according to

$$K_T(u) = \begin{cases} 1 & \text{if } \sum_{t=1}^{T} \alpha_t k_t(u) \geq \frac{1}{2} \sum_{t=1}^{T} \alpha_t, \\ 0 & \text{otherwise,} \end{cases} \quad (10)$$

where $\frac{1}{2} \sum_{t=1}^{T} \alpha_t$ is the AdaBoost threshold and α_t is a voting coefficient computed based on the classification error in each stage t of the T stages of the cascade [47].

Next, the lighting-adaptable strong super-classifier $\mathcal{K}(u)$ is defined as

$$\mathcal{K}(u) = \begin{cases} K_L(u) & \text{if } I_{g_{AVG}} > I_{g_{th}}, \\ K_D(u) & \text{if } I_{g_{AVG}} \leq I_{g_{th}}, \end{cases} \quad (11)$$

where $I_{g_{th}}$ is the global image intensity threshold and where D and L (with $D \geq L$) are the numbers of the weak classifiers for dark and light images, respectively.

In this way, the system allows to automatically select the number of stages of the AdaBoost cascade accordingly to the lighting conditions expressed in Eq. (11) by $I_{g_{AVG}}$ and could be applied on any new input frame [28].

D. Hand Detection and Tracking

Information fusion between the skin-colour detection, foreground detection, and additional face detection is performed in order to detect the hand(s) as illustrated in Fig. 1. Indeed, the skin detection (Section II-A) allows the HD system to remove the forearm from the detected foreground (Section II-B), while the hand blob complements the skin-detected region in order to have a precise detection of the human hand(s) as shown in Fig. 3 (d). In case a face is detected (Section II-C) in a frame containing the full body, this head region is discarded to avoid confusion between regions containing skin colour and to extract only hands' regions as in Fig. 4 (d).

As the detection method is fast enough to operate at image acquisition frame rate, it can be used for hand tracking. Tracking hands is difficult since hands can move very fast and their appearance can change vastly within a few frames. Tracking hands is defined in the HD system as the frame-to-frame correspondence of the segmented hand regions (see Fig. 1). Tracking provides the inter-frame linking of hand/finger appearances. This provides trajectories of features in time. These trajectories convey essential information regarding the gesture [38] and might be used either in a raw form (e.g. in hand-guided control applications) or after further analysis (e.g. in hand gesture recognition).

III. Experiments and Discussion

Our HD system and its components have been tested on real-world videos captured at 25f/s and with a resolution of 640x360 pixels. All the experiments have been carried out on a computer with an Intel Core (TM) 2 Duo CPU @2.5GHz processor, with a 2Gb RAM, and running MatLab software.

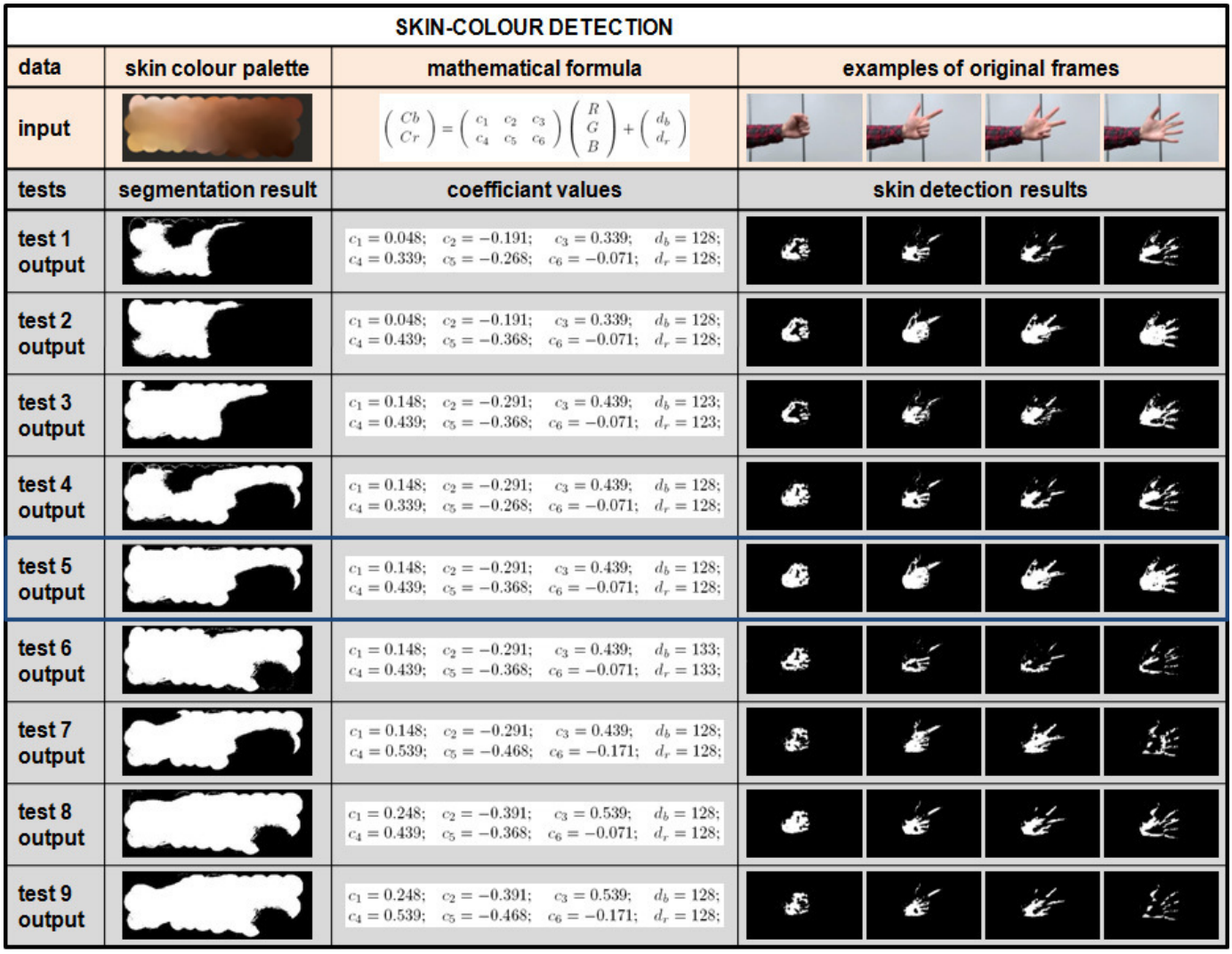


Fig. 2. Results of the skin-colour modeling and detection tests with different coefficient values.

A. *Experiment 1*

The first experiment consists in skin-colour detection and it evaluates the colour distributions from the hand in order to improve overall detection performance as illustrated in Fig. 2. Indeed, in the YCbCr colour space, Cb and Cr channels are employed owing to their relative insensitivity to the lighting variations [49].

In test 1, decreasing both $c1 - c3$ and $c4 - c6$ coefficient values came to no positive effect, showing a decrease in lighter and darker skin tones compared to the original palette, and resulting in poor hand detection. In test 2, decreasing $c1 - c3$ threshold values reduces system's capability to detect darker skin regions, and thus is not recommended to use with different skin tones and environmental changes e.g. light, shading and backgrounds. The purpose of the test 3 is to determine if changing the end values for both $c1 - c3$ and $c4 - c6$ thresholds would have any effect on results. The difference is easy to see between this test and the original, as the darker skin region is not significantly detected. In test 4, decreasing the $c4 - c6$ values demonstrated no positive changes to that of the original one. Lighter skin tones are not as visible and darker ones have not particularly changed. Once again, this can be demonstrated within frame results as, in this case, half of the hand seems missing in some frames. In test 5, this model accepts a wide range of skin colours, although it can be seen within the skin colour palette that it has had issues with detecting darker skin regions. In test 6, most of skin regions of the palette have been detected. However, within the frames tested, one can see that less of the hand is detected. Within test 7, darker regions have had minimal change between this and the original palette, but it is the lighter skin tones that have lost parts. This is demonstrated within the frame results as the hand in some frames becomes barely visible. In test 8, it is noticeable that the detection of darker region has increased, although it has also decreased detected regions within lighter skin tones. This can also be highlighted within the same test on the running frames, as in many parts of the hands, detection is missed. In test 9, by increasing both $c1 - c3$ and $c4 - c6$ coefficient

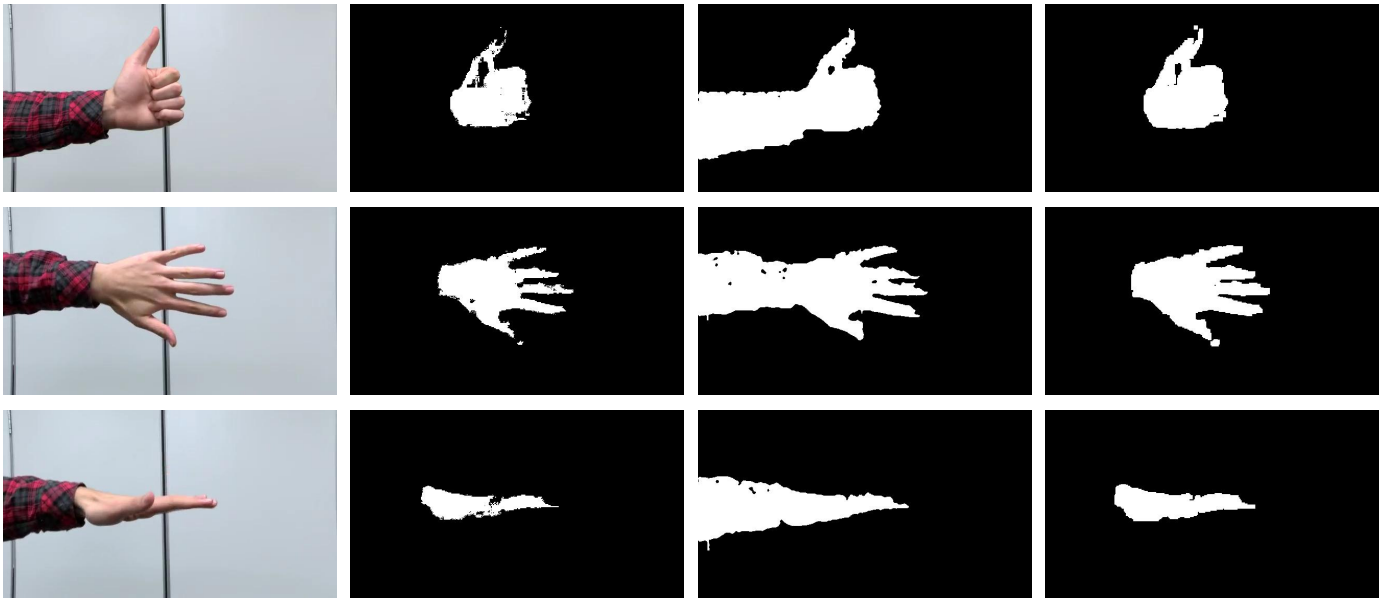


Fig. 3. Samples of our HD system for (a) input frame (1st row: frame 91, 2nd row: frame 288, 3rd row: frame 374); (b) skin-colour detection; (c) foreground detection; (d) hand detection.

values at the same time, there is a decrease within the detection of lighter skin tones and an increase in detection of darker skin tones. However, it is very prone to noise as demonstrated within each of the tested frames.

Finally, the adopted model is the one from the test 5, since the palette has been well respected and the hand has been the most clearly detected in all frames.

B. Experiment 2

Experiment 2 considers the hand detection in sequences without visible face. So, it tests the skin-colour detection and the foreground detection as well as the fusion of the related information for the hand detection.

Moreover, the background noise removal value was tested between the range of 0 and 1000, and morphological methods were also attempted. The skin colour was set to range of values in proportion to standard skin colours as described in Section III-A, the challenge being to detect darker shades of skin, nails, red tints within skin and shadows.

Results demonstrate the effectiveness of our HD system, since at all time the hand (Fig. 3(a)) is correctly detected (Fig. 3(d)), leading to 100% accuracy. Figure 3 clearly shows the blobs of the skin-detected wrist (Fig. 3(b)) and the foreground forearm (Fig. 3(c)) as well as the combined final result provind the detected hand (Fig. 3(d)) in each frame.

C. Experiment 3

The third experiments are testing the entire HD system for scenarios where other commonly visible, skin-colour-like body parts appear in the frames (Fig.4(a)), the face for instance. Hence, the HD system computes the skin colour objects (Fig. 4(b)), the foreground (Fig. 4(c)) as well as the face (Fig. 4(d)) which is detected within the frame by the lighting-variation-robust cascade object detector. Then, the HD system uses this information to remove the detected face blob, while the remaining skin-colour objects and the foreground are combined together in order to allow accurate hand detection (Fig. 4(e)).

The hand itself has been detected at all times, which gives a 100% detection rate, which is an excellent results compared to typical values for hand detection using blob approach (96.77%) [50], Histogram of Gradient (HOG) method (94.40%) [51], and Skin Colour Histogram of Gradient (SCHOG) technique (97.80%) [51]. Moreover, the running time of our overall HD system has been assessed, and it allows real-time hand tracking.

IV. Conclusions

The human hand has crucial importance to many actions that may occur in a person's day-to-day life, e.g. person-to-person interactions or person-to-object interactions. Hence, this paper focuses on the automatic detection of hands, with the use of background subtraction, foreground detection, and skin-colour thresholding methods. With these techniques combined, our HD system is capable of detecting human hands in a number of different scenarios such as person hand detection and tracking for Parkinson's disease diagnosis or for stroke recovery monitoring. Other potential applications could be 3D hand tracking [52] and gesture control for serious game [53], HRI and rehabilitation interactions [54] as well as individual finger processing for sign-language communication between deaf/non-deaf people [55].

(a) (b) (c) (d) (e)

Fig. 4. Samples of our HD system for (a) input frame; (b) skin-colour detection; (c) foreground detection; (d) face detection; (e) hand detection.

REFERENCES

[1] V. I. Pavlovic, R. Sharma, and T. S. Huang, "Visual interpretation of hand gestures for human-computer interaction: A review," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 677–695, 1997.

[2] R. Suriya and V. Vijayachamundeeswari, "A survey on hand gesture recognition for simple mouse control," in *Proceedings of the IEEE International Conference on Information Communication and Embedded Systems*, 2014, pp. 1–5.

[3] A. Aristidou and J. Lasenby, "Motion capture with constrained inverse kinematics for real-time hand tracking," in *Proceedings of the IEEE International Symposium on Communications, Control and Signal Processing*, 2010, pp. 1–5.

[4] N. J. Kim, S. Suh, and C. Choi, "Robust finger contact detection with majority quadrant search for interactive tabletop displays," in *Proceedings of the IEEE International Conference on Consumer Electronics*, 2014, pp. 518–519.

[5] M. Refice, M. Savino, M. Adduci, and M. Caccia, "Automatic classification of gestures: A context-dependent approach," in *Proceedings of the IEEE Federated Conference on Computer Science and Information Systems (FedCSIS)*, 2011, pp. 743–750.

[6] A. T. S. Chan, H. V. Leong, and S. H. Kong, "Real-time tracking of hand gestures for interactive game design," in *Proceedings of the IEEE International Symposium on Industrial Electronics*, 2009, pp. 98–103.

[7] A. H. J. Moreira, S. Queiros, J. Fonseca, P. L. Rodrigues, N. F. Rodrigues, and J. L. Vilaca, "Real-time hand tracking for rehabilitation and character animation," in *Proceedings of the IEEE International Conference on Serious Games and Applications for Health*, 2014, pp. 1–8.

[8] S. Ben Jemaa, M. Hammami, and H. Ben-Abdallah, "Contactless hand detection in complex image based on data-mining process," in *Proceedings of the IEEE Conference on Computer Systems and Applications*, 2013, pp. 1–4.

[9] A. N. Kataria, D. M. Adhyaru, A. K. Sharma, and T. H. Zaveri, "A survey of automated biometric authentication techniques," in *Proceedings of the IEEE Nirma University International Conference on Engineering*, 2013, pp. 1–6.

[10] W. Kong, A. Hussain, M. H. M. Saad, and N. M. Tahir, "Hand detection from silhouette for video surveillance application," in *Proceedings of the IEEE International Colloquium on Signal Processing and its Applications*, 2012, pp. 514–518.

[11] E. Dente, J. Ng, A. Vrij, S. Mann, A. Bull, and A. Bharath, "Tracking small hand movements in interview situations," in *Proceedings of the IEEE International Symposium on Imaging for Crime Detection and Prevention*, 2005, pp. 55–60.

[12] C. D. Lim, C.-M. Wang, C.-Y. Cheng, Y. Chao, S.-H. Tseng, and L.-C. Fu, "Sensory cues guided rehabilitation robotic walker realized by depth image-based gait analysis," *IEEE Transactions on Automation Science and Engineering*, vol. 13, no. 1, pp. 171–180, January 2016.

[13] N. Nordin, M. R. Arshad, U. Soori, and N. M. Yin, "Virtual input using skin color model for robotic platform control," in *Proceedings of the IEEE International Conference on Signal and Image Processing Applications*, 2009, pp. 305–311.

[14] G. P. Rosati Papini, M. Fontana, and M. Bergamasco, "Desktop haptic interface for simulation of hand-tremor," *IEEE Transactions on Haptics*, vol. 9, no. 1, pp. 33–42, 2016.

[15] L. Shires, S. Battersby, J. Lewis, D. Brown, N. Sherkat, and P. Standen, "Enhancing the tracking capabilities of the Microsoft Kinect for stroke rehabilitation," in *Proceedings of the IEEE International Conference on Serious Games and Applications for Health*, 2013, pp. 1–8.

[16] G. V. Kondraske and R. M. Stewart, "Web-based evaluation of Parkinson's disease subjects: Objective performance capacity measurements and subjective characterization profiles," in *Proceedings of the IEEE Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2008, pp. 799–802.

[17] R. LeMoyne, C. Coroian, and T. Mastroianni, "Quantification of Parkinson's disease characteristics using wireless accelerometers," in *Proceedings of the IEEE International Conference on Complex Medical Engineering*, 2009, pp. 1–5.

[18] J. Ide, T. Sugi, N. Murakami, F. Shima, H. Shibasaki, and M. Nakamura, "Quantitative evaluation of hand movement on visual target tracking for patients with Parkinson's disease," in *Proceedings of the IEEE International Conference on Complex Medical Engineering*, 2007, pp. 1896–1900.

[19] S. Das, L. Trutoiu, A. Murai, D. Alcindor, M. Oh, F. De la Torre, and J. Hodgins, "Quantitative measurement of motor symptoms in Parkinson's disease: A study with full-body motion capture data," in *Proceedings of the IEEE International Conference on the Engineering in Medicine and Biology Society*, 2011, pp. 6789–6792.

[20] J. Chiang, Z. J. Wang, and M. J. McKeown, "A generalized multivariate autoregressive (GmAR)-based approach for EEG source connectivity analysis," *IEEE Transactions on Signal Processing*, vol. 60, no. 1, pp. 453–465, 2012.

[21] S. S. Rautaray and A. Agrawal, "Design of gesture recognition system for dynamic user interface," in *Proceedings of the IEEE International Conference on Technology Enhanced Education*, 2012, pp. 1–6.

[22] V. Viitaniemi, M. Karppa, and J. Laaksonen, "Experiments on recognising the handshape in blobs extracted from sign language videos," in *Proceedings of the IEEE International Conference on Pattern Recognition (ICPR'14)*, 2014, pp. 2584–2589.

[23] R. Z. Khan and N. A. Ibraheem, "Survey on gesture recognition for hand image postures," *Computer and Information Science*, vol. 5, no. 3, pp. 110, 2012.

[24] A. El-Sawah, C. Joslin, N. D. Georganas, and E. M. Petriu, "A framework for 3D hand tracking and gesture recognition using elements of genetic programming," in *Proceedings of the IEEE Canadian Conference on Computer and Robot Vision*, 2007, pp. 495–502.

[25] S. Lu, G. Tsechpenakis, D. N. Metaxas, M. L. Jensen, and J. Kruse, "Blob analysis of the head and hands: A method for deception detection," in *Proceedings of the IEEE Annual Hawaii International Conference on System Sciences*, 2005, pp. 20c–20c.

[26] A. Thangali and S. Sclaroff, "An alignment based similarity measure for hand detection in cluttered sign language video," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2009, pp. 89–96.

[27] E.-J. Ong and R. Bowden, "A boosted classifier tree for hand shape detection," in *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition*, 2004, pp. 889–894.

[28] R. Wood and J. I. Olszewska, "Lighting-variable AdaBoost based-on system for robust face detection," in *Proceedings of the International Conference on Bio-Inspired Systems and Signal Processing*, 2012, pp. 494–497.

[29] S. Berman and H. Stern, "Sensors for gesture recognition systems," *IEEE Transactions on Systems, Man, and Cybernetics - Part C: Applications and Reviews*, vol. 42, no. 3, pp. 277–290, 2012.

[30] J. Kumar, Q. Li, S. Kyal, E. A. Bernal, and R. Bala, "On-the-fly hand detection training with application in egocentric action recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW' 15)*, 2015, pp. 18–27.

[31] D. D. Nguyen, T. C. Pham, X. D. Pham, S. H. Jin, and J. W. Jeon, "Finger extraction from scene with grayscale morphology and BLOB analysis," in *Proceedings of the IEEE International Conference on Robotics and Biomimetics*, 2009, pp. 324–329.

[32] J. Suarez and R. R. Murphy, "Hand gesture recognition with depth images: A review," in *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication*, 2012, pp. 441–417.

[33] S. Rungruangbaiyok, R. Duangsoithong, and K. Chetpattananondh, "Ensemble threshold segmentation for hand detection," in *Proceedings of the IEEE International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology*, 2015, pp. 1–5.

[34] E. Marilly, A. Gonguet, O. Martinot, and F. Pain, "Gesture interactions with video: From algorithms to user evaluation," *Bell Labs Technical Journal*, vol. 17, no. 4, pp. 103–118, 2013.

[35] Y. R. Wang, J. L. Syu, H. T. Li, and L. Yang, "Fast hand detection and gesture recognition," in *Proceedings of the IEEE International Conference on Machine Learning and Cybernetics*, 2015, vol. 1, pp. 408–413.

[36] Z. Musa, K. Jumari, and N. Zainal, "A method of human skin detection base on background subtraction and color enhancement," in *Proceedings of the IEEE Symposium on Business, Engineering and Industrial Applications*, 2011, pp. 498–502.

[37] D. Xu, Y. L. Chen, X. Wu, Y. Ou, and Y. Xu, "Integrated approach of skin-color detection and depth information for hand and face localization," in *Proceedings of the IEEE International Conference on Robotics and Biomimetics*, 2011, pp. 952–956.

[38] S. Rautaray, S. Siddharth, and A. Agrawal, "Vision based hand gesture recognition for human computer interaction: A survey," *Artificial Intelligence Review*, vol. 43, no. 1, pp. 1–54, 2015.

[39] V. Vezhnevets, V. Sazonov, and A. Andreeva, "A survey on pixel-based skin color detection techniques," in *Proceedings of the IEEE Graphicon*, 2003, vol. 3, pp. 85–92.

[40] S. Bilal, R. Akmeliawati, M. J. E. Salami, A. A. Shafie, and E. M. Bouhabba, "An hybrid method using Haar-like and skin-color algorithm for hand posture detection, recognition and tracking," in *Proceedings of the IEEE International Conference on Mechatronics and Automation*, 2010, pp. 934–939.

[41] S. Xie and J. Pan, "Hand detection using robust color correction and Gaussian mixture model," in *Proceedings of the IEEE International Conference on Image and Graphics*, 2011, pp. 553–557.

[42] B. Junxia, Y. Jianqin, W. Jun, and Z. Ling, "Hand detection based on depth information and color information of the Kinect," in *Proceedings of the IEEE Chinese Control and Decision Conference*, 2015, pp. 4205–4210.

[43] J. Rajan, "Pantech Solution,," 2013, Available online at: `https://www.pantechsolutions.net/blog/matlab-code-for-background-subtraction/`.

[44] J. I. Olszewska, "Multi-camera video object recognition using active contours," in *Proceedings of the International Conference on Bio-Inspired Systems and Signal Processing*, 2015, pp. 379–384.

[45] C. R. Wren, A. Azarbayejani, T. Darrell, and A. P. Pentland, "Real-time tracking of the human body," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 780–785, 1997.

[46] J. Sonkusare, N.B. Chopade, R. Sor, and S. L. Tade, "A review on hand gesture recognition system," in *Proceedings of the IEEE International Conference on Computing, Communication, Control, and Automation*, 2015, pp. 790–794.

[47] P. Viola and M. J. Jones, "Robust real-time face detection," *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, January 2004.

[48] N. Soontranon, S. Aramvith, and T. H. Chalidabhongse, "Face and hands localization and tracking for sign language recognition," in *Proceedings of the IEEE International Symposium on Communications and Information Technology*, 2004, vol. 2, pp. 1246–1251 vol.2.

[49] S. Kang, J. Oh, and H. Hong, "Human gesture detection based on 3D blobs and skeleton model," in *Proceedings of the IEEE International Conference on Information Science and Applications*, 2013, pp. 1–4.

[50] H. S. Park and K. H. Jo, "Real-time hand gesture recognition for augmented screen using average background and Camshift," in *Proceedings of the IEEE Korea-Japan Joint Workshop on Frontiers of Computer Vision*, 2013, pp. 18–21.

[51] X. Meng, J. Lin, and Y. Ding, "An extended HOG model: SCHOG for human hand detection," in *Proceedings of the IEEE International Conference on Systems and Informatics*, 2012, pp. 2593–2596.

[52] H. Cheng, L. Yang, and Z. Liu, "A survey on 3D hand gesture recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 9, pp. 1659–1673, 2016.

[53] L. Evett, A. Burton, S. Battersby, D. Brown, N. Sherkat, G. Ford, H. Liu, and P. Standen, "Dual camera motion capture for serious games in stroke rehabilitation," in *Proceedings of the IEEE International Conference on Serious Games and Applications for Health*, 2011, pp. 1–4.

[54] W.-H. Chen, Y.-H. Lin, and S.-J. Yang, "A generic framework for the design of visual-based gesture control interface," in *Proceedings of the IEEE Conference on Industrial Electronics and Applications*, 2010, pp. 1522–1525.

[55] H. V. Verma, E. Aggarwal, and S. Chandra, "Gesture recognition using Kinect for sign language translation," in *Proceedings of the IEEE International Conference on Image Information Processing*, 2013, pp. 96–100.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 299–302
DOI: 10.15439/2016F274
ACSIS, Vol. 8. ISSN 2300-5963

Random Forest Feature Selection for Data Coming from Evaluation Sheets of Subjects with ASDs

Krzysztof Pancerz, Wiesław Paja
University of Rzeszów, Poland
Email: {kpancerz,wpaja}@ur.edu.pl

Jerzy Gomuła
Cardinal Stefan Wyszyński University
Warsaw, Poland
Email: jerzy.gomula@wp.pl

***Abstract*—We deal with the problem of initial analysis of data coming from evaluation sheets of subjects with Autism Spectrum Disorders (ASDs). In our research, we use an original evaluation sheet including questions about competencies grouped into 17 spheres. In the paper, we are focused on a feature selection problem. The main goal is to use appropriate data to build simpler and more accurate classifiers. The feature selection method based on random forest is used.**

I. Introduction

AUTISM is a brain development disorder that impairs social interaction and communication, and causes restricted and repetitive behaviors. Autism spectrum disorders can dramatically affect a child's life, as well as that of their families, schools, friends and the wider community.

The main aim of our research is to adapt computational intelligence methods for computer-aided decision support in diagnosis and therapy of persons with ASDs. In the first step of our research, we are interested in initial analysis of data coming from evaluation sheets of subjects with ASDs. The evaluation sheet, we use in the research, is an original sheet including questions (more than 300) about competencies of the subjects grouped into 17 spheres, among others, self-service, communication, cognitive, physical, as well as the sphere responsible for functioning in the social and family environment.

An initial analysis is focused on the data preprocessing step. The preprocessed data can be used to build simpler and more accurate classifiers. It is obvious, that an increasing number as well as complexity of classification rules make it difficult to be validated by domain experts. Experiments showed that in case of our evaluation sheet, over 300 features corresponding to questions (even divided into spheres) lead to less accurate classifiers with complex classification rules. Therefore, there is an important problem to select appropriate data to build (train) classifiers. In general, there is a variety of data preprocessing operations concerning both cases (instances) and features in datasets (cf. [1], [2], [3]). In [4], our consideration was focused on the case selection problem. Now, we deal with the feature selection problem.

Efficient analysis and retrieval of regularity from data is an extremely important task in the case of aggregation of vast amounts of data. Data mining processes are exposed to many aspects which cause failures. The large number of objects and variables, insignificance of some variables for the classification, interdependences between some part of variables, uneven distribution of target classes, and other difficulties are the reason to develop methods for effective selection of significant feature subsets.

There are three major categories of feature selection methods: filter, wrapper and embedded methods. The first one scores variables individually using different measures and eliminates some of them before a model is constructed [5]. In turn, wrapper methods investigate the prediction accuracy of a model directly measuring the value of a feature set. Although effective, the exponential number of possible subsets places computational limits for the wide data sets that are the focus of this work. The last type, embedded methods firstly develop a learning model and then analyze the model to estimate the relevance of a feature. Effects are dependent on methods used for model generation. During our experiments the Boruta algorithm [6] for feature selection was used.

Experiments showed that selected datasets enabled us to build simpler and more accurate classifiers, both decision tree based and rule based ones.

II. Input Data

Experiments which test the relative effectiveness of our approach have been performed on data describing over 70 cases (subjects) classified into three categories: high-functioning ($HIGH$), medium-functioning ($MEDIUM$), or low-functioning (LOW) autism. Each subject has been evaluated using an original sheet including questions about competencies grouped into 17 spheres marked with Roman numerals (only spheres used in our experiments are listed):

- VI. Support for active communication.
- VII. Active communication concerning objects, people, parts of the body.
- VIII. Imitation, the length and complexity of the utterance.
- IX. Needs, emotions, moods.
- X. Object communication (the level of specific symbols).
- XI. Symbolic communication.
- XII. Requests.
- XIII. Choices.
- XIV. Communication in a pair (with a contemporary, with an adult).
- XV. Social communication competences.
- XVI. Communication in a group and in social situations (in a team, at school, in the closest social environment).

978-83-60810-90-3/$25.00

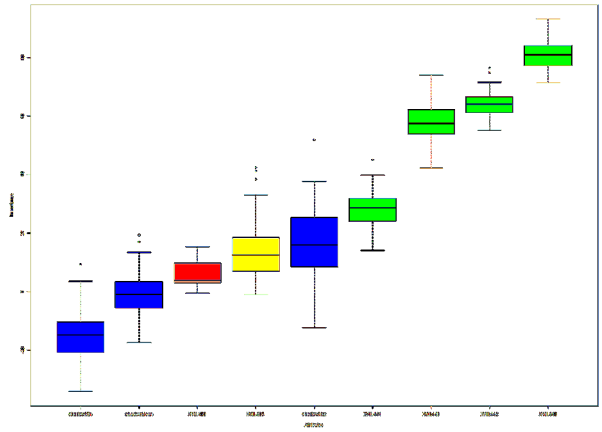


Fig. 2. Results of the feature selection process for sphere XVIII

TABLE I
A NUMBER OF FEATURES IN DATASETS

Dataset	#All features	#Confirmed features	#Tentative features
VI	18	5	4
VII	14	11	1
VIII	87	29	10
IX	51	21	6
X	3	1	0
XI	12	8	1
XII	9	1	2
XIII	14	11	3
XIV	13	11	1
XV	34	11	7
XVI	25	10	9
XVIII	6	4	1
XIX	7	6	1
XX	9	6	2
XXI	8	3	3
XXII	13	10	1
XXIII	13	8	2

- XVIII. Vocabulary.
- XIX. The degree of effectiveness of information.
- XX. The degree of motivation to communicate.
- XXI. The degree and type of hint in communication.
- XXII. Building the utterance - the degree of its complexity and functionality.
- XXIII. Dialogues.

Each case is described by over 300 features. Four values of features are possible, namely 0, 25, 50, and 100. They have the following meaning:

- 0 - not performed,
- 25 - performed after physical help,
- 50 - performed after verbal help/demonstration,
- 100 - performed unaided.

III. TOOLS

To solve a feature selection problem, we have used the Boruta algorithm. This algorithm applies random forest to determine all-relevant feature subset from datasets. It was designed as a wrapper method. Trees are independently developed on different bagging samples of the training set. The importance estimation of an attribute is gathered as the loss of accuracy of classification caused by a random permutation of attribute values between objects. It is computed separately for all trees in the forest which use a given attribute for classification. After that, the average and standard deviation of the accuracy loss are computed. Thus, the Z score computed by dividing the average loss by its standard deviation can be used as an importance measure [6], [7]. Boruta separates attributes into three categories:

- confirmed,
- tentative,
- rejected.

Figures 1 and 2 show some examples of results of feature selection processes. The confirmed attributes are marked with green, tentative - with yellow, and rejected with red.

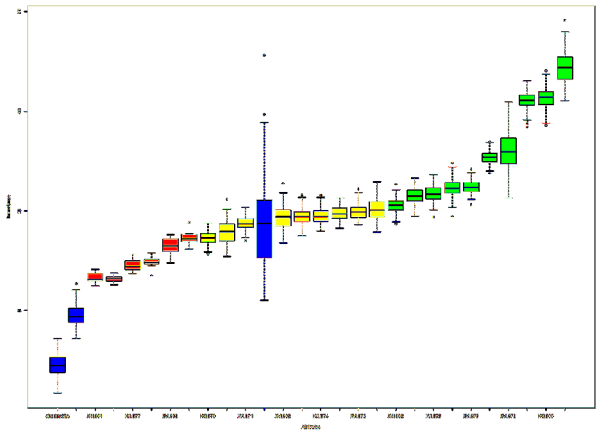


Fig. 1. Results of the feature selection process for sphere XVI

The datasets, after the feature selection processes, have been used to build decision tree and rule based classifiers.

For building classifiers, we have used two machine learning computer tools:

- RSES - a toolset for analyzing data with the use of methods coming from rough set theory [8].
- Orange - a comprehensive, component-based software suite for machine learning and data mining [9].

In RSES, we have used the LEM2 algorithm [10] for rule generation. LEM2 is most frequently used for rule induction. LEM2 explores the search space of feature-values pairs. It is based on lower and upper approximations of decision classes defined in rough set theory [11]. The expected degree of coverage of the training set by derived rules was set to 0.9. In a classification process, conflicts were resolved by standard voting (each rule has as many votes as supporting cases).

In Orange, we have used an algorithm for generation of decision trees based on the Gini criterion [1]. The following values of pruning parameters were set:

- minimum instances in leaves: 2,
- limit of the depth: 100.

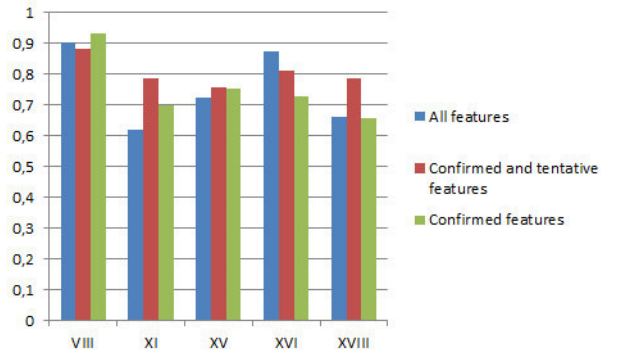


Fig. 5. Selected results of experiments with LEM2: classification accuracy

IV. Results

In this section, we give selected results of experiments with the Boruta feature selection algorithm and classification algorithms (the algorithm of decision tree generation implemented in Orange and the LEM2 algorithm for rule generation implemented in RSES).

In our experiments, each data set has been treated separately. It enabled us to assess the evaluation sheet with respect to individual spheres. The results can be used in further development of the sheet. In the future, any adding, removing, and modifying of questions are allowed. Especially, the questions corresponding to rejected features should be checked.

Table I shows the effects of applying a feature selection procedure in terms of a number of features. Next, we present the results of assessment of classifiers for selected datasets (spheres), see Figures from 3 to 8. To estimate the accuracy of classifiers, ten-fold cross-validation method was used.

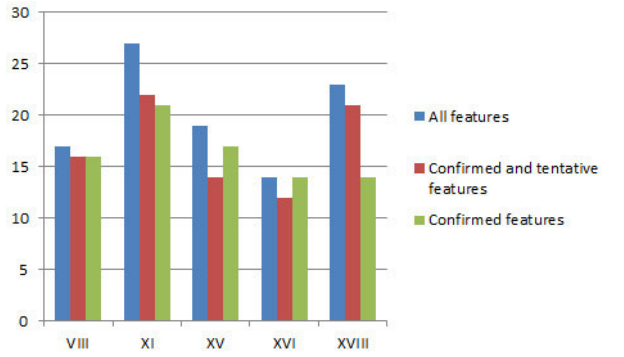


Fig. 3. Selected results of experiments with LEM2: a number of rules

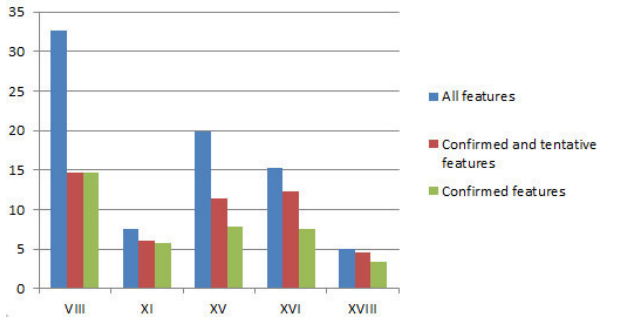


Fig. 4. Selected results of experiments with LEM2: mean of rule premise length

In case of complexity of classifiers, we have taken into consideration:

- a number of rules and mean of rule premise length (for a rule based classifier),
- a number of nodes and a number of leaves (for a decision tree based classifier).

In general, a feature selection procedure in the preprocessing step causes the decrease in the complexity of classifiers. In case of decision trees, a feature selection procedure positively influences the classification accuracy. In the case of rules generated by LEM2, taking into consideration the confirmed and tentative features seems to be more appropriate.

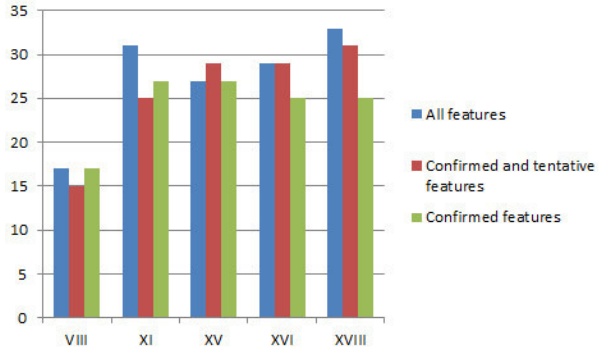


Fig. 6. Selected results of experiments with a decision tree: a number of nodes

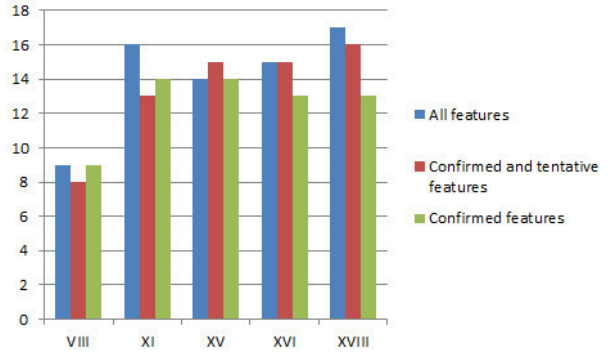


Fig. 7. Selected results of experiments with a decision tree: a number of leaves

V. Conclusions and Further Work

In the paper, we have examined the Boruta algorithm to solve the feature selection problem for data coming from evaluation sheets of subjects with Autism Spectrum Disorders (ASDs). Simultaneously our research is also focused on the case selection problem [4]. Our main goal is to create hybrid

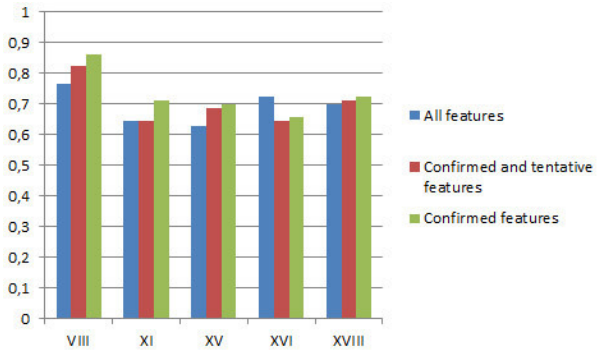


Fig. 8. Selected results of experiments with a decision tree: classification accuracy

classifiers combining a wide range of approaches that will be implemented in a dedicated computer tool supporting diagnosis and therapy of persons with ASDs.

REFERENCES

[1] K. Cios, W. Pedrycz, R. Swiniarski, and L. Kurgan, *Data mining. A knowledge discovery approach*. New York: Springer, 2007.

[2] S. García, J. Luengo, and F. Herrera, *Data Preprocessing in Data Mining*, ser. Intelligent Systems Reference Library. Switzerland: Springer International Publishing, 2015, vol. 72.

[3] N. Jankowski and M. Grochowski, "Comparison of instances selection algorithms I. Algorithms survey," in *Artificial Intelligence and Soft Computing - ICAISC 2004*, ser. Lecture Notes in Computer Science, L. Rutkowski, J. H. Siekmann, R. Tadeusiewicz, and L. A. Zadeh, Eds. Berlin, Heidelberg: Springer-Verlag, 2004, vol. 3070, pp. 598–603.

[4] K. Pancerz, A. Derkacz, and J. Gomuła, "Consistency-based preprocessing for classification of data coming from evaluation sheets of subjects with ASDs," in *Position Papers of the 2015 Federated Conference on Computer Science and Information Systems (FedCSIS'2015)*, ser. Annals of Computer Science and Information Systems, M. Ganzha, L. Maciaszek, and M. Paprzycki, Eds., vol. 6, Lodz, Poland, 2015. doi: 10.15439/2015F393 pp. 63–67.

[5] E. Tuv, A. Borisov, G. Runger, and K. Torkkola, "Feature selection with ensembles, artificial variables, and redundancy elimination," *Journal of Machine Learning Research*, vol. 10, pp. 1341–1366, 2009.

[6] W. R. Rudnicki, M. Wrzesień, and W. Paja, "All relevant feature selection methods and applications," in *Feature Selection for Data and Pattern Recognition*, ser. Studies in Computational Intelligence, U. Stańczyk and C. L. Jain, Eds. Berlin, Heidelberg: Springer-Verlag, 2015, vol. 584, pp. 11–28.

[7] M. Kursa and W. Rudnicki, "Feature selection with the Boruta package," *Journal of Statistical Software*, vol. 36, no. 1, 2010. doi: 10.18637/jss.v036.i11

[8] J. G. Bazan and M. S. Szczuka, "The Rough Set Exploration System," in *Transactions on Rough Sets III*, ser. Lecture Notes in Artificial Intelligence, J. Peters and A. Skowron, Eds. Berlin Heidelberg: Springer-Verlag, 2005, vol. 3400, pp. 37–56.

[9] J. Demšar, T. Curk, A. Erjavec, Črt Gorup, T. Hočevar, M. Milutinovič, M. Možina, M. Polajnar, M. Toplak, A. Starič, M. Štajdohar, L. Umek, L. Žagar, J. Žbontar, M. Žitnik, and B. Zupan, "Orange: Data mining toolbox in Python," *Journal of Machine Learning Research*, vol. 14, pp. 2349–2353, 2013.

[10] J. Grzymala-Busse, "A new version of the rule induction system LERS," *Fundamenta Informaticae*, vol. 31, pp. 27–39, 1997.

[11] Z. Pawlak and A. Skowron, "Rudiments of rough sets," *Information Sciences*, vol. 177, pp. 3–27, 2007. doi: 10.1016/j.ins.2006.06.003

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 303–308
DOI: 10.15439/2016F146
ACSIS, Vol. 8. ISSN 2300-5963

A Conception of Pairwise Comparisons Model for Selection of Appropriate Body Surface Area Calculation Formula

Grzegorz Redlarski*†, Waldemar W. Koczkodaj‡, Marek Krawczuk*, Janusz Siebert§¶,
Katarzyna E. Mrozik*, Aleksander Palkowski*, and Piotr M. Tojza*
*Department of Mechatronics and High Voltage Engineering
Gdansk University of Technology, Gdansk, 80–233, Poland
Email: {grzegorz.redlarski, marek.krawczuk}@pg.gda.pl,
katmrozi@student.pg.gda.pl, {aleksander.palkowski, piotr.tojza}@pg.gda.pl
†Faculty of Technical Sciences
University of Warmia and Mazury, Olsztyn, 10–719, Poland
‡Laurentian University, 935 Ramsey Lake Rd ON P3E 2C6 Sudbury, Canada
Email: wkoczkodaj@cs.laurentian.ca
§Department of Family Medicine
Medical University of Gdansk, Gdansk, 80–211, Poland
Email: kmr@gumed.edu.pl
¶University Center for Cardiology
Gdansk, Poland

***Abstract*—Body surface area (BSA) may be computed using a variety of formulas, but the computed BSA differs from real BSA values for particular subjects. This is presented in the paper by computing BSA values for selected subject and comparing them to the real BSA value obtained with the use of a 3D body scanner. The results show inequalities in the relevant BSA computing formulas. Hence, there is a need to determine a method that will allow to select the best formula for calculating BSA in a particular case. For this purpose, the pairwise comparisons (PC) method is suggested. This article presents a proposition of using consistency-driven PC, as well as the basic and most important aspects of using PC to determine the appropriate BSA calculation formula.**

I. Introduction

During the last century, many different body surface area (BSA) formulas have been developed for use as the indicator of patient-focused health outcomes [1]. Commonly used BSA formulas had been reported as having different properties and accuracy in determination of real BSA values.

BSA is a parameter commonly used in medicine, mainly in oncology and burns treatment [1], [2]. It is crucial to determine the exact BSA value of the patient with minimal error using only the knowledge of patient's height and weight. Height and weight can be relatively easy to obtain especially when working under time pressure, which makes those factors suitable for medical examination. The existing methods used to compute BSA prove to be inaccurate and may cause ineffective chemotherapy or burns treatment. This article presents a proposition for evaluation and selection of appropriate BSA formula. The consistency-driven pairwise comparisons (PC) method is proposed to be used as an examination method for choosing a BSA formula for individuals based on a series of weighted parameters. Pairwise comparisons were also used in [3]–[5].

II. BSA calculation methods

Since 1879 many BSA calculation formulas were developed, 25 of which are most common. To determine the BSA value these formulas use the patient's weight W (in kilograms) and in most cases also patient's height H (in centimetres). The formulas under question are listed in Table I.

III. BSA calculation errors

The early BSA calculation formulas were developed using different coating methods in order to obtain the patient's real BSA value. The accuracy of the methods depended not only on the type of mathematical approximation but also on the quality of the measuring methods. In the process of obtaining real BSA values the authors used subjects from a limited range of race, sex, and age, often generalizing the outcome BSA formula for an entire population.

In modern times, different methods of obtaining real BSA values were used. Still, when using the BSA calculation formulas the computed BSA values for the same patient vary in an extensive way. The results are not equal and differ one from another. This should not surprise when using old calculation formulas, but this regularity is true even for modern formulas.

To elucidate this phenomena a real BSA value was obtained from a female patient with the use of a body scanning device, specifically build for the purpose of scanning human bodies and obtaining high precision BSA. The scan was performed

978-83-60810-90-3/$25.00

TABLE I
BODY SURFACE AREA FORMULAS CONSIDERED

Authors	Formula	Reference
Meeh (1879)	$0.1053 \cdot W^{2/3}$	6
DuBois & DuBois (1916)	$0.007184 \cdot W^{0.425} \cdot H^{0.725}$	7
Faber & Melcher (1921)	$0.00785 \cdot W^{0.425} \cdot H^{0.725}$	8
Takahira (1925)	$0.007246 \cdot W^{0.425} \cdot H^{0.725}$	9
Breitmann (1932)	$0.0087 \cdot (W + H) - 0.26$	10
Boyd (1935)	$0.0003207 \cdot (W \cdot 1000)^{0.7285 - 0.0188 \cdot \log_{10}(W \cdot 1000)} \cdot H^{0.3}$	11
Stevenson (1937)	$0.0128 \cdot W + 0.0061 \cdot H - 0.1529$	12
Sendroy & Cecchini (1954)	$0.0097 \cdot (W + H) - 0.545$	13
Banerjee & Sen (1955)	$0.007466 \cdot W^{0.425} \cdot H^{0.725}$	14
Choi (1956)	men: $0.005902 \cdot W^{0.407} \cdot H^{0.776}$ women: $0.008692 \cdot W^{0.442} \cdot H^{0.678}$	15
Mehra (1958)	$0.01131 \cdot W^{0.4092} \cdot H^{0.6468}$	16
Banerjee & Bhattacharya (1961)	$0.007 \cdot W^{0.425} \cdot H^{0.725}$	17
Fujimoto et al. (1968)	$0.008883 \cdot W^{0.444} \cdot H^{0.663}$	18
Gehan & George (1970)	$0.0235 \cdot W^{0.51456} \cdot H^{0.42246}$	19
Haycock et al. (1978)	$0.024265 \cdot W^{0.5378} \cdot H^{0.3964}$	20
Mosteller (1987)	$\sqrt{W \cdot H \,/\, 3600}$	21
Mattar (1989)	$(W + H - 60)/100$	22
Nwoye (1989)	$0.001315 \cdot W^{0.262} \cdot H^{1.2139}$	23
Shuter & Aslani (2000)	$0.00949 \cdot W^{0.441} \cdot H^{0.655}$	24
Livingston & Lee (2001)	$0.1173 \cdot W^{0.6466}$	25
Tikuisis (2001)	men: $0.01281 \cdot W^{0.44} \cdot H^{0.6}$ women: $0.01474 \cdot W^{0.47} \cdot H^{0.55}$	26
Nwoye & Al-Sheri (2003)	$0.02036 \cdot W^{0.427} \cdot H^{0.516}$	27
Yu, Lo, Chiou (2003)	$0.015925 \cdot (W \cdot H)^{0.5}$	28
Schlich (2010)	men: $0.000579479 \cdot W^{0.38} \cdot H^{1.24}$ women: $0.000975482 \cdot W^{0.46} \cdot H^{1.08}$	29
Yu, Lin, Yang (2010)	$0.00713989 \cdot W^{0.404} \cdot H^{0.7437}$	30

with the use of Artec Eva 3D Scanner. The testes subject was a young 22-year-old female, of a body height of 171 cm and weight of 55.8 kg (Fig. 1a). After obtaining her real BSA value (1.633 m^2), formulas presented in Table I were used to calculate individual BSA values. The results are shown in Fig. 1b. For most cases the results are greatly inconsistent with the real BSA value. The calculated BSA values span from 1.538 m^2 (for the Meeh method) to 1.937 m^2 (for the Nwoye method). Fig. 2 presents percentage values of errors between the real BSA value obtained through scanning and the formulas shown in Table I. The errors show that in most cases the formulas indicate BSA values lower than the real one. When taking into consideration the Meeh and Nwoye formulas, the maximum error that can be made in calculating BSA is 24.46%. Therefore, it is important to develop the best method to select the right BSA calculation formula. For this purpose the PC model was selected.

To evaluate the above mentioned observation, a similar procedure concerning BSA calculation was performed in the case of 42 patients. The results are shown in Fig. 3. The patients used for this study were 20 to 28 year old, healthy Caucasians, both males and females. As it can be seen, the highest error values were obtained for the Nwoye method (12-15 percent in general) whereas the lowest error values characterizes the Yu, Lo and Chiou BSA calculation method (about 5 percent).

IV. DEFINITION OF WEIGHTS USED FOR BSA FORMULA SELECTION

The PC method requires describing weights that are used in the selection process. They should be distinctive to the analysed process and represent the subjective assessment performed by a specialist (e.g., a physician) on the scale from 0 to 5.

In the research process all scanned subjects can be divided to five groups based on their physique, age or medical history: normal, obese, after/during chemotherapy, elderly, and children. To each of these groups five factors are used in order to describe individual patients: degree of obesity, height to volume ratio, anthropological ancestry and race, type of body physique (athletic, deformed or similar), and degree of skin corrugation. The above mentioned classification and rating process is shown in Fig. 4.

V. THE PC METHOD PRELIMINARIES

The pioneer of PC is Condorcet [34]. He used PC in 1785 in the context of counting political ballots. In 1860, however, Fechner provided further, yet limited, psychometric information about this method. By way of refining the method, Thurstone [36] described the PC method as a statistical analysis and proposed a solution. In 1977 Saaty [37] introduced a hierarchy instrument for practical applications.

The PC method is outlined in Appendix A [31]–[33], [39]. It creates a matrix A of values a_{ij} of the i-th candidate (or alternative) compared with the j-th candidate. A scale $[1/c, c]$

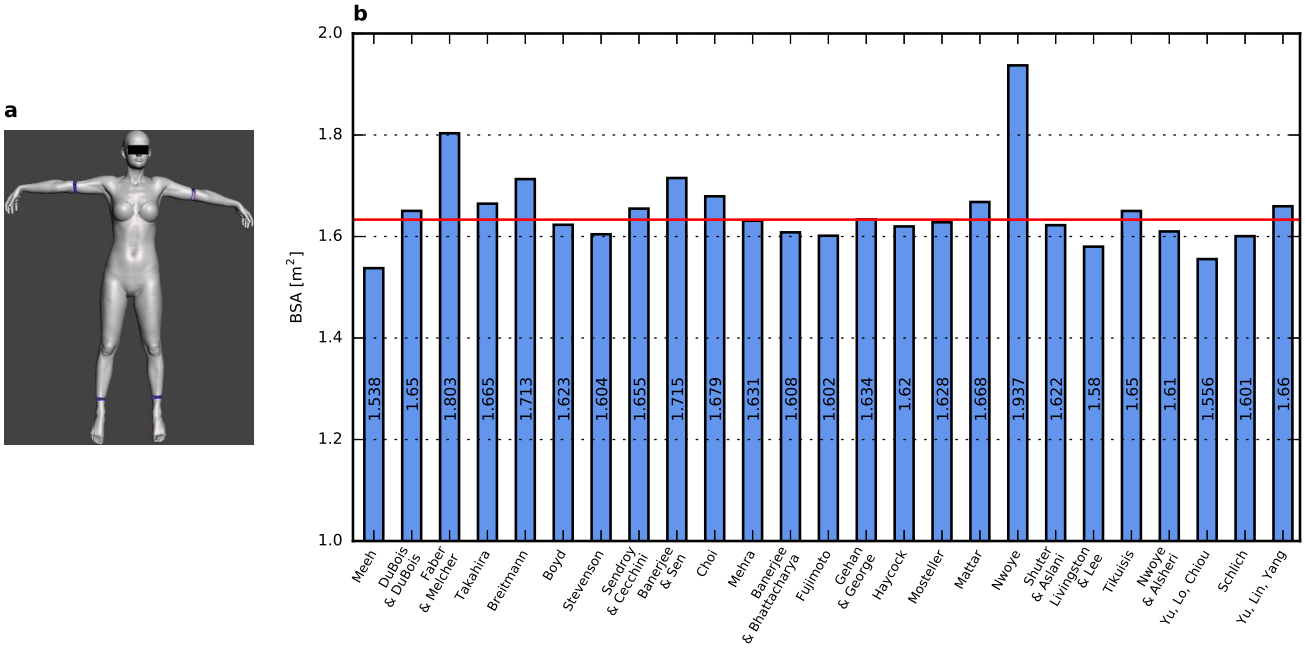


Fig. 1. BSA of a selected patient. **a**: 3D model of the patient. **b**: Calculated BSA values. The red line indicates the real BSA.

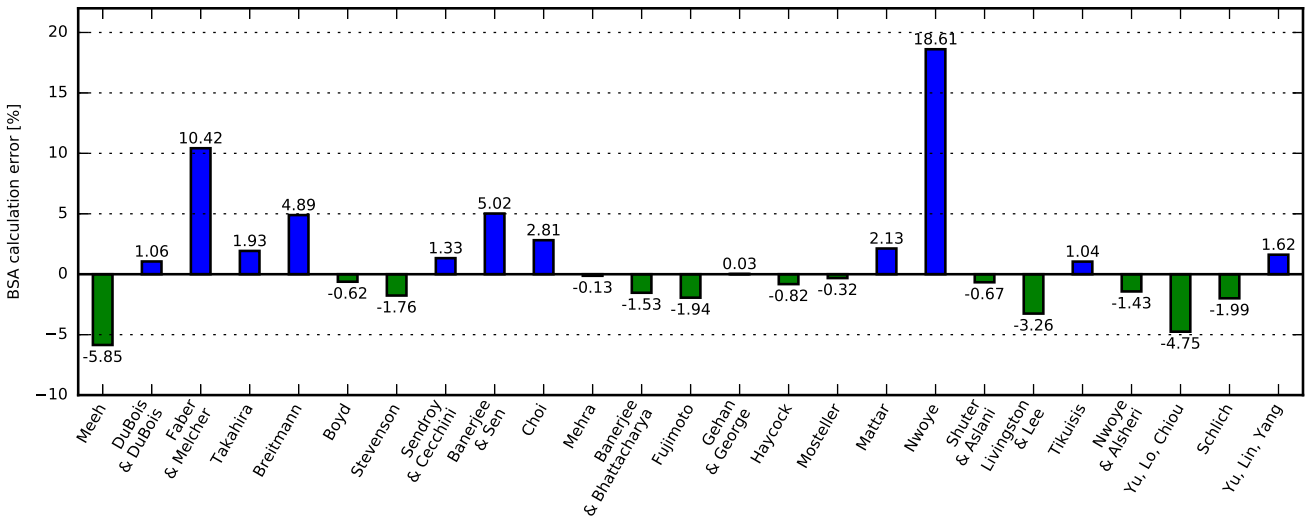


Fig. 2. Percentage error between the real BSA for a selected patient and BSA values calculated using the formulas shown in Table I.

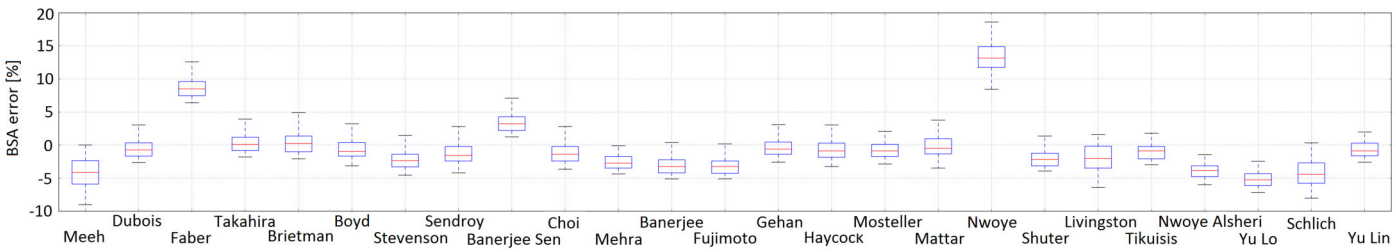


Fig. 3. Percentage error between the real BSA for a selected patient and BSA values calculated using the formulas shown in Table I for 42 patients.

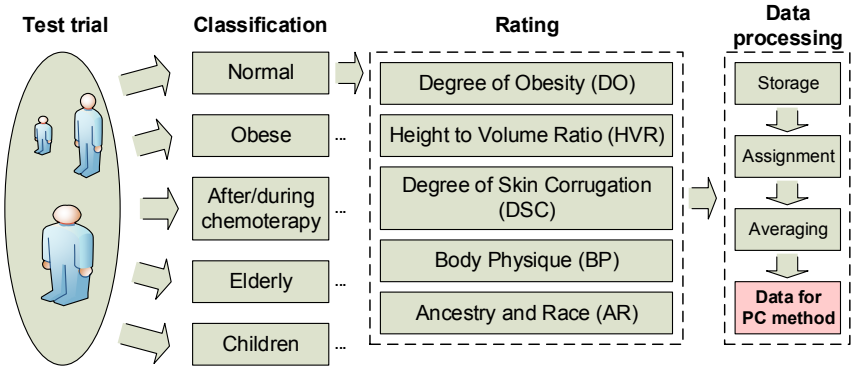


Fig. 4. Process of obtaining weights for factors used in the PC method.

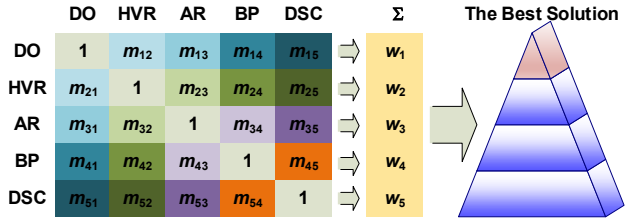


Fig. 5. The PC table of criteria and selection process.

is used for i to j comparisons where $c > 1$ is a small real number (5 to 9 in most practical applications). It is usually assumed that all the values a_{ij} on the main diagonal are equal to 1.

Using a scale of 1 to 5, the relative importance of each of the five groups are entered and objects are compared in the smallest subgroup. For example, degree of obesity and height to volume ratio are compared to each other in the subgroup and given 4 out of 5 (which can be changed for every clinical case to which this instrument is applied). In the case of inconsistency of the resulting matrix, the following formula [32] can be applied:

$$ii = \min(|1 - a_{ij}/(a_{ik} \cdot a_{kj})|,\ |1 - a_{ik} \cdot a_{kj}/a_{ij}|) \quad \text{for } i = 1,\ j = 2, \text{ and } k = 3 \qquad (1)$$

An example of a PC table of criteria and the selection process is presented in Fig. 5. The above presented factors are compared and relevant comparison values are assigned. Then the summed up weights of particular rows are compared with each other, thus providing an indicator for which of the tested solution is the best in the case.

As explained by Koczkodaj [32], [39], the weights w_1 to w_5 are computed as normalized geometric means of the matrix rows. The example is presented to illustrate the method, not the real instrument.

VI. Conclusion

BSA is often a major factor in determining of the course of treatment. A series of formulas to simplify the process have been developed throughout the years. However, the choice of a particular formula is a difficult task. Therefore, there is a need to develop additional methods to help in the selection of an appropriate formula for individuals.

Although the PC method was originally used over 200 years ago, it has not been utilised to refine the properties of quality of life instruments. The method can strengthen the BSA calculation instrument by providing an additional layer of selection.

Evidently, not all objects on the BSA instrument are of equal importance. Appreciation of their relative differences adds to the measure's precision. The inconsistency analysis further strengthens the measure by bringing the most problematic but often crucial comparisons of the instrument items. A challenge to the multiple experts in this tool's development can be "averaging" their individual assessments in the assumed model. Clinical trials and statistical analysis need to follow the model enhancement.

The proposition presented in this paper show that it is possible to use the pairwise comparison in order to select a BSA calculation formula conformed to a specific situation. The enhancement may be a challenging undertaking for years to come. Refinement of the BSA may improve understanding of physiology as well as improve health care professional practices in their efforts to assess quality of life.

Acknowledgment

The research was supported by National Science Centre grant 2014/15/B/NZ7/01018.

References

[1] Redlarski, G., Tojza, P.M.,Computer Supported Analysis of the Human Body Surface Area, International Journal of Innovative Computing, Information and Control. Vol 9, Issue 5, pp 1801-1818
[2] G. Redlarski, A. Palkowski, and M. Krawczuk, Body surface area formulae: an alarming ambiguity, Sci. Rep., vol. 6, p. 27966, Jun. 2016
[3] Mohammed Alqarni, Yassen Arabi, Tamar Kakiashvili, Mohammed Khedr, Waldemar W. Koczkodaj, J. Leszek, Artur Przelaskowski, K. Rutkowski: Improving the predictiveness of ICU medical scales by the method of pairwise comparisons. FedCSIS 2011: 11-17

[4] Tamar Kakiashvili, Waldemar W. Koczkodaj, Phyllis Montgomery, Kalpdrum Passi, Ryszard Tadeusiewicz: Assessing the properties of the World Health Organization's Quality of Life Index. IMCSIT 2008: 151-154
[5] Waldemar W. Koczkodaj, Vova Babiy, Agnieszka D. Bogobowicz, Ryszard Janicki, Alan Wassyng: Selecting the best strategy in a software certification process. IMCSIT 2010: 53-58
[6] Meeh, K. Oberachenmessungen des menschlichen korpers. Zeitschrift fur Biologie 15, 425485 (1879).
[7] Du Bois, D. and Du Bois, E. F. A formula to estimate the approximate surface area if height and weight be known. Arch 117 Intern Med 17, 863871 (1916).
[8] Faber, H. K. and Melcher, M. S. A modication of the Du Bois height-weight formula for surface areas of newborn infants. 172 Experimental Biology and Medicine 19, 5354 (1921).
[9] Takahira, H. Metabolism of the Japanese. Imperial Government Institute of Nutrition. Report (The Institute, 1925).
[10] Breitmann, M. Eine vereinfachte Methodic der Korperoberachbestimmung. Zeitschrift fur Konstitutionslehre 17, 211 175 214 (1932).
[11] Boyd, E. Growth of Surface Area in Human Body. In Institute of Child Welfare Monograph Series, vol. 10, 5360 177 (University of Minnesota Press, Minneapolis, 1935), 3rd edn.
[12] Stevenson, P. H. Height-weight-surface formula for the estimation of body surface area in Chinese subjects. The Chinese 179 Journal of Physiology 12, 327334 (1937).
[13] Sendroy, J. and Cecchini, L. P. Determination of Human Body Surface Area From Height and Weight. Journal of Applied 181 Physiology 7, 112 (1954).
[14] Banerjee, S. and Sen, R. Determination of the Surface Area of the Body of Indians. Journal of Applied Physiology 7, 183 585588 (1955)
[15] Choi, W. R. The body surface area of Koreans. Ph.d. dissertation, Seoul National University (1956).
[16] Mehra, N. C. Body Surface Area of Indians. Journal of Applied Physiology 12, 3436 (1958).
[17] Banerjee, S. and Bhattacharya, A. K. Determination of body surface area in Indian Hindu children. Journal of Applied 187 Physiology 16, 969970 (1961).
[18] Fujimoto, S., Watanabe, T., Sakamoto, A., Yukawa, K. and Morimoto, K. Studies on the physical surface area of Japanese. 189 18. Calculation formulas in three stages over all ages. Japanese Journal of Hygiene 23, 443450 (1968).
[19] Gehan, E. A. and George, S. L. Estimation of human body surface area from height and weight. Cancer chemotherapy 191 reports. Part 1 54, 225235 (1970).
[20] Haycock, G. B., Schwartz, G. J. and Wisotsky, D. H. Geometric method for measuring body surface area: A height-weight 193 formula validated in infants, children, and adults. The Journal of Pediatrics 93, 6266 (1978).
[21] Mosteller, R. D. Simplied calculation of body-surface area. The New England journal of medicine 317, 1098 (1987).
[22] Mattar, J. A. A Simple Calculation to Estimate Body Surface Area in Adults and Its Correlation with the Du Bois Formula. 196 Critical Care Medicine 17, 846853 (1989).
[23] Nwoye, L. O. Body Surface Area of Africans: A Study Based on Direct Measurements of Nigerian Males. Human 198 Biology 61, 439457 (1989).
[24] Shuter, B. and Aslani, A. Body surface area: Du Bois and Du Bois revisited. European Journal of Applied Physiology 82, 200 250254 (2000).
[25] Livingston, E. H. and Lee, S. Body surface area prediction in normal-weight and obese patients. American Journal of 202 Physiology - Endocrinology and Metabolism 281, E586E591 (2001).
[26] Tikuisis, P., Meunier, P. and Jubenville, C. Human body surface area: measurement and prediction using three dimensional 204 body scans. European Journal of Applied Physiology 85, 264271 (2001).
[27] Nwoye, L. O. and Al-Shehri, M. A. A formula for the estimation of the body surface area of Saudi male adults. Saudi 206 Medical Journal 24, 13411346 (2003).
[28] Yu, C.-Y., Lo, Y.-H. and Chiou, W.-K. The 3D scanner for measuring body surface area: a simplied calculation in the 119 Chinese adult. Applied Ergonomics 34, 273278 (2003).
[29] Schlich, E., Schumm, M. and Schlich, M. 3D-Body-Scan als anthropometrisches Verfahren zur Bestimmung der spezis- 121 chen Korperoberache. Ernahrungs Umschau 4, 178183 (2010).
[30] Yu, C.-Y., Lin, C.-H. and Yang, Y.-H. Human body surface area database and estimation formula. Burns 36, 616629 123 (2010).
[31] Kakiashvili, T., Kielan, K., Koczkodaj, W.W., Passi, K., Tadeusiewicz, R., Supporting the Asperger Syndrome Diagnostic Process by Selected AI Methods, proceedings of artificial intelligence studies. Vol. 4, pp21-27, 2007.
[32] Koczkodaj, W.W., A new definition of consistency of pairwise comparisons, Mathematical and computer modelling, (18)7, pp, 79-84, 1993.
[33] Koczkodaj, W.W., Herman, M.W., Orlowski, M. Using Consistency-driven Pairwise Comparisons in Knowledge-based Systems, International Conference on Information and Knowledge Management, 1997.
[34] Condorcet, M., The Essay on the Application of Analysis to the Probability of Majority Decisions, Paris: Imprimerie Royale, 1785.
[35] Fechner, G., Elemente der Psychophysik (1860, 2nd ed., 1889)
[36] Thurstone, L.L. (1927). A law of comparative judgement. Psychological Review, 34, 278-286.
[37] Saaty, T.L. (1977). A scaling method for priorities in hierarchical structures. Journal of Mathematical Psychology, 15, 234-281.
[38] McDowell. I., Measuring health : a guide to rating scales and questionnaires, 3rd ed., New York : Oxford University Press, 2006, 748 p.
[39] Koczkodaj, W.W., Redlarski, G., Szybowski, J., Wajch, E., Mikhailov, L., Soltys, M., Tamazian, G., Yuen, K.K.F., Important Facts and Observations about Pairwise Comparisons (the special issue edition), Fundamenta Informaticae 144, 1–17, 2016.

APPENDIX A
BASIC CONCEPTS OF PAIRWISE COMPARISONS

An n by n pairwise comparisons matrix is defined as a square matrix $A = [a_{ij}]$ such that $a_{ij} > 0$ for every $i, j = 1, ..., n$. Each a_{ij} expresses a relative preference of criterion (or stimulus)s_i over criterion s_j for $i, j = 1, ..., n$ represented by numerical weights (positive real numbers) and w_i and w_j respectively. The quotients $a_{ij} = w_i/w_j$ form a pairwise comparisons matrix:

$$A = \begin{bmatrix} 1 & a_{12} & \cdots & a_{1n} \\ \frac{1}{a_{12}} & 1 & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{a_{1n}} & \frac{1}{a_{2n}} & \cdots & 1 \end{bmatrix}$$

A pairwise comparisons matrix A is called *reciprocal* if $a_{ij} = 1/a_{ji}$ for every $i, j = 1, ..., n$ (then automatically $a_{ii} = 1$ for every $i = 1, ..., n$ because they represent the relative ratio of a criterion against itself). A pairwise comparisons matrix A is called *consistent* if $a_{ij} \cdot a_{jk} = a_{ik}$ holds for every $i, j, k = 1, ..., n$ since $w_i/w_j \cdot w_j/w_k$ is expected to be equal to w_i/w_k. Although every consistent matrix is reciprocal, the converse is not generally true. In practice, comparing of s_i to s_j, s_j to s_k, and s_i to s_k often results in inconsistency amongst the assessments in addition to their inaccuracy; however, the inconsistency may be computed and used to improve the accuracy.

The first step in pairwise comparisons is to establish the relative preference of each combination of two criteria. A scale from 1 to 5 can be used to compare all criteria in pairs. Values from the interval $[1/5, 1]$ reflect inverse relationships between criteria since $s_i/s_j = 1/(s_j/s_i)$. The consistency driven approach is based on the reasonable assumption that by finding the most inconsistent judgments, one can then reconsider one's own assessments. This in turn contributes to the improvement of judgmental accuracy. Consistency analysis is a dynamic process which is assisted by the software.

The central point of the inference theory of the pairwise comparisons is Saaty's Theorem [37], which states that for every n by n consistent matrix $A = [a_{ij}]$ there exist positive real numbers $w_1, ..., w_n$ (weights corresponding to criteria $s_1, .n.., s$) such that $a_{ij} = w_i/w_j$ for every $i, j = 1, ..., n$. The weights w_i are unique up to a multiplicative constant. Saaty (1977) also discovered that the eigenvector corresponding to the largest eigenvalue of A provides weights w_i which we wish to obtain from the set of preferences a_{ij}. This is not the only possible solution to the weight problem. In the past, a least squares solution was known, but it was far more computationally demanding than finding an eigenvector of a matrix with positive elements. Later, a method of row geometric means was proposed (Jensen, 1984), which is the simplest and most effective method of finding weights. A statistical experiment demonstrated that the accuracy, that is, the distance from the original matrix A and the matrix AN reconstructed from weights with elements $[a_{ij}] = [w_i/w_j]$, does not strongly depend on the method. There is, however, a strong relationship between the accuracy and consistency. Consistency analysis is the main focus of the consistency driven approach.

An important problem is how to begin the analysis. Assigning weights to all criteria (e.g., $A = 18, B = 27, C = 20, D = 35$) seems more natural than the above process. In fact it is a recommended practice to start with some initial values. The above values yield the ratios: $A/B = 0.67$, $A/C = 0.9$, $A/D = 0.51$, $B/C = 1.35$, $B/D = 0.77$, $C/D = 0.57$. Upon analysis, these may look somewhat suspicious because all of them round to 1, which is of equal or unknown importance. This effect frequently arises in practice, and experts are tempted to change the ratios by increasing some of them and decreasing others (depending on knowledge of the case). The changes usually cause an increase of inconsistency which, in turn, can be handled by the analysis because it contributes to establishing more accurate and realistic weights. The pairwise comparisons method requires evaluation of all combinations of pairs of criteria, and can be more time consuming because the number of comparisons depends on n^2 (the square of the number of criteria). The complexity problem has been addressed and partly solved by the introduction of hierarchical structures [37]. Dividing criteria into smaller groups is a practical solution in cases in which the number of criteria is large.

APPENDIX B
CONSISTENCY ANALYSIS

Consistency analysis is critical to the approach presented here because the solution accuracy of *not-so-inconsistent* matrices strongly depends on the inconsistency. The consistency driven approach is, in brief, the next step in the development of pairwise comparisons.

The challenge to the pairwise comparisons method comes from a lack of consistency in the pairwise comparisons matrices which arises in practice. Given an n by n matrix A that is not consistent, the theory attempts to provide a consistent n by n matrix AN that differs from matrix A "as little as possible". In particular, the geometric means method produces results similar to the eigenvector method (to high accuracy) for the ten million cases tested. There is, however, a strong relationship between accuracy and consistency.

Unlike the old eigenvalue based inconsistency, introduced in [37], the triad based inconsistency locates the most inconsistent triads [32]. This allows the user to reconsider the assessments included in the most inconsistent triad.

Readers might be curious, if not suspicious, about how one could arrive at values such as 1.30 or 1.50 as relative ratio judgments. In fact the values were initially different, but have been refined and the final weights have been computed by the consistency analysis. It is fair to say that making comparative judgments of rather intangible criteria (e.g., overall alteration and/or mineralization) results not only in imprecise knowledge, but also in inconsistency in our own judgments. The improvement of knowledge by controlling inconsistencies in the judgments of experts, that is, the consistency driven approach, is not only desirable but is essential.

In practice, inconsistent judgments are unavoidable when at least three factors are independently compared against each other. For example, let us look closely at the ratios of the four criteria A, B, C, and D in Figure $C1$. Suppose we estimate ratios A/B as 2, B/C as 3, and A/C as 5. Evidently something does not add up as $(A/B)@\ (B/C) = 2 \cdot 3 = 6$ is not equal to 5 (that is A/C). With an inconsistency index of 0.17, the above triad (with highlighted values of 2, 5, and 3) is the most inconsistent in the entire matrix (reciprocal values below the main diagonal are not shown in Figure $C1$). A rash judgment may lead us to believe that A/C should indeed be 6, but we do not have any reason to reject the estimation of B/C as 2.5 or A/B as 5/3. After correcting B/C from 3 to 2.5, which is an arbitrary decision usually based on additional knowledge gathering, the next most inconsistent triad is (5,4,0.7) with an inconsistency index of 0.13. An adjustment of 0.7 to 0.8 makes this triad fully consistent ($5 \cdot 0.8$ is 4), but another triad (2.5,1.9,0.8) has an inconsistency of 0.05. By changing 1.9 to 2 the entire table becomes fully consistent. The corrections for real data are done on the basis of professional experience and case knowledge by examining all three criteria involved.

An acceptable threshold of inconsistency is 0.33 because it means that one judgment is not more than two grades of the scale 1 to 5 away (an off-by-two error) from the remaining two judgments. There was no need to continue decreasing the inconsistency, as only its high value is harmful; a very small value may indicate that the artificial data were entered hastily without reconsideration of former assessments.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 309–316
DOI: 10.15439/2016F80
ACSIS, Vol. 8. ISSN 2300-5963

Leukocyte subtypes classification by means of image processing

Oleg Ryabchykov*†, Anuradha Ramoji*‡, Thomas Bocklitz*†, Martin Foerster§ Stefan Hagel‡¶, Claus Kroegel§, Michael Bauer‡, Ute Neugebauer*†‡, Juergen Popp*†‡

Emails: oleg.ryabchykov@uni-jena.com, thomas.bocklitz@uni-jena.de, juergen.popp@leibniz-ipht.de
* Leibniz Institute of Photonic Technology, Albert-Einstein-Str. 9, 07745 Jena, Germany
† Institute of Physical Chemistry and Abbe Center of Photonics, FSU Jena, Helmholtzweg 4, 07743 Jena, Germany
‡ Center for Sepsis Control and Care (CSCC), Jena University Hospital, Erlanger Allee 101, 07747 Jena, Germany
§ Clinic for Internal Medicine I, Department of Pneumology and Allergy/Immunology, Jena University Hospital, Erlanger Allee 101, 07747 Jena, Germany
¶ Center for Infectious Diseases and Infection Control, Jena University Hospital, Erlanger Allee 101, 07747 Jena, Germany

***Abstract*—The classification of leukocyte subtypes is a routine method to diagnose many diseases, infections, and inflammations. By applying an automated cell counting procedure, it is possible to decrease analysis time and increase the number of analyzed cells per patient, thereby making the analysis more robust. Here we propose a method, which automatically differentiate between two white blood cell subtypes, which are present in blood in the highest fractions. We apply generalized pseudo-Zernike moments to transfer morphological information of the cells to features and subsequently to a classification model. The first results indicate that information from the morphology can be used to obtain efficient automatic classification, which was demonstrated for the leukocyte subtype classification of neutrophils and lymphocytes. The approach can be extended to other imaging modalities, like different types of staining, spectroscopic techniques, dark field or phase contrast microscopy.**

I. Introduction

WHITE blood cells (WBCs) are also called leukocytes. These cells protect the body from infections caused by viruses and other foreign invaders like bacteria or fungi, which make WBCs an important part of the immune system. Leukocytes are produced and derived from the bone marrow and circulate through the bloodstream. A change of the number of different WBC subtypes in the blood is utilized as marker for various diseases. Therefore a blood cell count is often utilized for a routine health examination or diagnosis of specific conditions of a patient. There are five major subtypes of WBCs [1], [2]:

- neutrophils (50-70%);
- lymphocytes (25-30%);
- monocytes (3-9%);
- eosinophils (0-5%);
- basophils (0-1%).

The ranges within the brackets display the percentage of the corresponding cell subtypes in the blood, which are typical ratios for a healthy person. There are various classification approaches, which can be roughly divided into manual and automated methods of cell classification.

The manual classification is performed by a pathologist through the subjective recognition of cell subtypes on microscopic images of stained cells. This type of analysis does not require complex equipment or highly specialized chemical reagents. To simplify the identification, cells are usually stained with the Kimura stain, which colors cell nuclei in blue. Manual differentiation between varying subtypes is accomplished based on characteristics of the cell morphology, like cell size, transparency, granularity, and the shape of the cell nucleus, which are the major differences between the subtypes. Manual classification is widely used in some specific cases of diagnosis and as a “gold standard” for scientific purposes. However, variation of cell morphology within the same cell subtype is very high, and manual classification efficiency is dependent on the pathologist’s qualification and experience.

On the other side, there are various automated classification methods, based on different physical and chemical characteristics of the cells. The main advantage of the automated devices is that they efficiently analyze large number of cells in a short time. Unfortunately, their analyzing workflows include very specific combinations of chemical and physical processes. The complexity of the analysis does not allow the design of a simple portable device. Therefore, automated blood cell counting machines are usually big and expensive.

An alternative approach is an automatic image analysis of microscopic images of stained cells. In a combination with a small camera this method can become a useful tool for doctors, providing them an instant access to the information about WBCs population at bedside of a patient. There are some studies that show efficient leukocyte identification [3], [4] and segmentation [5], [6] within microscopic images. However, these studies are focused on the leukocyte count without the classification of the leukocytes into subtypes. That leads to the loss of important information about the proportions of each cell subtype. In distinction to the mentioned studies, the current manuscript describes an algorithm for the classification

of WBCs, focusing on the textural features analysis of single cell images.

The concept of the work is to extract quantitative features related to the cell morphology from the microscopic images. Subsequently, these features are used to train and evaluate a statistical model for cell subtype identification. Moreover, the same type of images as for manual classification is used, therefore, this approach allows a direct comparison to the "gold standard". In order to use these images for an automated image analysis, standardization and preprocessing have to be carried out. However, during the pretreatment step, it is important to eliminate corrupting effects, such as uniformities in staining and lighting, but to keep the morphological information for further analysis steps.

The textural information extraction from preprocessed images can be carried out by various methods [7], [8]. However, image description by means of pseudo-Zernike (PZ) moments [9] was chosen for the cell subtype identification because it was proven to be a reliable method for the recognition of shapes [10], characters [11], [12], faces [13], [14], [15], and viruses [16]. An advantage of the representation by PZ-moments is that their absolute values are independent from image rotation, which is necessary due to random orientation of the cells on a microscopic slide. The PZ-moments are derived from PZ-polynomials, which are orthogonal to each other and can be used in further statistical analysis, thus an automated classification technique can be established.

The proposed automated cell classification method is aimed to combine the simplicity of the manual classification and the advantages of automatization. The approach is based on the analysis of images, which are similar to the images used for manual "gold standard" method and are produced by common microscopy from a blood sample after non-complicated preparation. On the other side, due to automatization, extremely short classification times and objectivity, comparable with a human observer, can be achieved.

II. Materials and Methods

A. Sample preparation

Leukocytes were isolated from the venous blood of patients admitted to the intensive care unit with informed consent according to the Ethics Committee of the Jena University Hospital (Ethic vote n 4004-02/14). Briefly, 2.7ml of blood in ethylenediaminetetraacetic acid (EDTA) was drawn freshly from an existing catheter using the BD monovettes. In case of healthy donor, blood (about 100μl) was collected from fingertip using lancet. Red blood cell lysis was carried out by mixing the blood with an ammonium chloride solution with a ratio of 1:5 in a 50ml falcon tube. After 5 minutes of incubation at room temperature (RT), the mixture was centrifuged for 10 minutes at 400g at RT. The WBC pellet at the bottom of the falcon tube was collected by discarding the supernatant and suspending it in a phosphate buffer solution (PBS). The WBCs were chemically fixed with 4% formaldehyde for 10 minutes, followed by washing the cells successively with PBS and 0.9% NaCl. The cells were coated on slides using cytospin and stained with a Kimura staining solution (which stains only the cellular nucleus) and washed with distilled water. The slides were dried at RT and stored at $4\,^{\circ}\mathrm{C}$ for maximum one hour until further use. The Kimura stained images of the WBCs (Fig. 1 *a,b*) were captured with an upright epifluorescence microscope (Axioplan 2, Carl Zeiss, Germany) equipped with an AxioCam HRc camera (Carl Zeiss, Germany). Images were acquired using Zeiss Axio Vert software (Carl Zeiss, Germany).

B. Calculations

All calculations reported in this work were carried out in Gnu R (version 3.0.2) [17] running on a Windows 7 Professional 64-bit system (Intel® Core™ i5-4570 CPU @ 3.20 GHz 2.70 GHz with 8GB RAM). In addition to the base R package, which contains the input/output, basic programming support, and arithmetic functions, some more specific algorithms were utilized from other packages. For orthogonal moment analysis the "IM" package [18] was used. A support vector machine (SVM) classification model was built with the "e1071" package [19]. Parallel computing was obtained by functions from "foreach" [20] and "doParallel" [21] package. K-means clustering from the "stats" package [17] was utilized for the background removal. The functions for principal component analysis (PCA), nonlinear least squares estimation, and the fast Fourier transform (FFT) are all contained in the base package [17]. JPEG files were loaded into the R environment via the "jpeg" package [22].

Prior to analysis, each image was converted from *sRGB* color space to *Lab* color space, one of the most common color spaces for image analysis applications. It was chosen due to the fact that, unlike additive or subtractive color models (for example *RGB* or *CMYK*), it is not optimized for image representation on a screen or for printing, but is adapted to cover the entire range of colors distinguishable by the human eye and to match the perception of these colors. In this color space, a and b components are related to chromatic color values. The L component of *Lab* color space closely matches the human perception of lightness, which allows to expect that in this representation cell subtypes can be identified based on their morphology. The conversion of the color space was performed by base R function "convertColor".

Subsequently to the color space conversion, other steps, such as noise reduction, background removal and intensity normalization were performed. The details of these preprocessing steps are described in the "Results and discussion" section.

C. Pseudo-Zernike (PZ) Moments

As mentioned previously, PZ-moments were chosen for feature extraction from the images. These orthogonal, complex-valued moments are defined on a unit disk and are widely used for pattern recognition. The PZ-moments can describe a 2-dimensional function on the unit circle. However, the function $f(x, y)$ can represent an image if two arguments, x and y, are related to a pixel position and the function value is related to

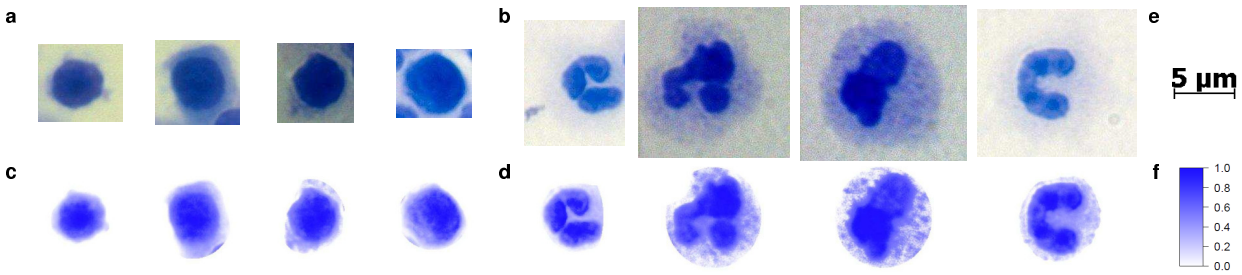


Fig. 1. Original images of two Kimura stained cell subtypes from the patients are displayed in the first row: lymphocytes (a), which are characterized by deep staining of the nuclei and a relatively small amount of cytoplasm, and neutrophils (b), which are the most common subtype that normally contain a nuclei divided into 2-5 lobes. All images are sized according to the scale (e). At the bottom preprocessed false-color equivalents of the presented images (c, d) normalized to the unit scale (f) are shown.

lightness or another color component in that pixel. The PZ-moments (A_{nl}) of an image on a unit disk are defined in radial coordinates by [23]:

$$A_{nl} = \frac{n+1}{\pi} \int_0^{2\pi} \int_0^1 [V_{nl}(r\cos\theta, r\sin\theta)]^* f(r\cos\theta, r\sin\theta)\, r dr d\theta \ .$$

In this equation $n = 0, \ldots, \infty$ represents the order, the repetition is denoted by $l \leq n$, and f is the value related to the current pixel position: $0 \leq r \leq 1$ and $0 \leq \theta \leq 2\pi$ (polar coordinates of the pixel). V_{nl} is the orthogonal set of complex-valued PZ-polynomials, which can be written as:

$$V_{nl}(r,\theta) = R_{nl}(r)\, e^{jl\theta} \ ,$$

where R_{nl} represents radial polynomials with integer coefficients $D_{n,m,s}$:

$$R_{nl}(r) = \sum_{s=0}^{n-|l|} D_{n,|l|,s} r^{(n-s)} \ ,$$

$$D_{n,m,s} = \frac{(-1)^s (2n+1-s)!}{s!(n-m-s)!(n+m-s+1)!} \ .$$

Both the order n and the repetition l are related to the spatial frequencies of the image. However, the order n represents the spatial frequency along the unit disk's radius, while the repetition l represents the spatial frequency along the unit disk's angular coordinate. Moreover, by clarifying the idea behind order n and repetition l, the respective moments can be interpreted. Therefore, the classification model can be checked, analyzed and the morphological differences between the cell subtypes can be examined.

As it is seen from the formulas, the angular coordinate is included in the PZ-moments only within the multiplier $e^{jl\theta}$, which is related to the phase of the complex value [9], [10], [12]. Due to this fact, the absolute values of moments are independent from a rotation of the coordinate system. Thus, they are independent from the spatial alignment of the cell within the image and from the orientation on the microscopic slide. Other advantages of these particular moments are their low sensitivity to noise [10] and that the PZ-moments are orthogonal to each other.

III. Results and Discussion

A. Data set

Taking into account the extremely low number of monocytes, eosinophils, and basophils in the data, only two major subtypes could be investigated in the current study. These both subtypes represent about 90% of WBCs in the blood and were included in the statistical evaluation. Thus, the training data included 28 lymphocytes and 45 neutrophils from 6 patients which were showing signs of inflammation. On the other side, the test data included 128 cells from two healthy volunteers. Unlike the training set, where some cell subtypes were sorted out, the test data included randomly selected cells without presorting or labeling according to their subtypes.

The cell subtypes included in the training data are different in sizes and cell nuclei morphology (see Fig. 1). Most notable is that the neutrophils are relatively big and have multi-lobed nuclei, while lymphocytes have almost round nuclei and are smaller. Other WBC subtypes, which were not included in the training data, are characterized by their granularity and the following properties of the cell nuclei: monocytes have kidney shaped nuclei, eosinophils have relatively small bi-lobed nuclei, and basophils have bi-lobed or tri-lobed nuclei. Although each subtype has a typical average cell size and other specific characteristics, each single cell varies from that average characteristics, which make some of its parameters dissimilar to the typical characteristics of its subtype.

B. Workflow

To obtain a stable and efficient analytical system, an image processing workflow was developed and optimized for the specific task of leukocyte subtype classification. The data was loaded, preprocessed, and represented as a set of pseudo-Zernike moments based invariants for further analysis. The workflow is presented in more detail in Fig. 2.

Important and nontrivial steps are the image preprocessing and standardization, which have to be optimized. These procedures should reduce the variations of brightness and color tones between the images of cells within the same sample and occasional appearing variations caused by the sample preparation routine for images taken from different samples. If

the workflow presented here is applied to other imaging modalities, like holographic imaging and phase contrast microscopy, these variations are expected to be less significant. Therefore, the preprocessing procedure has to be modified individually for each microscopic imaging technique and classification task.

For the construction of the classification model based on image analysis, the measured cells were labeled according to the classification made by the pathologist. The labeled and preprocessed training data were subsequently divided into three batches for cross-validation of the model. This step of the workflow was of enormous importance for setting model parameters and estimating the model quality. Thereafter, it should not be underestimated.

Leave-batch-out-cross-validation of SVM classification was performed on the training data with different combinations of input variables. This cross-validation procedure was designed to avoid any relations between different batches of cells. Therefore, the data splitting into three batches was arranged so, that the batch reflect the measurement dates and patient's origin. Thus, the generalization performance for the prediction of an independent dataset is well estimated by the leave-one-patient-out-cross-validation. Consequently, classification models with various numbers of PZ-moments' orders and principal components were compared. The variable selection was carried out according to the highest sensitivity for cross-validation of SVM classification model. The model with highest sensitivity was chosen as an optimal one and further used for the test data prediction.

Besides high identification efficiency, the proposed algorithm has to be suitable for real-life applications. Therefore, the workflow was optimized by parallelization of each single image loading, preprocessing, and calculation of the moments. Thereby, the parallelization on hardware with a multi-core processor should decrease the calculation time for a large amount of data roughly by a factor related to the number of calculation units. We chose the number of clusters for parallel calculation as one less than the number of processor cores, which was three for the PC on which the analysis was performed. During the preliminary study stage, the amount of data was relatively small, and thus, the parallelization of calculations had a negligible effect. However, despite the insignificant improvement on a small data set, parallelization is highly important for further applications and implementation of the algorithm, especially for the case if the number of analyzed cells is on the order of thousands.

C. Preprocessing

Examples of WBC images are shown in Fig. 1 *a,b*. As it can be seen by naked eye, differences between some images, which are not related to the cell's morphology, occur. These fluctuations originate from the sample preparation procedure, which is simple and standardized. There are some systematic deviations between the cells of different patients, but also the images of cells from the same patient can differ due to the spatial alignment of the cells and non-uniform coloring of samples along microscopic slides. Moreover, parts of other

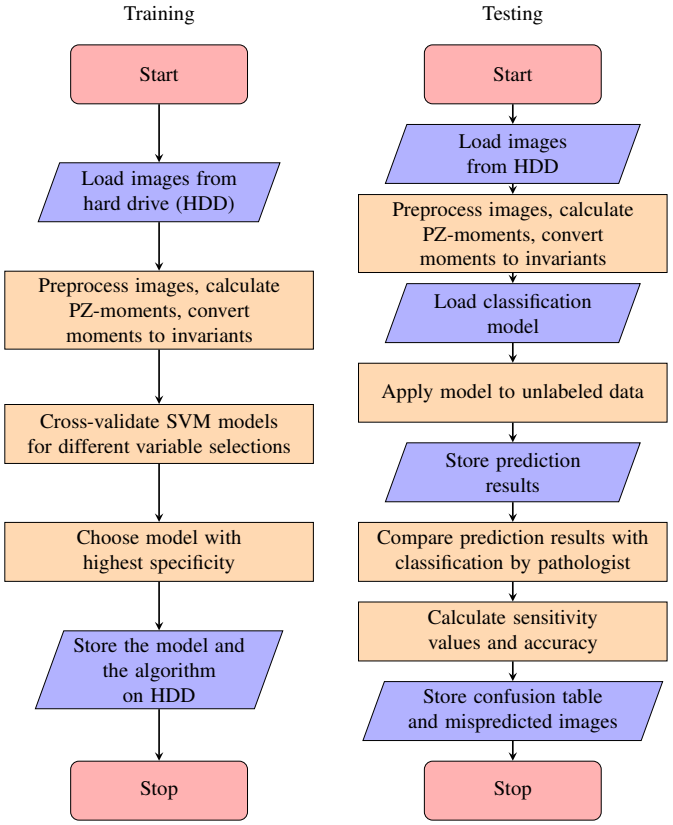


Fig. 2. Schematic workflow of the presented algorithm and the model validation.

cells are visible within some images and, additionally, other influences on the brightness, contrast, and tone are present on the microscopic images. To reduce the discussed corrupting effects, an advanced preprocessing has to be carried out before the feature extraction procedure.

According to the chosen concept of the analysis, it was important to keep the morphological features which can be distinguished visually. The automation of the preprocessing procedure took an important part in the development of the algorithm. The original images were stored in the standard *sRGB* representation, which is designed to display images in electronic systems, such as a computer's screens. However, analysis of the color channels separately from each other can be problematic and leads to a high complexity of the classification model. Switching to a single component can be circumvented by applying a more convenient color space. As it was mentioned in "Materials and methods", the lightness L of *Lab* color space is closely related to the human visual perception of images. In order to keep the features used for manual classification, the *Lab* color space was used in the further analysis. Moreover, the cells used for analysis were colored by Kimura staining, which highlights the cell nucleus in blue. Due to monochromatic coloring, all variations of the chromatic values are only related to the deviations of the sample preparation process and staining. Thus, related color components (a and b) were skipped and only the lightness

L was analyzed. However, for staining procedures which stain different cell organelles or cytoplasm in different colors, normalized a and b components should be also included in the analysis.

Due to high variations between different images, even for the single L-component, the automation of preprocessing took an important part in the analysis development. Pretreatment was aimed to decrease deviations of the features extracted from images within the same cell subtype and to increase the overall identification accuracy. Consequently, the background, or non-cell area of the images, was cut off via the unsupervised k-means clustering of lightness values within each image. In order to improve the background removal, an FFT-filter was applied to the images prior to the clustering. After the background removal, the lightness distribution within each cell was standardized by means of normalization to the unit interval and equalization of the histogram.

Subsequently to the lightness standardization of the images, a 2-dimensional Gaussian function was fitted to each cell image using nonlinear least squares. Based on the coefficients of the fitted function, centers and estimated radii were determined for each cell. As the next step, background-free images of the single cells were cropped according to the estimated cells' radii. This procedure was performed, to preserve the full region of the stained nucleus with a cytoplasm area and to exclude regions of other cells, non-cell area, or unexpected artifacts which were present in some images outside of the cell area. After cropping, images were placed on frames with a determined preset size, which was chosen to fit the biggest cell expected among the analyzed cell subtypes: 13x13 μm, which was equivalent to 200x200 pixels. On this step the centers of the cells were also matched to the centers of the frames. Pretreated images are shown in Fig. 1 *c,d*.

D. Features extraction

As quantitative features which can be used to describe the morphology of cell images, the complex-valued pseudo-Zernike moments where chosen. However, the position of each individual cell on a slide is random and it is necessary to operate with rotationally independent features. Since the phase of the moment is related to the angular coordinate within the image plane, complex-valued moments were converted to absolute PZ-moments and then normalized to the zero-order moment. Therefore, invariants, which are not dependent on the image rotation and scale, were produced. These invariants skip all information about the phase (angular coordinate), and thus, the obtained variables are independent of the image rotation.

Unfortunately, as it is shown above in the "Materials and methods" section, the calculation of PZ-moments requires a double integration of a two-dimensional function which is a costly CPU process. Because the pre-computed images were transferred to a frame with a preset size, the algorithm for the PZ- moment calculation can be simplified. Instead of the integration, the sum of a scalar product of the image with a pre-computed complex matrix can be used. The matrices related to each moment can be generated once and then stored on a hard disk drive for the further use.

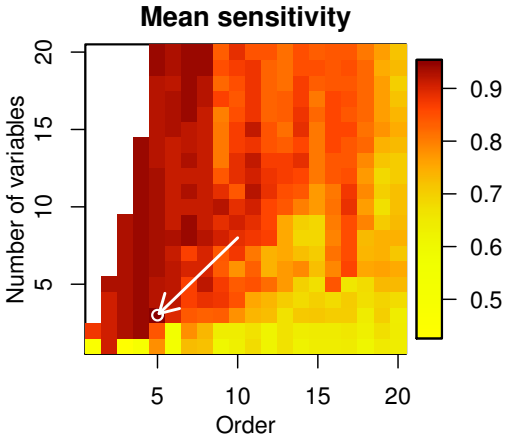


Fig. 3. Mean sensitivity of SVM leave-batch-out-cross-validation of training data. Classification models were created for a different number of selected orders of moments and for a different number of principal components (called variables in the image). The maximum value, which is related to the optimal model, is indicated with a white arrow.

TABLE I
CONFUSION TABLE FOR THE LEAVE-BATCH-OUT-CROSS-VALIDATION OF THE SVM MODEL WITH OPTIMAL VARIABLE SELECTION.

		Predicted		
		Lymphocytes	**Neutrophils**	**Sensitivity**
True	**Lymphocytes**	26	2	0.893
	Neutrophils	1	44	0.978

E. Statistical model establishment and evaluation

To avoid an overfitting of the statistical model, the dimensionality of the data was reduced. A dimension reduction was obtained via a principal component analysis (PCA). The dimensionality of the retaining data set was optimized based on a leave-batch-out-cross-validation of the training data set. The parameter intervals checked for the feature extraction was 1 to 20 for orders, while repetition was chosen maximal. The score dimension of the PCA was evaluated from 1 to 20. For each parameter set the model performance was estimated based on the mean sensitivity. These values are summarized in plot Fig 3. The maximal sensitivity is marked on the plot with an arrow. This parameter set defines the optimal combination of input variables (3 principal components, based on PZ-moments up to 5th order). The model trained with these parameters was further analyzed and visualized. In table I a confusion table of training data cross-validation is given. In Fig. 4 a histogram of its probability scores, which represents SVM decision values rescaled to the unit range, is plotted.

F. Blind prediction

Model validation was performed by applying the established model to the independent data, which contained 163 microscopic images of stained WBCs. All preprocessing and feature extracting steps were performed on these unlabeled

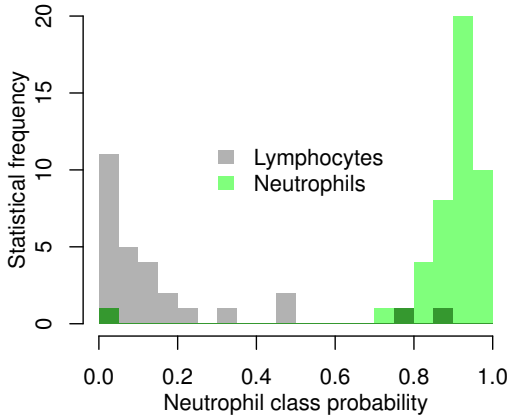


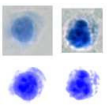
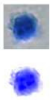
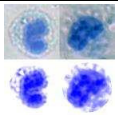
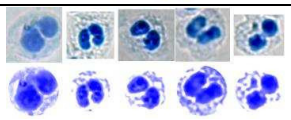
Fig. 4. Histogram for SVM posterior probabilities calculated by a leave-batch-out-cross-validation of the training data with the optimal number of variables are shown. Classification was performed between lymphocytes (gray bars) and neutrophils (green bars). The overlap of the groups is indicated within the histogram by dark green bins.

images in the same way as for the training data. In order to avoid the influence of the operator's subjectivity, a double blind prediction was carried out. Images were classified in manual mode by an experienced pathologist independently from the automated prediction. Subsequently, the statistical predictions were compared with the manual classification results. A summary of the results is visualized by a confusion table (see table II). Another representation of the classification performance is shown by means of a ROC curve in Fig 5. This curve, built for the threshold of the SVM decision values of the test data prediction, illustrates the high performance of the prediction. Moreover, the area under ROC curve (AUC) is about 0.984, which indicates an almost perfect classification. A perfect binary classification is characterized by an AUC equal to 1. Among 155 cells, which were classified as lymphocytes or neutrophils in manual mode, three images were wrongly identified by the statistical model. Such a low misclassification rate of independent test data corresponds to a high accuracy of the 2-class prediction. This accuracy was higher as 97%. Additionally, cells of the subtypes, which were not included in the training set, were present in the test data. These cells (five eosinophils and two monocytes), were predicted within the same class as neutrophils. This behavior was expected, since they feature a similar morphology as neutrophils compared to lymphocytes. Additionally, neutrophils, eosinophils, and monocytes feature a higher biological similarity and higher subjective similarity of the images. These classification results of the eosinophils and monocytes indicate that an extension of the presented model may be possible. A hierarchic layout of the classification seems optimal to incorporate eosinophils and monocytes.

IV. Conclusion

In this work, we presented an algorithm for a highly efficient classification between two dominant subtypes of leukocytes. The special feature of the proposed method is that by means

TABLE II
CONFUSION TABLE FOR THE PREDICTION OF THE UNLABELED TESTING DATA. CORRECT PREDICTED CELLS ARE SPECIFIED ONLY WITH THE QUANTITY OF THE IDENTIFIED CELLS. ALL INCORRECTLY PREDICTED CELLS AND CELLS, THAT RELATE TO OTHER SUBTYPES, WHICH WERE NOT INCLUDED IN THE TRAINING DATA, ARE SHOWN IN THE TABLE AS UNTREATED MICROSCOPIC (UPPER ROWS) AND PREPROCESSED (BOTTOM ROWS) IMAGES.

		Predicted (assorted by statistical model)	
		Lymphocytes	Neutrophils
True (assorted by pathologist)	Lymphocytes	17	
	Neutrophils		101
	Monocytes	0	
	Eosinophils	0	

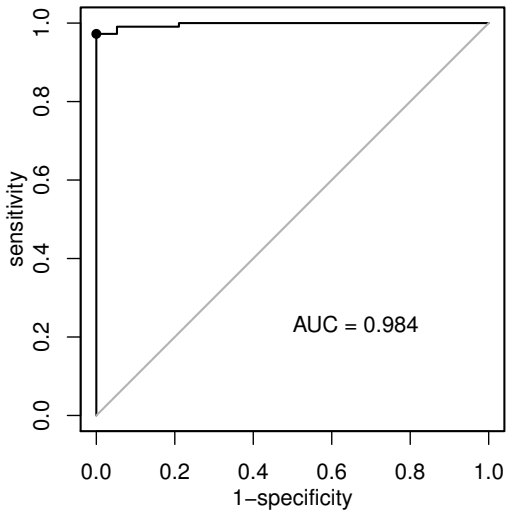


Fig. 5. The ROC curve and area under the curve (AUC) illustrate the high performance of the SVM prediction of the binary classification model between two WBC subtypes (lymphocytes and neutrophils) for independent unlabeled testing data.

of PZ-invariants the cell morphology is represented as a quantitative marker for the cell subtypes. Therefore, a combination of such common statistical methods as principal component analysis and support vector machine classification was applied to build the classification model. This approach showed a high stability against patient to patient and sample to sample variations. Moreover, an advanced image preprocessing made a further contribution to the robustness of the model. The standardization of the images decreased deviations, which occur between samples due to the sample preparation routine. Additionally, the automated framing and centering of the analyzed images of cells led to the replacement of the double numerical integration, performed for PZ-moment calculation, with a matrix product. This simplification of the calculation procedure resulted in the reduction of computation time and allowed the analysis to be performed in real-time. The classification results showed that WBCs subtypes as monocytes and eosinophils (which were not included in the model due to their low quantity in the training data) were predicted within the same class. Due to this fact, it can be assumed that the classification can be improved and extended to other cell types by a multilevel model. However, that requires a statistically significant amount of microscopic images for each leukocyte subtype in the training data set. The described approach can be applied for microscopy images taken of other staining types. Only important is that the images display the cell morphology. The method presented here may be also applied to images obtained with techniques such as fluorescence, dark field, or phase contrast microscopy.

ACKNOWLEDGMENT

Financial support of the BMBF via the integrated research and treatment center "Center for Sepsis Control and Care" (FKZ 01EO1502) and from the EU via the project "Hemo-Spec" (FP 7, CN 611682) is highly acknowledged. The authors thank Katrin Ludewig and Frank Brunkhorst for data collection and their support.

REFERENCES

[1] A. Kratz, M. Ferraro, P. M. Sluss, and K. B. Lewandrowski, "Normal reference laboratory values," *New England Journal of Medicine*, vol. 351, no. 15, pp. 1548–1563, 2004. doi: 10.1056/NEJMcpc049016 PMID: 15470219. [Online]. Available: http://dx.doi.org/10.1056/NEJMcpc049016

[2] A. Ramoji, U. Neugebauer, T. Bocklitz, M. Foerster, M. Kiehntopf, M. Bauer, and J. Popp, "Toward a spectroscopic hemogram: Raman spectroscopic differentiation of the two most abundant leukocytes from peripheral blood," *Analytical Chemistry*, vol. 84, no. 12, pp. 5335–5342, 2012. doi: 10.1021/ac3007363 PMID: 22721427. [Online]. Available: http://dx.doi.org/10.1021/ac3007363

[3] S. Khan, A. Khan, F. S. Khattak, and A. Naseem, "An accurate and cost effective approach to blood cell count," *International Journal of Computer Applications*, vol. 50, no. 1, 2012. doi: 10.5120/7734-0682. [Online]. Available: http://dx.doi.org/10.5120/7734-0682

[4] L. Putzu and C. Di Ruberto, "White blood cells identification and counting from microscopic blood image," *International Journal of Medical, Health, Biomedical, Bioengineering and Pharmaceutical Engineering*, vol. 7, no. 1, pp. 20 – 27, 2013. [Online]. Available: http://waset.org/Publications?p=73

[5] M. M. G. Bhamare and D. Patil, "Automatic blood cell analysis by using digital image processing: A preliminary study," in *International Journal of Engineering Research and Technology*, vol. 2, no. 9, ESRSA Publications. ESRSA Publications, 2013. [Online]. Available: http://www.ijert.org/view-pdf/5460/

[6] F. Sadeghian, Z. Seman, A. R. Ramli, B. A. Kahar, and M.-I. Saripan, "A framework for white blood cell segmentation in microscopic blood images using digital image processing," *Biological procedures online*, vol. 11, no. 1, pp. 196–206, 2009. doi: 10.1007/s12575-009-9011-2. [Online]. Available: http://dx.doi.org/10.1007/s12575-009-9011-2

[7] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *Systems, Man and Cybernetics, IEEE Transactions on*, vol. SMC-3, no. 6, pp. 610–621, 1973. doi: 10.1109/TSMC.1973.4309314. [Online]. Available: http://dx.doi.org/10.1109/TSMC.1973.4309314

[8] M. Habibzadeh, A. Krzyżak, and T. Fevens, *Comparative study of feature selection for white blood cell differential counts in low resolution images*, ser. Lecture Notes in Computer Science. Springer International Publishing, 2014, vol. 8774, book section 20, pp. 216–227. ISBN 978-3-319-11655-6. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-11656-3_20

[9] T. Xia, H. Zhu, H. Shu, P. Haigron, and L. Luo, "Image description with generalized pseudo-Zernike moments," *Journal of the Optical Society of America A*, vol. 24, no. 1, pp. 50–59, 2007. doi: 10.1364/JOSAA.24.000050. [Online]. Available: http://josaa.osa.org/abstract.cfm?URI=josaa-24-1-50

[10] S. O. Belkasim, M. Shridhar, and M. Ahmadi, "Pattern recognition with moment invariants: A comparative study and new results," *Pattern Recognition*, vol. 24, no. 12, pp. 1117–1138, 1991. doi: 10.1016/0031-3203(91)90140-Z. [Online]. Available: http://www.sciencedirect.com/science/article/pii/003132039190140Z

[11] C. Kan and M. D. Srinath, "Invariant character recognition with Zernike and orthogonal Fourier-Mellin moments," *Pattern Recognition*, vol. 35, no. 1, pp. 143–154, 2002. doi: 10.1016/S0031-3203(00)00179-5. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0031320300001795

[12] C. W. Chong, P. Raveendran, and R. Mukundan, "The scale invariants of pseudo-Zernike moments," *Pattern Analysis and Applications*, vol. 6, no. 3, pp. 176–184, 2003. doi: 10.1007/s10044-002-0183-5. [Online]. Available: http://dx.doi.org/10.1007/s10044-002-0183-5

[13] Y.-H. Pang, A. T. B. J, and D. N. C. L, "Enhanced pseudo Zernike moments in face recognition," *IEICE Electronics Express*, vol. 2, no. 3, pp. 70–75, 2005. doi: 10.1587/elex.2.70. [Online]. Available: http://dx.doi.org/10.1587/elex.2.70

[14] E. Walia, C. Singh, and N. Mittal, "Discriminative Zernike and pseudo Zernike moments for face recognition," *Int. J. Comput. Vis. Image Process.*, vol. 2, no. 2, pp. 12–35, 2012. doi: 10.4018/ijcvip.2012040102. [Online]. Available: http://dx.doi.org/10.4018/ijcvip.2012040102

[15] J. Haddadnia, M. Ahmadi, and K. Faez, "An efficient feature extraction method with pseudo-Zernike moment in rbf neural network-based human face recognition system," *EURASIP Journal on Advances in Signal Processing*, vol. 2003, pp. 890–901, 2003. doi: 10.1155/s1110865703305128. [Online]. Available: http://dx.doi.org/10.1155/s1110865703305128

[16] T. Bocklitz, E. Kämmer, S. Stöckel, D. Cialla-May, K. Weber, R. Zell, V. Deckert, and J. Popp, "Single virus detection by means of atomic force microscopy in combination with advanced image analysis," *Journal of Structural Biology*, vol. 188, no. 1, pp. 30 – 38, 2014. doi: 10.1016/j.jsb.2014.08.008. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1047847714001841

[17] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2015. [Online]. Available: https://www.R-project.org/

[18] B. Rajwa, M. Dundar, A. Irvine, and T. Dang, *IM: Orthogonal Moment Analysis*, 2013, R package version 1.0. [Online]. Available: https://CRAN.R-project.org/package=IM

[19] D. Meyer, E. Dimitriadou, K. Hornik, A. Weingessel, and F. Leisch, *e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien*, 2015, R package version 1.6-7. [Online]. Available: https://CRAN.R-project.org/package=e1071

[20] Revolution Analytics and S. Weston, *foreach: Provides Foreach Looping Construct for R*, 2015, R package version 1.4.3. [Online]. Available: https://CRAN.R-project.org/package=foreach

[21] ——, *doParallel: Foreach Parallel Adaptor for the 'parallel' Package*, 2015, R package version 1.0.10. [Online]. Available: https://CRAN.R-project.org/package=doParallel
[22] S. Urbanek, *jpeg: Read and write JPEG images*, 2014, R package version 0.1-8. [Online]. Available: https://CRAN.R-project.org/package=jpeg
[23] C. H. Teh and R. T. Chin, "On image analysis by the methods of moments," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 10, no. 4, pp. 496–513, Jul 1988. doi: 10.1109/34.3913. [Online]. Available: http://dx.doi.org/10.1109/34.3913

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 317–321
DOI: 10.15439/2016F333
ACSIS, Vol. 8. ISSN 2300-5963

Random Subspace Ensemble Artificial Neural Networks for First-episode Schizophrenia Classification

Roman Vyškovský
Masaryk University, Research Centre for Toxic Compounds in the Environment (RECETOX), Kamenice 3, 625 00 Brno, Czech Republic
Email: vyskovsky@recetox.muni.cz

Daniel Schwarz, Eva Janoušová
Masaryk University, Institute of Biostatistics and Analyses, Kamenice 3, 625 00 Brno, Czech Republic
Email: {schwarz, janousova}@iba.muni.cz

Tomáš Kašpárek
Masaryk University and University Hospital Brno, Department of Psychiatry, Jihlavska 20, Brno, Czech Republic
Email: tkasparek@fnbrno.cz

***Abstract*—Computer-aided schizophrenia diagnosis is a difficult task that has been developing for last decades. Since traditional classifiers have not reached sufficient sensitivity and specificity, another possible way is combining the classifiers in ensembles. In this paper, we take advantage of random subspace ensemble method and combine it with multi-layer perceptron (MLP) and support vector machines (SVM). Our experiment employs voxel-based morphometry to extract the grey matter densities from 52 images of first-episode schizophrenia patients and 52 healthy controls. MLP and SVM are adapted on random feature vectors taken from predefined feature pool and the classification results are based on their voting. Random feature ensemble method improved prediction of schizophrenia when short input feature vector (100 features) was used, however the performance was comparable with single classifiers based on bigger input feature vector (1000 and 10000 features).**

I. Introduction

SCHIZOPHRENIA (SZ) is a severe and chronic neurodevelopmental disorder with unknown etiology. Patient's response to the treatment is uncertain and early diagnosis could increase the probability of remission. Since nowadays the diagnostics is based on interview, self-report and psychiatrist's observation, there are efforts to develop a diagnostic tool that could support establishing diagnosis of the first episode of schizophrenia in a more objective way.

Involvement of modern imaging methods in the last decades has opened up new possibilities in brain research. These methods include for instance, magnetic resonance imaging (MRI), computed tomography or positron emission tomography. Especially MRI techniques offer good contrast and spatial resolution. Thus, morphological abnormalities and relations between brain structures and functions can be studied with the use of imaging data. Since the differences between schizophrenia patients and healthy controls (HC) have been already found with the use of manual segmentation of region of interests [1] or automated morphometry methods such as voxel-based morphometry [2] or deformation-based morphometry [3], many scientists have been recently trying to create computer-aided diagnostic tools based on neuroimaging data. The outcomes of such tools have not reached sufficient sensitivity and specificity for implementation into the psychiatric clinical practice yet, and hence the demand still persists. The application of classification methods with self-adapting strategies known as machine learning is a challenging task in the schizophrenia research.

Artificial neural network (ANN) is a model inspired by how the brain works. Since the backpropagation algorithm [4] was invented as a technique for learning ANNs, they have been used widely in many applications and have achieved success at least in two areas of brain image processing: segmentation [5], [6] and classification [7]–[11]. Those results have shown that ANNs deserve attention of neuroimaging community investigating how to recognize mental diseases in imaging data.

Several authors have already tried to classify schizophrenia based on diffusion tensor data [7] using several types of neural networks: backpropagation neural networks, radial basis function networks, learning vector quantization neural networks and probabilistic neural networks. Other studies [8], [12] used backpropagation neural network on functional magnetic resonance imaging (fMRI) data or resting-state fMRI. Other papers discussed the use of ANN for classification of other brain diseases such as Alzheimer's disease [9], [10] or brain cancer [11] – based on structural MRI data.

The power of such single models can be further improved by the means of ensemble learning. This approach employs set of classifiers to determine the object's class by voting. To ensure a necessary assumption, which is the disparity among the classifiers, variability in the training process must be somehow acquired. Many methods have been invented for this purpose, such as random subspace ensemble [13], random forests [14], bagging [15], boosting (e.g. AdaBoost [16]), rotation forests [17] and others. The first one is explored in this paper in combination with ANNs. Several ensemble methods were applied to investigate neuroimaging problems. For instance, Yang et al. [18] classified schizophrenia using ensembles of support vector machines (SVM) trained by modified AdaBoost algorithm that besides boosting performed also simultaneous feature selection from

fMRI and single nucleotide polymorphism data. They reached an overall accuracy of 87% when both data sets were combined. Janousova et al. [19] based their ensemble for schizophrenia prediction on different image features extracted from MRI data (MR intensities, grey matter densities and local deformations) and on three various types of classifiers, and achieved the accuracy of 81.6%. Lebedev et al. [20] used random forests for Alzheimer's disease (AD) detection and achieved overall accuracy of 91%. Liu et al. [21] proposed local patch-based subspace ensemble method combined with a classifier based on sparse representation of data and improved the performance of classification performance of AD and mild cognitive impairment up to 3% compared to the use of a single classifier.

To the best of our knowledge, this is the first time that random subspace ensemble ANNs are used for schizophrenia classification based on the structural MRI data. The paper is organized as follows: Section II outlines the basics of ANNs and random subspace ensemble. The experiment and results of classification are summarized in section III. Section IV discuss the results and concludes the paper.

II. METHODS

A. Dataset and Image Processing

We used the same dataset as in [22]. It consisted of MRI data of 52 schizophrenia patients and 52 age- and sex-matched (only men) healthy control subjects without family history or personal history of axis I psychiatric conditions. Patients' and healthy controls' median age was 22.9 years (range 17-40 years) and 23.0 years (range 18.2-37.8 years) respectively. Images of all subjects were acquired on 1.5 T magnetic resonance imaging machine Siemens Symphony at University Hospital Brno, using a 3-D acquisition with IR/GR sequence, TR 1700 ms, TE 3.93 ms, TI 1100 ms, flip angle 15°, 160 slices, voxel size 1.17 × 0.48 × 0.48 mm, FOV 246 × 246 mm, and matrix size 512 × 512 voxels .

Gray matter tissue segments were obtained from all images after a correction of bias-field inhomogeneity, spatial normalization, segmentation, modulation and Gaussian smoothing. All the steps followed the pipeline of the optimized voxel-based morphometry [23]. The last step - smoothing the gray matter segments with an isotropic Gaussian kernel - spreads the information to the neighboring voxels and compensates the inexact nature of the spatial normalization step [24]. Furthermore it ensures more normally distributed data for further parametric statistical analyses used for feature selection [25]. After this image preprocessing, a binary mask of a brain normalized to the stereotactic space was used to erase the voxels representing extracerebral tissues.

B. Feature Pool Preparation

Before applying the classifiers, it was necessary to select only those features - gray matter density voxels - which represented the information useful to discriminate between the two classes, and conversely to exclude those voxels without any helpful information, and thus to improve signal-to-noise ratio and create more accurate classifier. The feature selection step is also important to tackle the problem of the curse of dimensionality – well-known in neuroimaging community [26].

Voxel-wise two-sample t-tests were the instrument for feature selection. This method selected only those voxels which showed significant differences in gray matter density between the groups. Such a very naïve approach might be less prone to overfitting than selection the features with the highest discrimination power derived from correlation with the classification outcome. In addition, it is important that the whole volume of gray matter was explored with no arbitrarily predefined regions of interests, as morphological abnormalities in schizophrenics' brains have been uncovered with automated brain morphometry methods in many different brain regions [2].

C. Multi-layer Perceptron

Multi-layer perceptron (MLP) is the most traditional type of ANN. It maps relations between inputs and desired outputs. MLP consists of three or more layers formed by neurons - basic computational units - and are adapted in a supervised learning manner using the backpropagation algorithm. The information is passed through the layers in input-output direction. The equation of a neuron is:

$$y = \left(w_0 + \sum_{i=1}^{n} w_i x_i\right) \tag{1}$$

where y is the output, w_i are weights, x_i are inputs, n is the number of neuron inputs, is the activation function – hyperbolic tangent for hidden layers and linear functions in the output layer.

The output layer comprises two neurons; each represents one class, so the subject is classified according to the comparison of their values after the excitation. Furthermore, the softmax function is applied on these two neurons:

$$y_i = \frac{e^{\xi_i}}{\sum_{j=1}^{n} e^{\xi_j}}, \tag{2}$$

where n is the number of outputs, ξ_i and ξ_j are net activations of $i-th$ and $j-th$ output neurons. This function ensures that the outputs are non-negative and their sum is equal to 1. Such results can be interpreted as posterior probabilities [27].

In this project, the MLPs are trained by minimization of cross-entropy using the scaled conjugate gradient backpropagation algorithm that has been found to be fast in preliminary experimentation. Since neural networks have many parameters to be predefined - both for architecture creation and for adaption - it is difficult to set them to the optimal configuration. We kept the implicit learning parameters i.e. learning rate 0.01, maximal number of epochs 1000, minimum gradient 10^{-6} and regarding architecture and we used two-layer network with 10 hidden neurons, which achieved a good performance during experimentation. Since the weights were initialized

randomly, adaptation of the network was repeated 11 times and the classification of the testing subject was based on voting.

The Neural Network Toolbox for Matlab R2014b (The MathWorks, Inc.) was used here for all the described experiments.

D. Random Subspace Ensemble Principle

The datasets in neuropsychiatric research usually suffer from small sample size while their dimensionality is huge - often hundreds of thousands or even millions of voxels are used to describe each of subjects in the dataset. Hence, it is a hardly feasible task to avoid overfitting on a validation set by termination of neural network adaptation. Ensemble methods help to tackle overfitting and to improve the generalization ability.

In random subspace ensemble method, the variability is reached by a random selection of features from the set of all preselected features – a feature pool (FP). Each single classifier is adapted on one subset of feature which is less dimensional than the original feature space.

E. Validation and Evaluation

In order to report unbiased results, we used leave-one-out cross-validation strategy: one subject was left as a testing subject and the rest subjects composed a training set. This process was repeated *n* times, where *n* was the sample size. Since even only one testing subject withdrawn from the feature selection might have influenced the result, especially when the sample size was small, it was necessary to validate both features selection and classification.

For evaluation of the classification results, we used overall accuracy, sensitivity and specificity. The overall accuracy refers what proportion of subjects was classified correctly, whereas sensitivity and specificity say what the ability of diagnostic test is to reveal diseased and healthy person respectively.

III. Experiment and Results

The experiment was configured with many parameters – particularly size of the feature pool, number of chosen features and ensemble size. In order to achieve a reasonable computation time, we investigated only several predefined configurations. First, we defined a size of the feature pool as the 10000, 20000, 50000 and 100000 most significant voxels based on two-sample t-test criterion. This FP was used as the bag from which the features were chosen randomly.

The second parameter was a number of features used for the adaptation of the classifiers. We trained both MLP and SVM classifier on the different number of the most significant features - as shown in Fig. 1 - to find the optimum. The best performance was reached with the use of MLP on 1000 features and with the use of SVM on 100 features. Adding more features to the models did not reveal any trend in the classification performance - increasing or decreasing - so we later experimented with 3 options - 100, 1000 and 10000 - in order to explore both small and big dimensionality of the feature space. The MLP revealed higher accuracies compared to the SVM in most configurations.

The last investigated parameter was the number of voting classifiers in ensemble, which was set to 31. This odd number ensures that the subject is always assigned to one of the classes – SZ or HC.

Since outcomes from the whole ensemble were available, i.e. 31 units, we could use for evaluation any combination of 1, 3, 5 etc. classifiers from this ensemble. Hence, we computed performance measures on all, but maximally 10000 combinations in order to gain nonrandom and computationally accessible outcomes. Furthermore, the subsets of features were chosen randomly, all experiments were repeated 10 times and the measures of the classification performances were averaged.

Figures 2-5 show the experimental results. In each figure, the outcomes computed on the different size of the feature pool are displayed.

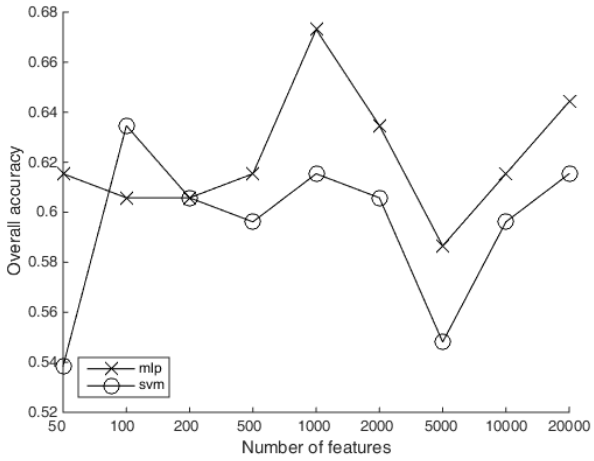


Fig. 1 The overall accuracy of the MLP and SVM dependent on the number of the most significant features selected with voxel-wise two-sample t-tests.

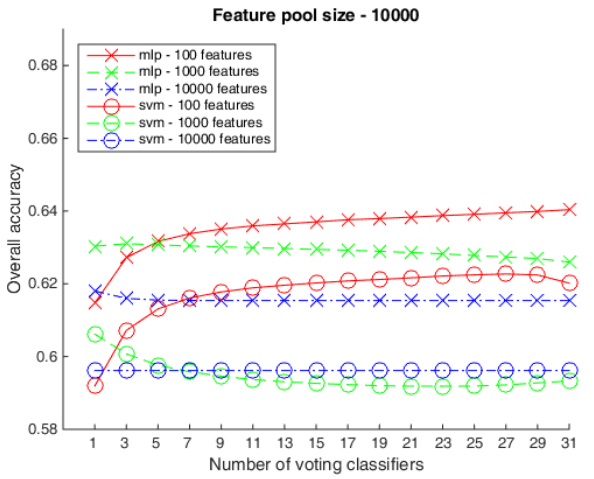


Fig. 2 Results of ensemble voting based on the feature pool with the size of 10000 selected features.

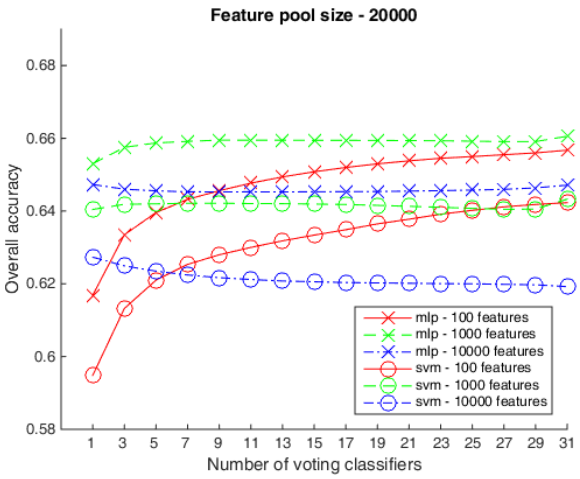


Fig. 3 Results of ensemble voting based on the feature pool with the size of 20000 selected features.

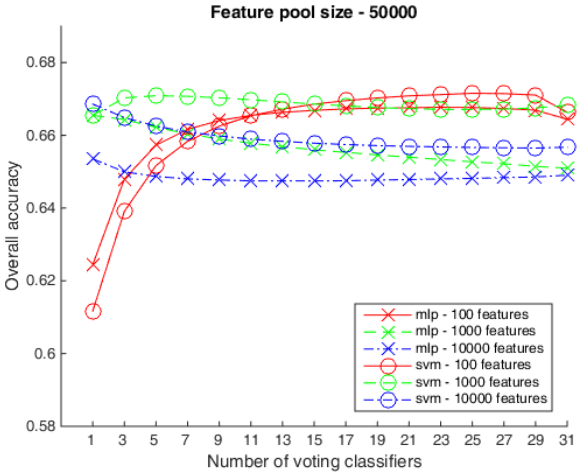


Fig. 4 Results of ensemble voting based on the feature pool with the size of 50000 selected features.

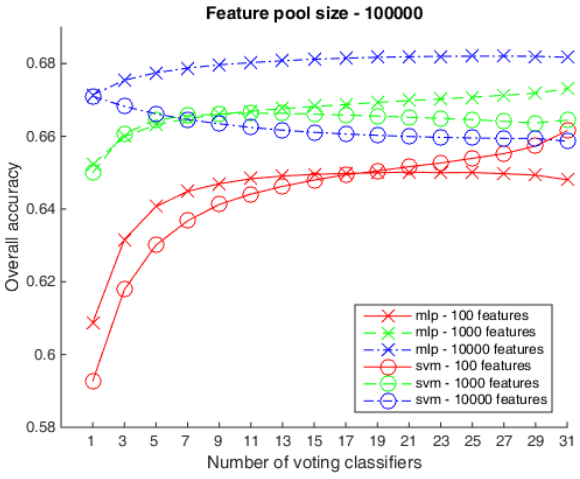


Fig. 5 Results of ensemble voting based on the feature pool with the size of 100000 selected features.

IV. Discussion and Conclusions

In this paper, we explored the random subset ensemble method for first-episode schizophrenia classification with the use of multi-layer perceptron and performed a comparison to ensembles of SVM classifiers.

We selected the features with a naïve approach based on voxel-wise two-sampled t-tests. We believe that this approach may be less prone to overfitting than other more complex methods. More sophisticated and often computationally more time-consuming methods such as sequential forward/backward feature selection or multivariate extraction methods are left for our future research.

Since the neighboring voxels could have been correlated due to the smoothing in the image preprocessing phase, we assumed that higher amount of features in the feature pool enabled to capture the discriminative information in more parts of the brain, and therefore improved the classification performance. This improvement is observable when the Fig. 2 and 4 are compared.

Multi-layer perceptron was more effective when smaller feature pool was used. With the use of bigger feature pools, both classification methods yielded more similar outcomes.

Increasing the number of voting classifiers surprisingly improved the classification accuracy only in case of small number of input features. The increasing trend in the accuracy reached a level similar to as the models with more inputs and the increasing trend did not continue. The SVM with 100 features adapted on FP with 100000 features was improved by 6.88% and MLP with the same number of input features adapted on FP of size 50000 achieved improvement of 4.33% when compared to single variants of the classification methods.

The best outcomes were revealed with the use of MLP on 100000 input features and 29 voting classifiers that have the highest overall accuracy of 68% (sensitivity 67%, specificity 69%).

Finally, we compare our results of schizophrenia prediction to other studies dealing with MLP or ensemble learning. Jafri and Calhoun [8] achieved 75.6% with MLP based on fMRI data (38 SZ + 31 HC). Savio et al. [7] reached 100% on diffusion imaging data containing 20 male subjects using both neural networks and SVM. Yang et al. [18] also reached much higher overall accuracy 87% with SVM classifier ensemble based on AdaBoost, but besides small sample size (20 SZ + 20 HC) they admit the patients were chronic and under an antipsychotic medication. These studies could suffer from small sample size, since models tend to reach more variable outcomes and therefore it is easier to reach good (as well as poor) classification performance [28]. Since Janousova et al. [19] exceeded 81% with a similar number of subjects, we propose that an involvement of other classifiers and various feature extraction methods may be helpful in our framework too.

We conclude that the random feature ensemble method in combination with MLP and SVM improved prediction of schizophrenia from MRI data only in case of short feature vectors (100 features) – when compared to the use of single MLP or SVM classifiers. The classification accuracy achieved with the ensembles was not much different from

the accuracy of the single classifiers with feature vectors of higher dimensionality (1000 and 10000 features).

REFERENCES

[1] J. Sun, J. J. Maller, L. Guo, and P. B. Fitzgerald, "Superior temporal gyrus volume change in schizophrenia: a review on region of interest volumetric studies," *Brain Res. Rev.*, vol. 61, no. 1, pp. 14–32, Jun. 2009. [Online]. Available: http://dx.doi.org/10.1016/j.brainresrev.2009.03.004

[2] D. C. Glahn, A. R. Laird, I. Ellison-Wright, S. M. Thelen, J. L. Robinson, J. L. Lancaster, E. Bullmore, and P. T. Fox, "Meta-Analysis of Gray Matter Anomalies in Schizophrenia: Application of Anatomic Likelihood Estimation and Network Analysis," *Biol. Psychiatry*, vol. 64, no. 9, pp. 774–781, Nov. 2008. [Online]. Available: http://dx.doi.org/10.1016/j.biopsych.2008.03.031

[3] C. Gaser, H.-P. Volz, S. Kiebel, S. Riehemann, and H. Sauer, "Detecting Structural Changes in Whole Brain Based on Nonlinear Deformations—Application to Schizophrenia Research," *NeuroImage*, vol. 10, no. 2, pp. 107–113, Aug. 1999. [Online]. Available: http://dx.doi.org/10.1006/nimg.1999.0458

[4] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 1," D. E. Rumelhart, J. L. McClelland, and C. PDP Research Group, Eds. Cambridge, MA, USA: MIT Press, 1986, pp. 318–362.

[5] J. Alirezaie, M. E. Jernigan, and C. Nahmias, "Neural network-based segmentation of magnetic resonance images of the brain," *IEEE Trans. Nucl. Sci.*, vol. 44, no. 2, pp. 194–198, Apr. 1997. [Online]. Available: http://dx.doi.org/10.1109/23.568805

[6] Y. Li, Z. Li, and Z. Xue, "Segmenting MR Images Using Fully-Tuned Radial Basis Functions (RBF)," in *9th International Conference on Control, Automation, Robotics and Vision, 2006. ICARCV '06*, 2006, pp. 1–6. [Online]. Available: http://dx.doi.org/10.1109/ICARCV.2006.345425

[7] C. J. Savio A, "Neural classifiers for schizophrenia diagnostic support on diffusion imaging data," *Neural Netw. World*, vol. 20, pp. 935–949, 2010.

[8] M. J. Jafri and V. D. Calhoun, "Functional classification of schizophrenia using feed forward neural networks," *Conf. Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Conf.*, vol. Suppl, pp. 6631–6634, 2006. [Online]. Available: http://dx.doi.org/10.1109/IEMBS.2006.260906

[9] C. Huang, B. Yan, H. Jiang, and D. Wang, "Combining Voxel-based Morphometry with Artifical Neural Network Theory in the Application Research of Diagnosing Alzheimer's Disease," in *International Conference on BioMedical Engineering and Informatics, 2008. BMEI 2008*, 2008, vol. 1, pp. 250–254. [Online]. Available: http://dx.doi.org/10.1109/BMEI.2008.245

[10] A. Savio, M. García-Sebastián, C. Hernández, M. Graña, and J. Villanúa, "Classification Results of Artificial Neural Networks for Alzheimer's Disease Detection," in *Intelligent Data Engineering and Automated Learning - IDEAL 2009*, E. Corchado and H. Yin, Eds. Springer Berlin Heidelberg, 2009, pp. 641–648. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-04394-9_78

[11] D. M. Joshi, N. K. Rana, and V. M. Misra, "Classification of Brain Cancer using Artificial Neural Network," 2010, pp. 112–116. [Online]. Available: http://dx.doi.org/10.1109/ICECTECH.2010.5479975

[12] M. R. Arbabshirani, K. Kiehl, G. Pearlson, and V. D. Calhoun, "Classification of schizophrenia patients based on resting-state functional network connectivity," *Brain Imaging Methods*, vol. 7, p. 133, 2013. [Online]. Available: http://dx.doi.org/10.3389/fnins.2013.00133

[13] T. K. Ho, "The random subspace method for constructing decision forests," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 8, pp. 832–844, Aug. 1998. [Online]. Available: http://dx.doi.org/10.1109/34.709601

[14] L. Breiman, "Random Forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001. [Online]. Available: http://dx.doi.org/10.1023/A:1010933404324

[15] L. Breiman, "Bagging Predictors," *Mach. Learn.*, vol. 24, no. 2, pp. 123–140, Aug. 1996. [Online]. Available: http://dx.doi.org/10.1023/A:1018054314350

[16] Y. Freund and R. E. Schapire, *A Short Introduction to Boosting*. 1999.

[17] J. J. Rodríguez, L. I. Kuncheva, and C. J. Alonso, "Rotation forest: A new classifier ensemble method," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 10, pp. 1619–1630, Oct. 2006. [Online]. Available: http://dx.doi.org/10.1109/TPAMI.2006.211

[18] H. Yang, J. Liu, J. Sui, G. Pearlson, and V. D. Calhoun, "A Hybrid Machine Learning Method for Fusing fMRI and Genetic Data: Combining both Improves Classification of Schizophrenia," *Front. Hum. Neurosci.*, vol. 4, Oct. 2010. [Online]. Available: http://dx.doi.org/10.3389/fnhum.2010.00192

[19] E. Janousova, D. Schwarz, and T. Kasparek, "Combining various types of classifiers and features extracted from magnetic resonance imaging data in schizophrenia recognition," *Psychiatry Res. Neuroimaging*, vol. 232, no. 3, pp. 237–249, Jun. 2015. [Online]. Available: http://dx.doi.org/10.1016/j.pscychresns.2015.03.004

[20] A. V. Lebedev, E. Westman, G. J. P. Van Westen, M. G. Kramberger, A. Lundervold, D. Aarsland, H. Soininen, I. Kłoszewska, P. Mecocci, M. Tsolaki, B. Vellas, S. Lovestone, and A. Simmons, "Random Forest ensembles for detection and prediction of Alzheimer's disease with a good between-cohort robustness," *NeuroImage Clin.*, vol. 6, pp. 115–125, 2014. [Online]. Available: http://dx.doi.org/10.1016/j.nicl.2014.08.023

[21] M. Liu, D. Zhang, and D. Shen, "Ensemble Sparse Classification of Alzheimer's Disease," *Neuroimage*, vol. 60, no. 2, pp. 1106–1116, Apr. 2012. [Online]. Available: http://dx.doi.org/10.1016/j.neuroimage.2012.01.055

[22] E. Janousova, D. Schwarz, G. Montana, and T. Kasparek, "Brain image classification based on automated morphometry and penalised linear discriminant analysis with resampling," in *2015 Federated Conference on Computer Science and Information Systems (FedCSIS)*, 2015, pp. 263–268. [Online]. Available: http://dx.doi.org/10.15439/2015F147

[23] J. Ashburner and K. J. Friston, "Unified segmentation," *NeuroImage*, vol. 26, no. 3, pp. 839–851, Jul. 2005. [Online]. Available: http://dx.doi.org/10.1016/j.neuroimage.2005.02.018

[24] M. Kubicki, M. E. Shenton, D. F. Salisbury, Y. Hirayasu, K. Kasai, R. Kikinis, F. A. Jolesz, and R. W. McCarley, "Voxel-Based Morphometric Analysis of Gray Matter in First Episode Schizophrenia," *NeuroImage*, vol. 17, no. 4, pp. 1711–1719, Dec. 2002.

[25] J. Ashburner and K. J. Friston, "Voxel-based morphometry--the methods," *NeuroImage*, vol. 11, no. 6 Pt 1, pp. 805–821, Jun. 2000. [Online]. Available: http://dx.doi.org/10.1006/nimg.2000.0582

[26] S. Lemm, B. Blankertz, T. Dickhaus, and K.-R. Müller, "Introduction to machine learning for brain imaging," *NeuroImage*, vol. 56, no. 2, pp. 387–399, May 2011. [Online]. Available: http://dx.doi.org/10.1016/j.neuroimage.2010.11.004

[27] R. O. Duda, *Pattern classification*, 2nd ed. New York: Wiley, 2001.

[28] H. G. Schnack and R. S. Kahn, "Detecting Neuroimaging Biomarkers for Psychiatric Disorders: Sample Size Matters," *Neuroimaging Stimul.*, p. 50, 2016. [Online]. Available: http://dx.doi.org/10.3389/fpsyt.2016.00050

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 323–329
DOI: 10.15439/2016F515
ACSIS, Vol. 8. ISSN 2300-5963

Supervised and Unsupervised Machine Learning for Improved Identification of Intrauterine Growth Restriction Types

Agnieszka Wosiak
Lodz University of Technology
Institute of Information Technology
ul. Wolczanska 215
90-924 Lodz, Poland
Email: agnieszka.wosiak@p.lodz.pl

Agata Zamecznik
Department of Pediatric
Cardiology and Rheumatology
2nd Chair of Pediatrics
Medical University of Lodz, Poland
Email: agazamek@gmail.com

Katarzyna Niewiadomska-Jarosik
Department of Pediatric
Cardiology and Rheumatology
2nd Chair of Pediatrics
Medical University of Lodz, Poland
Email: kasiajarosik@wp.pl

***Abstract*—This paper concerns automated identification of intrauterine growth restriction (IUGR) types by use of machine learning methods. The research presents a comparison of supervised and unsupervised learning covering single and hybrid classification, as well as clustering. Supervised learning techniques included bagging with Naïve Bayes, k-nearest neighbours (kNN), C4.5 and SMO as base classifiers, random forest as a variant of bagging with a decision tree as a base classifier, boosting with Naïve Bayes, SMO, kNN and C4.5 as base classifiers, and voting by all single classifiers using majority as a combination rule, as well as five single classification strategies: kNN, C4.5, Naïve Bayes, random tree and sequential minimal optimization algorithm for training support vector machines. Unsupervised learning encompassed k-means and expectation-maximization algorithms. The major conclusion drawn from the study was that hybrid classifiers have demonstrated their potential ability to identify more accurately symmetrical and asymmetrical types of IUGR, whereas the unsupervised learning techniques produced the worst results.**

I. Introduction

IN MEDICINE there are many diseases and diagnoses where identification of their subtypes affects medical treatment. Many research papers concern cancer diagnosis, appropriate feature selection techniques [1]–[3], and its classification based on gene expression [4]. A big challenge is an accurate classification of medical imaging and sound recordings (see [5] and [6]). Moreover, in many cases classification process is performed on labelled and unlabelled data [7]. It may take place in situations where a medical expert diagnosis is imprecisely outlined, as described in this paper.

Intrauterine growth restriction (IUGR) is a fetal growth disorder which is associated with fetal hypoxia and increased perinatal mortality. IUGR may cause a significant risk factor for the development of many cardiovascular, metabolic, and pulmonologic diseases in adult life ([8]–[10]). It is a challenging problem for obstetrician, neonatologists and pediatricians, as the diagnosis is based on non-consistent definitions (see [11] and [12]). It occurs in about 3-10% of live-born newborns, and the most serious problem of IUGR exists in developing countries where it concerns up to 20-30% of newborn infants [13]. The comparisons of absolute measurements of the fetuses with reference values, as well as birth weight percentiles, allow detection of deviations between expected and actual fetal growth and identification of newborns being possibly at risk for adverse health events [14].

Two types of IUGR can be distinguished: symmetrically impaired and asymmetrically impaired. Foetuses of the first type tend to have a decrease in all dimensions of the body and internal organs, and usually face a higher risk of reduction in growth potential. The problems occur in the first or second trimester of pregnancy and are often encountered in foetuses with infection or genetic and anatomic defects [15]. The second type - asymmetrical - constitutes 75-80% of all cases born as IUGR. It develops in the late second and third trimester of pregnancy and is a consequence of abnormal cell growth, rather than their quantity. In this type, infants have a low birth weight while body length and head circumference remain normal [16]. As asymmetric IUGR infants are more likely to have major anomalies than symmetric IUGR infants or non-IUGR infants [17], there is a need to distinguish between those two patterns of IUGR. Moreover symmetric and asymmetric growth restriction may have different influence on growth and development in preterms from birth to 4 years [18].

To discover risk factors and any parameters that impact IUGR, or to state the dependencies IUGR impacts on, it is necessary:

- to distinguish IUGR from normal fetuses,
- to identify the symmetrical or asymmetrical type of IUGR.

The problem of separating IUGR from normal fetuses has been the subject of analysis for researchers in the field of medicine, as well as computer science, including machine learning and artificial intelligence.

The authors of [19] used multiparametric classifier based on k-mean cluster analysis to separate pathological and normal fetuses. The identification of the intrauterine growth-restricted fetuses was performed on the basis of fetal heart rate variabil-

978-83-60810-90-3/$25.00

ity analysis in the antepartum period. The results attained up to 82.4% of accuracy.

In [20] an artificial neural network (ANN) classifier was developed to identify normal and abnormal fetuses based on features from ultrasound images. The accuracy of the classification equalled over 90%. Two ANN models, Multilayer Perceptron using Back propagation algorithm and Radial Basis Function, were also studied and used for IUGR identification in [12].

Lunghi et al. in [21] applied support vector machine algorithm for normal and pathological (IUGR) fetuses classification, based on the analysis of fetal heart rate recordings. The correct classification rate was high enough, above 84%. However, as a concluding remark for future work, the authors suggested using combined classifiers for better discrimination results.

In [22] statistical analysis (contingency tables, analyses of variance, and multiple regression) was applied to identify the problem of placental lesions associated with normal and abnormal fetal growth in infants delivered for obstetric indications at less than 32 weeks' gestation.

Although there are many studies that concern IUGR problem and its identification, few such studies explore different classification techniques. Therefore, automated or semi-automated identification of IUGR patterns is still an open topic.

The aim of this paper is to identify an appropriate classification technique as applied to the problem of intrauterine growth restriction types. Even though classification methods have been studied extensively over the past few years ([23]), no exact solution has been discovered. Moreover, the authors usually focus on one group of machine learning techniques: supervised or unsupervised, without comparisons between the groups. This research not only constitutes an independent contribution to the relevant literature , but also attempts to find a successful way to perform accurate classification of IUGR type.

The rest of the paper is organized as follows. Section II corresponds to the medical data used in this research and is followed by the description of methods used in the experimental part of the paper. Section III is dedicated to the experiments conducted on sample data and the results. Finally, in Section IV, the concluding remarks are discussed.

II. Materials and Methods

The proposed methodology of indicating the best machine learning method to use in IUGR types identification consists of three steps:

- applying supervised learning by single classification methods,
- performing multiple classification,
- carrying out clustering as an example of unsupervised learning,
- comparing results of classification techniques by methods of statistical analysis.

TABLE I: Characteristics of the groups

Parameter	IUGR-1 Avg ±SD (*)	IUGR-2 Avg ±SD (*)	p-value
Birth weight(g)	2556.91 ± 145.52	2516.77 ± 301.68	<0.001
Birth length(cm)	52.68 ± 1.51	50.06 ± 2.40	<0.001
Head circ.(cm)	33.37 ± 0.96	32.24 ± 1.13	<0.001

(*) described as average values ± standard deviations

A. Data Description

The research was based on a group of 68 children aged 5-10 years (average 7.4 ± 1.36) born on term with IUGR and birth weight below 10 percentile according to gestational age for the Polish population [24]. It consisted of 35 girls and 33 boys. All patients were selected during prospective studies at the Pediatric Cardiology and Rheumatology Department of Medical University of Lodz in 2010-2013. The study was approved by Medical Ethical Committee of the Health Sciences Faculty of Lodz University (No: RNN/760/10/KB).

Two subgroups were distingushed according to the type of hypotrophopy:

- IUGR-1 – asymmetrically impaired based on birth weight and an appropriate remainder of the parameters (body length and head circumference above 10 percentile),
- IUGR-2 – symmetrically impaired, where all parameters to be considered (birth weight, body length and head circumference) were below 10 percentile.

Both subgroups were equinumerous - consisted of 34 cases. The IUGR-1 group was constituted by 15 boys and 19 girls, whereas IUGR-2 included 18 boys and 16 girls.

The characteristics of all parameters subjected to further analysis differed significantly between IUGR-1 and IUGR-2 (see Table I).

B. Supervised Learning by Single Classification Method

Classification is the form of supervised learning, which means assigning objects into pre-defined sets of categories or classes. The main purpose of classification is to identify which set of categories a new observation belongs to. This is performed on the basis of a training set consisting of instances that are already labelled the known classes.

A classifier is a mapping function that can be defined by (1):

$$A^i \rightarrow C \tag{1}$$

where:

$a_1, \ldots, a_i \in A$ – are i features that characterize a set of n input instances $x_1, \ldots, x_n$

$y_j \in C = c_1, ..., c_m$ – are desired class labels.

C. Multiple Classification

Multiple classification combines individual classifiers in order to obtain a classifier that outperforms every single one.

There are two main questions that should be considered while performing multiple classification:

- the types of classifiers, that should be chosen,
- the way classifiers are combined to obtain a single classification result.

In the literature, there are two terms that refer to multiple classification: "ensemble methods" and "hybrid classifiers". The first one usually refers to collections of models that are minor variants of the same basic model, whereas hybridization allows combining classifiers from different families.

Regarding to combination rules for classifiers, in practice plurality voting is usually implemented (besides unanimity and simple majority) [25], [26]. It takes the result with the higher number of single classifiers' votes, which can be written as (2):

$$class(x) = \underset{c_i \in dom(y)}{arg\,max} \Big(\sum_k g(y_k(x), c_i) \Big) \tag{2}$$

where:

x – is an instance to be classified,
$dom(y) = \{c_1, c_2, ..., c_k\}$– constitutes the set of labels,
$y_k(x)$ – is the classification of the k^{th} classifier,
$g(y, c)$ – is an indicator function defined as:

$$g(y,c) = \begin{cases} 1 & y = c \\ 0 & y \neq c \end{cases}$$

Bagging and boosting are techniques that improve the accuracy of a classifier by generating a composite model that combines multiple classifiers derived from the same inducer.

The term bagging was introduced by Breiman in [27] as an acronym for Bootstrap AGGregatING. The idea of bagging is to create an ensemble classifiers based on bootstrap replicates of the training set. The classifier outputs are combined by the plurality vote [28].

A variant of bagging is a random forest [29]. It is a general class of ensemble building methods using a decision tree as the base classifier.

Boosting improves the performance of a weak learner as the method iteratively invokes a classifier on training data that is taken from various distributions. The classifiers are generated by resampling the training set and then combined into a single strong composite classifier. Boosting was based on an on-line learning algorithm called Hedge(β) [30]. This approach allocates weights to a set of strategies used to predict the outcome of a certain problem. The distribution is updated after each new outcome and strategies with the correct prediction receive higher weights while the impacts of the strategies with incorrect predictions are reduced.

One of the most popular ensemble algorithm that improves the simple boosting algorithm by an iterative process is AdaBoost (Adaptive Boosting). It was first introduced in [30]. The basic AdaBoost algorithm deals with binary classification. The classification of a new instance is performed according to (3):

$$class(x) = \underset{y \in dom(y)}{arg\,max} \Big(\sum_{t: M_t(x) = y} \log \frac{1}{\beta_t} \Big) \tag{3}$$

where:

x – is an instance to be classified,
$dom(y) = \{c_1, c_2, ..., c_k\}$– constitutes the set of labels,
M_t – is a base classifier,
β_t – is defined as: $\beta_t = \frac{\varepsilon_t}{1-\varepsilon_t}$,
ε_t – is defined as: $\varepsilon_t = \sum_{i: M_t(x_i) \neq y_i} D_t(i)$,
D_t – is a distribution defined as:
$D_1(i) = 1/m;\ i = 1, \ldots, m$
(m is a size of a training set)

$$D_{t+1}(i) = D_t(i) \cdot \begin{cases} \beta_t & M_t(x_i) = y_i \\ 1 & Otherwise \end{cases}$$

Bagging and boosting use votes to combine the outputs of different classifiers. However in boosting, each classifier is influenced by the performance of predecessors, which means that the new classifier pays more attention to classification errors that were done by the previously built classifiers. Besides in boosting, instances are chosen with a probability that is proportional to their weight, whereas in bagging, each instance is chosen with equal probability.

Hybrid classifiers [25], [26], [31], [32] (also named multiple classifier systems) are designed to increase the accuracy of a single classifier by training several different classifiers and combining their decisions to output a single class label. The hybridization exploits the strength of each component [33] and it prevents the need to try each classifier and simplifies the entire process [31].

For hybrid approach, the diversity is supposed to provide improved accuracy and classifier performance [34]. Therefore most works try to obtain maximum diversity by different means: introducing classifier heterogeneity, bootstrapping the training data, randomizing feature selection, randomizing subspace projections or boosting the data weights. Nevertheless, the diversity hypothesis has not been fully proven [34].

D. Unsupervised Learning with Clustering

Cluster analysis groups objects taking into account a certain similarity metric. The algorithms divide all objects into a predetermined number of groups in a manner that maximizes a similarity function. There are two different approaches, that are commonly used in medical studies ([35] and [36]): the Expectation Maximization (EM) probabilistic method and deterministic k-means algorithm.

An expectation-maximization (EM) algorithm performs repeatedly 2 steps: an expectation (E) and a maximization (M). The first step (E) results in an expectation of the likelihood for observed variables, whereas the second step - maximization (M) computes the maximum expected likelihood found during the E step. EM generates a probability distribution to each instance which indicates the likelihood of its belonging to each cluster [37]. The number of clusters can be designated by cross validation. It is worth emphasizing, that the EM algorithm computes classification probabilities, not exact assignments of observations to clusters.

The k-means algorithm divides a data set into k clusters, where k is a user-defined value. The algorithm starts with k random clusters, and next moves objects between those groups to minimize variability within each of them and maximize variability between clusters. Usually, the means for each cluster on every dimension are calculated to assign objects into the closest group [39]. In most of the cases Euclidean metric is considered as the distance function for k-means algorithm [37], [40].

E. Statistical Analysis

Statistical analysis is a required part of any research investigations, including proposing new methods or comparing existing ones in any field of science. Many researchers in machine learning confirmed the need for statistical validation of results.

In cases where comparison of two classifiers is performed, the McNemar test and 5x2 cross validation were recommended [41]. The situation where many classifiers are verified, is more complex from statistical point of view. Although many research papers draw conclusions based on matrix of tests comparing all pairs of classifiers (e.g. a matrix of the McNemar tests), an appropriate test for multiple comparisons should be used. The Friedman test with the corresponding post-hoc analysis was proved to be suitable for comparison of many classifiers [38], [45].

The Friedman test was firstly introduced in [42], [43] for non-parametric measures. The goal of the test is to determine - basing on samples - that there is a difference among classification results. The original results are changed into ranks starting from the best one and the null hypothesis states that all algorithms give same results and their ranks are identical. The Friedman statistics is computed as follows (4):

$$\chi_F^2 = \frac{12n}{k(k+1)} \left[\sum_j R_j^2 - \frac{k(k+1)^2}{4} \right] \tag{4}$$

where:

n – is the number of datasets being considered
k – is the number of algorithms
R_j – is the average rank of *j*th algorithm

If the null hypothesis is rejected, a post-hoc analysis should be performed to compare classifiers with each other and find statistically significant differences. The Nemenyi test can be applied. It states that two classifiers differ significantly if their ranks vary at least by the critical difference (5):

$$CD = q_\alpha \sqrt{\frac{k(k+1)}{6n}} \tag{5}$$

where:

q_α – are critical values based on the Studentized range statistic divided by $\sqrt{2}$.

TABLE II: Single classification results

Method	ACC [%]	PREC	SENS	AUROC
kNN	75.00	0.751	0.750	0.754
C4.5	76.47	0.765	0.765	0.697
Logistic	80.88	0.809	0.809	0.874
DTable	76.47	0.782	0.765	0.779
NaiveBayes	77.94	0.785	0.779	0.846
RandomTree	73.53	0.742	0.735	0.732
SGD	79.41	0.794	0.794	0.794
SMO	73.91	0.775	0.739	0.746

III. Results and Discussion

The purpose of experiments was to find the best method for IUGR type identification by examining the accuracy of different classification approaches, including single and hybrid classifiers, as well as clustering techniques.

The experiments were conducted on a real dataset consisted of 68 cases. Each case was described by 3 numerical attributes: birth weight, body length and head circumference according to the description presented in Section II-A.

The aim of the classification was to distinguish automatically symmetrical or asymmetrical type of IUGR. Therefore the set of labels consisted of two classes.

All experiments were based on WEKA Open Source Data Mining Tool [44].

In order to assess the performance of various classification methods, following comparison criteria have been used: accuracy (ACC), precision (PREC), sensitivity (SENS) and the area under ROC curve (AUROC). To verify experimental results, a detailed statistical evaluation was performed with use of Friedman test and post-hoc analysis.

A. Single Classification

In the first step of the experiments, single classification algorithms were applied. Eight approaches were considered: k-nearest neighbours (kNN), C4.5, logistic regression (Logistic), decision table, Naïve Bayes (NaiveBayes), random tree (a tree that considers K randomly chosen attributes at each node), stochastic gradient descent (SGD) and sequential minimal optimization algorithm for training support vector machines (SMO). The results of classification are presented in Table II.

The best single classification results attained 80% for logistic regression and 79% for SGD algorithm in terms of classification accuracy. Moreover, logistic regression and NaiveBayes gave the best results of AUROC (0.874 and 0.846 respectively) The average accuracy of single classification approach equalled 76.7%.

B. Multiple Classification

Next step of the experiments concerned performing classification using hybrid classifiers. Different combinations were applied:

- bagging with C4.5 and SMO as base classifiers,
- random forest as a variant of bagging with a decision tree (DTree) as a base classifier,

TABLE III: Hybrid classification results

Method	Base	ACC [%]	PREC	SENS	AUROC
Bagging	C4.5	79.41	0.799	0.794	0.852
Bagging	SMO	77.94	0.782	0.779	0.833
RandomForest	DTree	76.47	0.765	0.765	0.843
AdaBoost	DTable	80.88	0.824	0.809	0.835
AdaBoost	C4.5	80.88	0.809	0.809	0.826
AdaBoost	SGD	82.35	0.824	0.824	0.876
AdaBoost	SMO	77.94	0.780	0.799	0.806
Hybrid	all single	80.88	0.812	0.809	0.810

TABLE IV: Results of clustering

Method	No of cases in clusters	ACC [%]	PREC	SENS	AUROC
k-means	28 / 40	64.71	0.625	0.735	0.647
EM	63 / 5	57.35	1.000	0.147	0.574

- boosting with decision table (DTable), SMO, C4.5 and SGD as base classifiers, and
- hybridization by use of all single classifiers with majority voting as a combination rule.

The results of multiple classifications are shown in Table III. The best hybrid classification accuracy attained 82.35% for AdaBoost algorithm with SGD as a base classifier, whereas the worse one equalled 76.45% for RandomForest method. The average accuracy for all hybrid classification methods achieved 79.59%.

C. Clustering

The last step referred to clustering techniques. According to the methodology, two different approaches were considered: k-means algorithm and Expectation-Maximization method. Using EM algorithm we firstly used 10 fold cross-validation [37] to obtain clusters automatically, however it resulted in one cluster only. Therefore, for both techniques we defined 2 groups: for symmetrical and asymmetrical IUGR cases. The results of clustering are shown in Table IV.

One can notice that in the case of IUGR dataset, unsupervised techniques did not meet the expectations. Both algorithms resulted in accuracies below 65%, which is not satisfactory enough to implement this approach in practice.

D. Classification Comparison and Statistical Analysis

To compare the classifiers, the Friedman test and the corresponding post-hoc analysis were performed. The final results of absolute differences between average ranks for classifiers are presented in Table V where significant values are in bold, italic and underlined.

The results of the post-hoc tests can be clearly visualized with the diagram [38]. Figure 1 shows the results of the analysis of the data from Table V. The diagram compares all the algorithms against each other. The top line of the diagram is the axis on which we plot the average ranks of each method. Each number represents subsequent classification method sorted by the values of classification accuracy in the descending order, i.e. the lowest and best ranks are to the right. As a result we start with number 1 for boosted SGD and end with number 18 for EM clustering. The positions of average ranks for each classifier are marked with vertical lines and captioned with their names. Moreover, the groups of algorithms that are not significantly different in terms of accuracy are connected with horizontal lines. Consequently, we can easily notice, that there is no significant difference between boosted SGD and hybrid approach, however both of them achieved statistically better accuracies when compared with, inter alia, NaiveBayes, SMO or kmeans.

To summarize the experimental studies, one can see, that none of the classification techniques significantly outper-

TABLE V: Average ranks of post-hoc analysis

	kNN	C4.5	Log	DT	NB	RT	SGD	SMO	Bagg C4.5
kNN	0	2	***10.5***	2	5	2	7.5	1	7.5
C4.5	2	0	8.5	0	3	4	5.5	3	5.5
Logistic	***10.5***	8.5	0	8.5	5.5	***12.5***	3	***11.5***	3
DT	2	0	8.5	0	3	4	5.5	3	5.5
NB	5	3	5.5	3	0	7	2.5	6	2.5
RT	2	4	***12.5***	4	7	0	***9.5***	1	***9.5***
SGD	7.5	5.5	3	5.5	2.5	***9.5***	0	8.5	0
SMO	1	3	***11.5***	3	6	1	8.5	0	8.5
Bag. C4.5	7.5	5.5	3	5.5	2.5	***9.5***	0	8.5	0
Bag. SMO	5	3	5.5	3	0	7	2.5	6	2.5
RF	2	0	8.5	0	3	4	5.5	3	5.5
Boost DT	***10.5***	8.5	0	8.5	5.5	***12.5***	3	***11.5***	3
Boost C4.5	***10.5***	8.5	0	8.5	5.5	***12.5***	3	***11.5***	3
Boost SGD	***13***	***11***	2.5	***11***	8	***15***	5.5	***14***	5.5
Boost SMO	5	3	5.5	3	0	7	2.5	6	2.5
Hybrid	***10.5***	8.5	0	8.5	5.5	***12.5***	3	***11.5***	3
kmeans	3	5	***13.5***	5	8	1	***10.5***	2	***10.5***
EM	4	6	***14.5***	6	9	2	***11.5***	3	***11.5***

	Bagg SMO	RF	Boost DT	Boost C4.5	Boost SGD	Boost SMO	Hyb	km	EM
kNN	5	2	***10.5***	***10.5***	***13***	5	***10.5***	3	4
C4.5	3	0	8.5	8.5	***11***	3	8.5	5	6
Logistic	5.5	8.5	0	0	2.5	5.5	0	***13.5***	***14.5***
DT	3	0	8.5	8.5	***11***	3	8.5	5	6
NB	0	3	5.5	5.5	8	0	5.5	8	9
RT	7	4	***12.5***	***12.5***	***15***	7	***12.5***	1	2
SGD	2.5	5.5	3	3	5.5	2.5	3	***10.5***	***11.5***
SMO	6	3	***11.5***	***11.5***	***14***	6	***11.5***	2	3
Bag. C4.5	2.5	5.5	3	3	5.5	2.5	3	***10.5***	***11.5***
Bag. SMO	0	3	5.5	5.5	8	0	5.5	8	9
RF	3	0	8.5	8.5	***11***	3	8.5	5	6
Boost DT	5.5	8.5	0	0	2.5	5.5	0	***13.5***	***14.5***
Boost C4.5	5.5	8.5	0	0	2.5	5.5	0	***13.5***	***14.5***
Boost SGD	8	***11***	2.5	2.5	0	8	2.5	***16***	***17***
Boost SMO	0	3	5.5	5.5	8	0	5.5	8	9
Hybrid	5.5	8.5	0	0	2.5	5.5	0	***13.5***	***14.5***
kmeans	8	5	***13.5***	***13.5***	***16***	8	***13.5***	0	1
EM	9	6	***14.5***	***14.5***	***17***	9	***14.5***	1	0

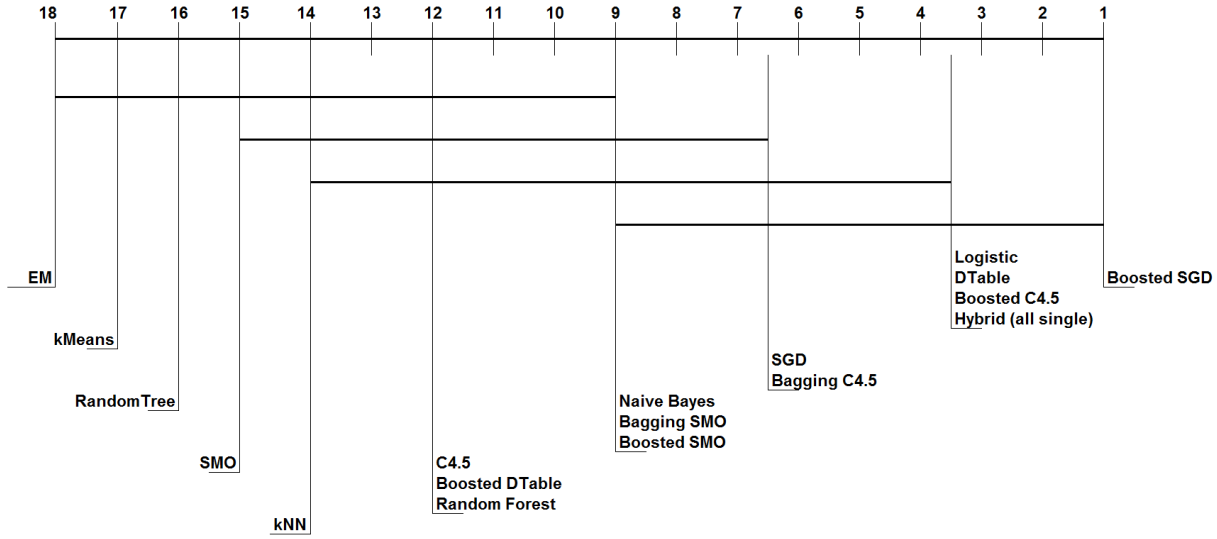


Fig. 1: Visualisation of post-hoc test for comparison of classifiers

formed the rest of them. However, it should be emphasized, that multiple classifications mostly exceeded single classifiers and grouping techniques in terms of classification accuracy. Even SGD - one of the best single classification method - when boosted, improved the accuracy to 82%.

IV. Conclusions

Classification of medical datasets is regarded as a challenging task, requiring extremely high accuracy. Therefore researches on finding the most appropriate methods for precise classification are conducted. Multiple classifiers constitute one of the most important advances in machine learning in recent years. In the absence of detailed a priori knowledge of the problem, they provide better performance.

The identification process of IUGR pattern (symmetrical or asymmetrical) is an important medical problem to solve, as symmetric and asymmetric growth restriction may have different influence on growth and development in childhood. Moreover asymmetric IUGR infants are more likely to have major anomalies than symmetric IUGR infants or infants appropriate for gestational age.

By comparing hybrid classifiers algorithms, single classification methods and clustering, it was demonstrated that the hybrid strategy resulted in the most satisfactory outcomes and confirmed other up-to-date researches on multiple classifier systems. Clustering, which is supposed to give good results in terms of unlabelled data and situations where label definitions are not precise, did not succeed in the case of IUGR classification.

In order to find the optimal solutions, future studies ought to involve other algorithms and strategies as well. Other combinations of various classifiers should be also investigated in depth. Furthermore, fuzzy logic can be applied to the problem of IUGR classification, as its results on medical data proved their efficiency [47]–[49].

References

[1] Paja W.: "Medical diagnosis support and accuracy improvement by application of total scoring from feature selection approach", Proceedings of the 2015 Federated Conference on Computer Science and Information Systems (FEDCSIS 2015), Annals of Computer Science and Information Systems, eds. M. Ganzha and L. Maciaszek and M. Paprzycki, IEEE, 2015, pp. 281–286, DOI: 10.15439/2015F361

[2] Gong, P. and Cheng, Y.-H. and Wang, X.-S.: "Benign or Malignant Classification of Lung Nodules Based on Semantic Attributes", Acta Electronica Sinica, 2015, vol. 43, no. 12, pp. 2476–2483

[3] Pérez, N. and Guevara, M.A. and Silva, A. and Ramos, I. and Loureiro, J.: "Improving the performance of machine learning classifiers for Breast Cancer diagnosis based on feature selection", Proceedings of the 2014 Federated Conference on Computer Science and Information Systems, Annals of Computer Science and Information Systems, eds. M. Ganzha, L. Maciaszek, M. Paprzycki, IEEE, 2014, vol. 2, pp. 209–217, DOI: 10.15439/2014F249

[4] Hijazi, H. and Chan, Ch.: "A Classification Framework Applied to Cancer Gene Expression Profiles" Journal of Healthcare Engineering, vol. 4, no. 2, pp. 255–283, 2013, DOI: 10.1260/2040-2295.4.2.255

[5] Sun, S., Wang, H., Jiang, Z., Fang, Y., Tao, T.: "Segmentation-based heart sound feature extraction combined with classifier models for a VSD diagnosis system", Expert Systems with Applications, 41(4), 2014, pp. 1769–1780, DOI: 10.1016/j.eswa.2013.08.076

[6] Montejo, L. D., Jia, J., Kim, H. K., Netz, U. J., Blaschke, S., Muller, G. A., Hielscher, A. H.: "Computer-aided diagnosis of rheumatoid arthritis with optical tomography, Part 2: image classification", Journal of biomedical optics, 2013, vol. 18(7), pp. 076002–076002, DOI: 10.1117/1.JBO.18.7.076002

[7] Stamatis, K. and Nikos, F. and Sotiris, K. and Kyriakos S.: "A Semisupervised Cascade Classification Algorithm", Applied Computational Intelligence and Soft Computing, Article ID 5919717, 14 pages, 2016, DOI: 10.1155/2016/5919717

[8] Baker, D.J.: "Maternal nutrition, fetal nutrition, and disease in later life", Nutrition, 1997, vol.13, pp. 807–813, DOI: 10.1016/S0899-9007(97)00193-7

[9] Mahajan, S.D. and Singh, S. and Shah, P. and Gupta, N. and Kochupillai, N.: "Effect of maternal malnutrition and anemia on the endocrine regulation of fetal growth", Endocrine research, 2004, vol. 30(2), pp. 189–203, DOI: 10.1081/ERC-200027380

[10] Mahajan, S.D. and Aalinkeel, R. and Singh, S. and Shah, P. and Gupta, N. and Kochupillai, N.: "Endocrine regulation in asymmetric intrauterine fetal growth retardation", Journal of Maternal-Fetal and Neonatal Medicine, 2006, vol. 19(10), pp. 615–623, DOI: 10.1080/14767050600799901

[11] Gadagkar, A.V. and Shreedhara, K.S.: "Fetal Growth Diagnosis using Re-Initialization Free Level Set Method and Classification using Radial Basis Function Neural Network", Proceedings of the International Conference on Multimedia Processing, Communication and Information Technology MPCIT 2013, 2013, pp. 137–144, DOI: 03.AETS.2013.4.81

[12] Bagi, K.S. and Shreedhara, K.S.: "Biometric measurement and classification of IUGR using neural networks", Proceedings of the International Conference on Contemporary Computing and Informatics (IC3I 2014), 2014, pp. 157–161, DOI: 10.1109/IC3I.2014.7019613

[13] Black, R.E. and Victora, C.G. and Walker, S.P. and Bhutta, Z.A. and Christian, P. and de Onis, M. and et al.: "Maternal and child undernutrition and overweight in low-income and middle-income countries", Lancet, 2013, vol. 382, pp. 427–451, DOI: 10.1016/S0140-6736(13)60937-X

[14] Gürgen, F. and Zeynep, Z. and Füsun, V.: "Intrauterine growth restriction (IUGR) risk decision based on support vector machines", Expert Systems with Applications, 2012, vol.39(3), pp. 2872–2876, DOI: 10.1016/j.eswa.2011.08.147

[15] Shreedhara, K.S. and Veena, A.: "Multiple sonographic features based IUGR diagnosis using artificial neural networks", International Journal of Information Technology and Knowledge Management, 2009, vol.2(1), pp. 73–78, DOI:10.1109/ICSIP.2014.54

[16] Zamecznik, A. and Niewiadomska-Jarosik, K. and Wosiak, A. and Zamojska, J. and Moll, J. and Stańczyk, J.: "Intra-uterine growth restriction as a risk factor for hypertension in children six to 10 years old", Cardiovascular Journal of Africa, 2014, pp.73–77, DOI: 10.5830/CVJA-2014-009

[17] Dashe, J.S. and McIntire, D.D. and Lucas, M.J. and Leveno, K.J.: "Effects of symmetric and asymmetric fetal growth on pregnancy outcomes", Obstetrics & Gynecology, 2000, vol. 96(3), pp. 321–327

[18] Bocca-Tjeertes, I. and Bos, A. and Kerstjens, J. and de Winter, A. and Reijneveld, S.: "Symmetrical and Asymmetrical Growth Restriction in Preterm-Born Children", Pediatrics, 2013, vol. 133(3), pp. e650–e656

[19] Ferrario, M. and Signorini, M.G. and Magenes, G.: "Complexity analysis of the fetal heart rate variability: early identification of severe intrauterine growth-restricted fetuses", Medical & Biological Engineering & Computing, 2009, vol.47(9), pp. 911–919, DOI: 10.1007/s11517-009-0502-8

[20] Gadagkar, A.V. and Shreedhara, K.S.: "Features Based IUGR Diagnosis Using Variational Level Set Method and Classification Using Artificial Neural Networks", Proceedings of the Fifth International Conference on Signal and Image Processing (ICSIP 2014), 2014, pp. 303–309, DOI: 10.1109/ICSIP.2014.54

[21] Lunghi, F. and Magenes, G. and Pedrinazzi, L. and Signorini, M.G.: "Detection of fetal distress though a support vector machine based on fetal heart rate parameters", Computers in Cardiology, 2005, vol. 32, pp. 247–250, DOI: 10.1109/CIC.2005.1588083

[22] Salafia, C.M. and Minior, V.K. and Pezzullo, J.C. and Popek, E.J. and Rosenkrantz, T.S. and Vintzileos, A.M.: "Intrauterine growth restriction in infants of less than thirty-two weeks' gestation: associated placental pathologic features", American Journal of Obstetrics and Gynecology, 1995, vol. 173(4), pp. 1049–1057, DOI: 10.1016/0002-9378(95)91325-4

[23] Jeetha Lakshmi, P. S. and Saravan Kumar S. and Suresh A.: "A Novel Hybrid Medical Diagnosis System Based on Genetic Data Adaptation Decision Tree and Clustering", ARPN Journal of Engineering and Applied Sciences, vol. 10, no. 16, 2015, pp. 7293–7299

[24] Malinowski, A. and Chlebna-Sokół, D.: "Dziecko łódzkie-metody badań i normy rozwoju biologicznego", Ankal, 1998, (In Polish)

[25] Kuncheva, L.I.: Combining Pattern Classifiers. Methods and Algorithms., John Wiley & Sons, Inc., 2004, Hoboken, New Jersey, USA

[26] Woźniak, M. and Graña, M. and Corchado, E.: "A survey of multiple classifier systems as hybrid systems", Information Fusion, 2014, pp. 3–17, DOI: 10.1016/j.inffus.2013.04.006

[27] Breiman, L.: "Bagging predictors", Technical Report 421, Department of Statistics, University of California, Berkeley, 1994

[28] Breiman, L.: "Bagging predictors", Machine Learning, 1996, vol. 26(2), pp. 123–140

[29] Breiman, L.: "Random forests", Machine Learning, 2001vol. 45, pp. 5–32

[30] Freund, Y. and Schapire, R.E.: "A decision–theoretic generalization of on-line learning and an application to boosting", Journal of Computer and System Sciences, 1997, vol. 55(1), pp. 119–139, DOI: 10.1006/jcss.1997.1504

[31] Rokach, L.: "Pattern Classification Using Ensemble Methods", World Scientific Publishing Co., Inc., 2010, River Edge, New Jork, USA

[32] Seni, G. and Elder, J.F.: "Ensemble Methods in Data Mining: Improving Accuracy Through Combining Predictions", Morgan & Claypool, 2010

[33] Michalski, R.S. and Tecuci, G.: "Machine Learning, A Multistrategy Approach", J. Morgan Kaufmann, 1994

[34] Wang, S.L. and Li, X.L. and Fang, J.: "Finding minimum gene subsets with heuristic breadth-first search algorithm for robust tumour classification", BMC Bioinformatics, 2012, vol. 13(178, pp. 1–26, DOI: 10.1186/1471-2105-13-178

[35] Pirooznia, M. and Yang, J. and Yang M.Q. and Deng, Y.: "A comparative study of different machine learning methods on microarray gene expression data", BMC Genomics, 2008 vol. 9, pp. 1471–2164, DOI: 10.1186/1471-2164-9-s1-s13

[36] Wosiak, A. and Zakrzewska, D.: "Feature Selection for Classification Incorporating Less Meaningful Attributes in Medical Diagnostics", Proceedings of the 2014 Federated Conference on Computer Science and Information Systems, Annals of Computer Science and Information Systems ACSIS, 2014, pp. 235–240, DOI: 10.15439/2014F296

[37] Witten, I.H. and Frank, E. and Hall, M.A.: "Data Mining: Practical Machine Learning Tools and Techniques" (3rd ed.), Morgan Kaufmann Publishers Inc., 2011

[38] Demsar, J.: "Statistical Comparisons of Classifiers over Multiple Data Sets", The Journal of Machine Learning Research, 2006, vol. 7, pp. 1–30

[39] MacQueen, J.B.: "Some Methods for Classification and Analysis of MultiVariate Observations", Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, 1967, pp. 281–297

[40] Ankita, V. and Satyanarayana, R.V. and Kamalakar, K.: "An Experiment with Distance Measures for Clustering", Proceedings of the International Conference on Management of Data, 2008

[41] Dietterich, T. G.: "Approximate statistical tests for comparing supervised classification learning algorithms", Neural Computation, 2006, vol. 10(7), pp. 1895–1923, DOI: 10.1162/089976698300017197

[42] Friedman, M.: "The use of ranks to avoid the assumption of normality implicit in the analysis of variance", Journal of the American Statistical Association, 1937, vol. 32, pp. 675–701, DOI: 10.1080/01621459.1937.10503522

[43] Friedman, M.: "A comparison of alternative tests of significance for the problem of m rankings", Annals of Mathematical Statistics, 1940, vol. 11, pp. 86–92

[44] Weka Data Mining Tool: http://www.cs.waikato.ac.nz/ml/weka/index.html

[45] Garcia, S. and Fernandez, A. and Luengo, J. and Herrera, F.: "Advanced nonparametric tests for multiple comparisons in the design of experiments in computational intelligence and data mining: Experimental analysis of power", Information Sciences, 2010, vol. 180(10), pp. 2044–2064, DOI: 10.1016/j.ins.2009.12.010

[46] Nemenyi, P.B.: "Distribution-free multiple comparisons", PhD thesis, Princeton University, 1963

[47] Widjaja, M. and Darmawan, A. and Mulyono, S.: "Fuzzy classifier of paddy growth stages based on synthetic MODIS data", Proceedings of the IEEE International Conference on Advanced Computer Science and Information Systems (ICACSIS 2012), 2012, pp. 239–244

[48] Grochowina, M. and Leniowska, L.: "Comparison of SVM and k-NN classifiers in the estimation of the state of the arteriovenous fistula problem", Proceedings of the 2015 Federated Conference on Computer Science and Information Systems (FEDCSIS 2015), Annals of Computer Science and Information Systems, eds. M. Ganzha and L. Maciaszek and M. Paprzycki, IEEE, 2015, pp. 249–254, DOI: 10.15439/2015F194

[49] Lakshmi, PS Jeetha, S. Saravan Kumar, and A. Suresh. "Intelligent Medical Diagnosis System Using Weighted Genetic and New Weighted Fuzzy C-Means Clustering Algorithm." Artificial Intelligence and Evolutionary Algorithms in Engineering Systems. Springer India, 2015. pp. 213-220

[50] Nawarycz, T. and Pytel, K. and Gazicki-Lipman, M. and Drygas, W. and Ostrowska-Nawarycz, L.: "A Fuzzy Logic Approach to The Evaluation of Health Risks Associated with Obesity", Proceedings of the 2013 Federated Conference on Computer Science and Information Systems, eds. M. Ganzha, L. Maciaszek, M. Paprzycki, IEEE, 2013, pp. 231–234

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 331–340
DOI: 10.15439/2016F150
ACSIS, Vol. 8. ISSN 2300-5963

Reliability Estimation of Healthcare Systems using Fuzzy Decision Trees

Vitaly Levashenko, Elena Zaitseva,
Miroslav Kvassay,
University of Zilina,
Department of Informatics
Zilina, Slovakia
Email: {vitaly.levashenko, elena.zaitseva,
miroslav.kvassay}@fri.uniza.sk

Thomas M. Deserno
Uniklinik RWTH Aachen
Department of Medical Informatics
52057 Aachen, Germany
Email: deserno@ieee.org

Abstract—**Reliability is an important characteristic of any system. Healthcare systems are typical examples of such systems. In reliability engineering, such systems are considered as complex, inhomogeneous, and uncertain, and require special mathematical representations. The structure function is a suitable model representing real systems. Methods of system reliability evaluation based on the structure function are well established but deterministic. This restricts its use for uncertain or incomplete data. A structure function can be created only for system in which correlations of all components are indicated and all component states are known. In this paper, a new method for structure function construction is proposed. Incomplete data is analysed using Fuzzy Decision Trees (FDTs), where input and output attributes are interpreted as component states and values of the structure function, respectively. This method is applied to reliability analysis of healthcare system. For illustration, we considered the system laparoscopic surgery that has 4 components and 36 state vectors. In addition we evaluate proposed method by 3 benchmark's systems with 243, 108, and 512 state vectors, respectively. Two of these benchmarks have 5 components and one has 4 components. Uncertainty is simulated by randomly deleting between 5% and 90% of all state vectors before constructing the structure function. With 50% of deleted stages, the error rate is below 0.2% for all three systems. We conclude that FDT-based reliability analysis is applicable for incomplete data in medical systems, too.**

I. Introduction

THE investigation in reliability engineering of healthcare system has started as analysis of medical devices and equipment [1]. Until recently, reliability quantification of equipment and devices has been a principal tendency in medicine [2, 3, 4]. Independently, the human factor in medicine has been investigated.

A special area in reliability engineering investigating the influence of human factor is named *Human Reliability Analysis* (HRA). HRA aims at identifying the potential failure of the system resulting from human errors, analyzing causes and identifying appropriate countermeasures to prevent and reduce the linked risks as much as possible. Failures in healthcare are called Medical Errors, if the patient's condition worsens or if patients develop additional illnesses. As documented in [5], the number of deaths due to these causes is 225,000 annually, with in-hospital medical errors causing 44,000 to 98,000 deaths, and 3,000,000 injuries annually. Of course, such situation must be changed.

A first step of improvement is to investigate causes and specifics of medical error. A human error in a healthcare system is considered as independent problem of reliability analysis [5, 6, 7, 8]. Authors of papers [7, 8, 9] have indicated methods that are most useful for medical error analysis. In these papers and other investigations, authors provide the adaptation of well-known and popular HRA-methods for the medical domain. The background of HRA methods is discussed in [9].

A healthcare system includes components of different types, such as technical components (equipment/devises) and the human factor. Therefore, this system is inhomogenous in the view of reliability engineering [3, 4, 7], and the construction of mathematical representation including all types of component become challenging [3, 7, 10]. The simplest decision is obtained for system of stationary states, where the time dependence of system behavior can be ignored. One of possible representation for this condition is a structure function. This is a deterministic model and all possible component states and performance levels must be indicated and reflected in the structure function. However, complete information of healthcare systems cannot be obtained, since the observation of all situations is impossible: some of them agree with hazard of patient's health. For example, author in [3] consider reliability and safety of pacemaker application that is interpreted as system with uncertainty. Furthermore, parts of the information are deducted from expert's experiences. This information is ambiguous and unequal, It can be considered with just some reliability or confidence. Therefore, the structure function of healthcare systems must be constructed based on uncertain data.

In this paper, a novel method for constructing the structure function is proposed. We take into account the uncertainties of initial data and suggest a method that is based on the *Fuzzy Decision Tree* (FDT). FDTs are widely used in data mining for analysis of uncertain data and

This work was supported by the grant of 7th RTD Framework Program No 610425 (RASimAs)

decision making in ambiguities [11, 12]. FDTs can be used naturally to analyze fuzzy data. Therefore, the initial data used to construct the structure function can be fuzzy. To transform a real system's performance level and its component states into an exact model, such as the structure function, fuzziness of bounds of performance levels/states are considered. In addition, FDTs allow to take into account uncertainties caused by incompletely specified data. This is still possible if it is expensive to obtain all data about the real system's behavior, if there is only sparse data, or if data s acquired incompletely due to poor documentation.

As a rule, if the exact value of the actual data about the system behavior cannot be determined, we need to rely on more data to give additional information necessary to correct the theoretical model used [13]. An FDT allows reconstructing these data with different levels of the confidence [14]. The use of FDTs for the construction of the structure function assumes the induction of a tree that is based on the data (fuzzy and/or crisp). The values of the structure function are then defined by the FDT for all combinations of the component states.

This paper is structured as follows. Section II discusses the concept of structure function. Principal steps of the proposed method are considered in Section III. The detail description of all steps is given in Section IV. The results of method evaluation are presented in Section V, and the discussion and conclusion are given in Section VI.

II. The Structure Function in Healthcare Analytics

A. Structure Function

The concept of structure function is introduced in reliability engineering in order to mathematically describe the real system that is studied. In this case, the system is represented as a mapping that assigns a system state to every possible profile of component states. Therefore, the system performance level is defined from the states of all its components, and all possible component states as well as all performance levels must be indicated and reflected in the structure function.

The structure function allows to represent the system's reliability behavior for two typical mathematical models, i.e. the *Binary-State System* (BSS) and the *Multi-State System* (MSS). BSS permits only two states to investigate the system and its components: perfect functioning and complete failure. However in practice, many systems can exhibit different performance levels between these two extremes of full function and fatal failure [15, 16]. MSS is a mathematical model that is used to describe a system with several (more than two) levels of performance [15, 17, 18].

The concept of the structure function is used to represent BSS and MSS, and associates the space of component states and system performance levels:

$$\phi(x_1,\ldots,x_n)=\phi(\boldsymbol{x}):\\ \{0,\ldots,m_1-1\}\times\ldots\times\{0,\ldots,m_n-1\}\to\{0,\ldots,M-1\}, \quad (1)$$

where $\phi(\boldsymbol{x})$ is the system state (performance level) from failure ($\phi(\boldsymbol{x})=0$) to perfect functioning ($\phi(\boldsymbol{x})=M$ -1); $\boldsymbol{x}=(x_1,\ldots,x_n)$ is a vector of system component's states (state vector). The state $\phi(\boldsymbol{x})=0$ represents the total failure of the component while state $\phi(\boldsymbol{x})=m_i$ -1 corresponds to perfect functioning of the i-th component.

Therefore, (1) defines the MSS structure function of n components (subsystems). The state of each component can be denoted by a random variable, x_i, and every component of the MSS is characterized by probabilities of its states:

$$p_{i,s}=\Pr\{x_i=s\},\quad s\in\{0,\ldots,m_i-1\}. \quad (2)$$

As it is shown in [10], reliability analysis of healthcare systems can be provided if the system is represented by a structure function (1). Then, the typical measures of reliability, e.g., the system availability (probabilities of the system performance levels), can be calculated by the structure function [18, 10]:

$$A_j=\Pr\{\phi(\boldsymbol{x})=j\},\ j=0,\ldots,M\text{-}1. \quad (3)$$

For example, consider a simple system to indicate medical errors of two components: a doctor (x_1) and diagnostic device (x_2). Suppose three levels in the doctor work: 0 is interpreted as fatal error, 1 is incorrect work (without fatal result) and 2 is perfect work. The devices can have two states only: 0 is failure and 1 is proper function. The system quantification has three levels: 0 is incorrect diagnosis with fatal consequence for the patient, 1 is incorrect diagnosis without fatal consequence for the patient, and 2 is correct diagnostics. This system (the medical error identification) is represented by the structure function in Table I. The probability of every performance level (availability) of this system is calculated according to (3):

$$A_0=\Pr\{\phi(\boldsymbol{x})=0\}=p_{1,0}\cdot p_{2,0},$$
$$A_1=\Pr\{\phi(\boldsymbol{x})=1\}=p_{1,0}\cdot p_{2,1}+p_{1,1}\cdot p_{2,0},$$
$$A_2=\Pr\{\phi(\boldsymbol{x})=2\}=p_{1,1}\cdot p_{2,1}+p_{1,2}\cdot p_{2,0}+p_{1,2}\cdot p_{2,1}$$

TABLE I.
THE STRUCTURE FUNCTION OF THE MEDICAL ERROR IDENTIFICATION

The system components		$\phi(\boldsymbol{x})$
x_1	x_2	
0	0	0
0	1	1
1	0	1
1	1	2
2	0	2
2	1	2

There are a lot of methods to estimate different aspects of the system's reliability based on the structure function [3, 10, 15, 17, 18]. Therefore, the structure function allows for investigating the system's reliability. In some applications – such as the medical domain – the construction of the structure function is a complex problem, because the structure function (1) usually is assumed to be exact and ambiguities are not taken into account. Since data about real

systems is uncertain, the structure function may not be realistic, which is – as a rule – typical for healthcare systems [7].

B. The Structure Function for Healthcare Systems

From the point of view of reliability analysis, estimating preventable medical errors in healthcare system is the principal goal.

During the past decade, healthcare delivery has seen the introduction of ever more sophisticated and complex equipment for preventable medical errors. Furthermore, the human factor still is an important component of such systems. The persistence of medical errors according to [7] suggests that there is either an absence of reliability engineering analysis or a gap in the reliability analysis currently being performed. The decision of this problem can be implemented by changing the process design of healthcare system and/or performing the reliability analysis of healthcare system's exploitation. It supposes the adaptation and application of reliability analysis methods for healthcare system.

The first step in reliability estimation of any system is to determine a mathematical model (representation). Such mathematical model of a healthcare system is:

- complex, the system cannot be represented by a typical structure (series, parallel, *k*-out-of-*n* etc.) only;
- inhomogeneous, it implies different types of the system components (hardware, software, human factor);
- uncertain, it is caused by human factor influence.

In case of the structure function, complexity and inhomogeneity of healthcare systems can be realized in a mathematical representation without special algorithms [15, 19]. In the construction of the structure function for healthcare system, the uncertainty of the initial data needs adaptation and the reasons of this uncertainty must be considered. According to (1), the structure function is defined if all of possible components and system states are indicated. As a rule, it is impractical to wait until all the component states are indicated. The first factor then is incomplete specification of data, because some values of the system component's states or performance levels cannot be obtained. In substitution, extra data on the healthcare system can be obtained through expert analysis. The second factor of uncertainty of initial data is the ambiguity and vagueness of collected data values. This type of ambiguity can be caused by an expert's subjective evaluation etc. For example, two experts can set different values of system performance level for equal situation [20, 21]. Therefore, uncertainty of initial data must be considered when constructing the structure function.

Uncertainties and ambiguities of a real system have been dealt with in reliability analysis using of the likelihood concept [13, 19]. However, uncertainties that are caused solely in evaluation by an expert's experience and judgement are not random in nature. This uncertainty can't be indicated in a quantitative form by probability theory.

Fuzzy logic, however, makes it possible to define the structure function in a more flexible form than the probabilistic approach. Consequently the structure function of a healthcare system must be constructed based on fuzzy data and incomplete samples. It is a typical problem in data mining. Therefore, we propose the use of the Fuzzy Decision Tree (FDT), which has been used widely in data mining, for analysis of uncertain data, and decision making with ambiguities. Here, collected data for structure function construction can be defined and characterized by likelihood (likelihood) or confidence if there is little data about the real system's behavior. An FDT allows to reconstruct these data with different levels of likelihood (confidence) [12, 14]. Applying FDT, we propose a new method for the construction of the structure function to mathematically represent healthcare system.

FDTs imply a tree based on the data about the system, which can be fuzzy and/or crisp, and the data can be specified incompletely. The values of the structure function are then defined for all combinations of component states by the FDT: component states are interpreted as FDT attributes and the structure function value agrees with one of the M values (classes) for the system performance level.

III. A Novel Method for Structure Function Construction by FDT

The structure function is a construct of mathematical representations that can be defined by a system structure analysis or that can be based on expert data [13, 22]. The structure function for system that includes a human factor is funded based on the evaluation by an expert's experience and judgment. Any healthcare system is a typical example of such a system.

The construction of a structure function representing expert's data requires special analysis and transformation of the initial system's data [7, 23], because expert's data is uncertain as a rule. This uncertainty can be caused by a lot of factors, but we have considered two of them. The first factor is incompletely specified data, because some values of system states or performance levels cannot be obtained. For example, it can be expensive or it might need unacceptable long time to get the data. The second factor is ambiguity and vagueness of collected data values. This type of ambiguity can be caused, for instance, by inaccuracies or errors of measurement, or subjectivity of expert's evaluations.

Therefore, the construction of the structure function must appreciate two aspects. The first is a mapping assigning the system performance level to each possible profile of component states (for example, see Table I). The second is addressed by interpreting it as classification problem for uncertain data, which is typical for data mining. One of the possible options is applying decision trees or FDT [31-34].

Our approach is based on FDTs [24, 25] and includes the following steps (Fig.1):

- Collection of data in the repository according to requests of FDT induction;

• Representation of the system model in the form of an FDT that classifies component states according to the system performance levels;

• Construction of the structure function as decision table that is created by the inducted FDT.

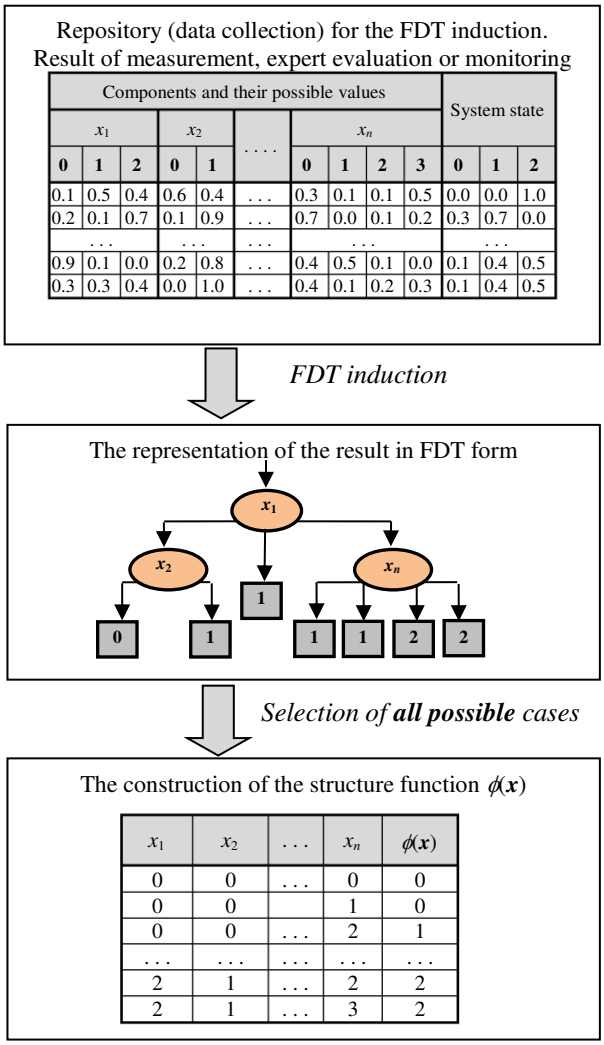


Fig. 1 Principal steps in the structure function construction based on FDT application

Therefore, the structure function is constructed as decision table that classifies the system performance level to each possible profile of components states. The decision table is formed based on a FDT that provides the mapping for all possible component states (input data) in M performance levels. FDT is inducted by uncertain data that is presented in form of specified repository.

IV. Principal Steps of the Method

A. The Repository for Data Collection

Collection of data in the form of a repository is provided by the monitoring or expert evaluation of values of the system's component states and the system's performance levels (Fig.1). Data for a repository is collected by monitoring if it is possible (for example, for some devises or equipment). It is expert evaluation dominantly in case of the system with analysis of human factor that is important component of any healthcare system. Therefore data for the construction of the repository for the analysis of healthcare system collects the expert experience. The experts propose an estimation of every situation (state vector) and resulting system state (system performance level) with some possibilities (likelihoods). The transformation of expert knowledge into quantification data is implemented based general rules that used for the analysis of experts' knowledge. Some of such methods are presented and discussed in papers [20, 21]. Need to note that initial data structuration and the repository preparation is important step and can be considered as separate problem for investigation and development.

The column number is n+1 (for indication of n components and the system's performance level). All of the n columns are separated into m_i sub-columns, and the column for the system's performance level has M sub-columns. The sub-column is assigned with one of the values for component states or performance levels. Every row of the table represents one monitoring situation or evaluation. The table cell includes number (from 0 to 1) that interpreted as the likelihood of this value. Note that the sum of these possibilities for each value equals 1. Such data can be obtained from experts' evaluations or possibility of fuzzy clustering [26, 27]. These possibilities correspond to a membership function of fuzzy data [28]. This demand for initial data representation is caused by the method of FDT induction. Therefore, values of the i-th component state and the performance levels are defined by possibilities. These possibilities indicate ambiguity of collected data values for the analysis. Having indicated and considered the uncertainty of the monitoring data, the resulting accuracy of data analysis is increased.

For example, consider a simple laparoscopic surgery procedure [7]. It is typical healthcare system with human component. Let us indicate 4 components (n = 4) of this system: device (a laparoscopic robotic surgery machine [29]), two doctors (anesthesiologist and surgeon) and a nurse. Let us interpret this system as *Multi-State System* MSS and introduce the numbers of states for every component and number of performance levels of the system. Let this system has three performance levels (M = 3):

0 – non-operational (fatal medical error),

1 – partially operational (some imperfection),

2 – fully operational (surgery without any complication).

The device has two states (m_1 = 2):

0 – failure, and

1 –functioning.

The work of anesthesiologist is indicated by two states (m_2 = 2):

0 – non-operational (medical error),

1 – fully operational (without any complication).

TABLE II.
THE COLLECTED DATA FOR ANALYSIS OF LAPAROSCOPIC SURGERY SUCCESSFUL

No	x_1		x_2		x_3			x_4			$\phi(x)$		
	0	**1**	**0**	**1**	**0**	**1**	**2**	**0**	**1**	**2**	**0**	**1**	**2**
1	0.6	0.4	0.9	0.1	0.1	0.9	0.0	0.2	0.6	0.2	0.9	0.1	0.0
2	0.7	0.3	1.0	0.0	0.0	0.9	0.1	0.1	0.8	0.1	0.8	0.1	0.1
3	0.5	0.5	0.9	0.1	0.8	0.2	0.0	0.8	0.1	0.1	0.9	0.1	0.0
4	1.0	0.0	0.1	0.9	1.0	0.0	0.0	0.1	0.9	0.0	0.8	0.2	0.0
5	0.9	0.1	0.0	1.0	0.1	0.2	0.0	0.1	0.9	0.0	1.0	0.0	0.0
6	1.0	0.0	0.0	1.0	0.0	0.0	1.0	0.0	1.0	0.0	0.0	1.0	0.0
7	1.0	0.0	0.0	1.0	0.0	0.1	0.9	0.0	0.3	0.7	0.1	0.8	0.1
8	0.0	1.0	1.0	0.0	0.0	1.0	0.0	0.0	0.6	0.6	0.1	0.9	0.0
9	0.1	0.9	0.1	0.9	0.1	0.1	0.8	1.0	0.0	0.0	0.1	0.8	0.1
10	0.3	0.7	0.9	0.1	0.0	0.0	1.0	0.0	0.5	0.5	0.0	0.1	0.0
11	0.2	0.8	0.0	1.0	0.9	0.1	0.0	0.0	1.0	0.0	1.0	0.0	0.0
12	0.0	1.0	0.0	1.0	0.1	0.9	0.0	0.8	0.2	0.0	0.0	1.0	0.0
13	0.1	0.9	0.2	0.9	0.1	0.8	0.1	0.0	0.6	0.4	0.0	0.0	1.0
14	0.2	0.8	0.0	1.0	0.0	0.1	0.9	1.0	0.0	0.0	0.1	0.8	0.1
15	0.3	0.7	0.0	1.0	0.0	0.1	0.9	0.1	0.8	0.1	0.0	0.1	0.9

The work of surgeon and the nurse can be modelled both by 3 levels ($m_3 = m_4 = 3$), i.e.:

0 – (the fatal error),

1 – (sufficient), and

2 – (perfect or the work without any complication).

The structure function of the system for analysis this simplified version of a laparoscopic surgery is composed of 36 situations (state vectors). The monitoring and expert analysis of this system permits to obtain some samples that represent the system behavior. In Table II, 15 samples are shown. In this table, 15 of 36 possible state vectors are indicated only. However, this information is uncertain because the data from real monitoring is incomplete and values are ambiguous. The monitoring does not allow obtaining information about all possible samples (situations) for fatal medical errors, because it would imply patient's health. So, the expert (or experts) adds some information about the system behavior, which can be uncertain, too. For example, an expert can indicate for the sample under the consideration the system performance level with likelihood only: as operational ($\phi(\boldsymbol{x}) = 2$) with the likelihood of 0.0, partly operational ($\phi(\boldsymbol{x}) = 1$) with likelihood 0.1 and non-operational ($\phi(\boldsymbol{x}) = 0$) with the likelihood of 0.9.

Table II illustrates the correlation of monitoring data and the structure function for this system. The monitoring data can be transformed into the structure function (1) for the system based on the following rule: only the value with the highest likelihood is considered. For example, the variable x_1 in Table II has value 0 with the likelihood of 0.6 and value 1 with the likelihood of 0.4. The resultant value is defined as 0 in this case, but some of the state vectors are absent in the repository. Traditional mathematical approaches for system reliability analysis are based on the structure function, which cannot be used in this case. Therefore, the construction of a structure function (1) based on incomplete data requires a special transformation and the development of new methods.

B. Representation of system model in the form of an FDT

A decision tree (and FDT in particular) can be considered as an alternative form of the structure function. The structure function maps state vectors to each equivalence class of the system's performance levels. At the same time, a decision tree is a formalism for expressing mappings of input attributes (component's states) and output attribute/attributes (system performance level/s), consisting of an analysis of attribute nodes, which are linked to two or more sub-trees and leaves or decision nodes that are labeled with a class (in our case it is the system performance level) [14]. The outcome of the analysis is based on attribute values of a sample, where each possible outcome is associated with one of the sub-trees. A sample is classified by the starting at the root node of the tree. If this node is not a leaf, the outcome for the instance is determined and the process continues using the appropriate sub-tree. If a leaf is encountered, its label directs to the predicted class of sample. The system's component states are interpreted as values of the input attributes. The system's performance levels are considered as an instance that is classified into M classes.

FDT is one of the possible types of decision trees that permit to operate with fuzzy data (attributes) and methods of fuzzy logic. The construction of a FDT-based structure function assumes ambiguous data and that the analysis of such data can be implemented based on the methods of fuzzy logic [11, 30, 31]. The ambiguity of data values may be present in the attributes (system components states) and the exact class of the instance (system performance level).

There are different methods to induct a FDT [11, 24, 25, 31]. The principal goal of all methods is to select expanded attributes and determine the leaf node. The FDT induction is implemented based on some initial data that is interpreted as a training test. Every training sample includes n attributes A_1, ..., A_n and an output attribute B. The construction of the structure function supposes its correlation with the FDT (Table III): the system performance level is the output

attribute and the component's states (state vectors) are input attributes. Each input attribute (component state) A_i ($1 \le i \le n$) is measured by a group of discrete values from 0 to m_i-1 that agree with the values of the i-th component states: $\{A_{i,0},\ldots,A_{i,j},\ldots,A_{i,m_i-1}\}$. The FDT assumes that the input set A_1, ..., A_n is classified as one of the output value attributes B. The output attribute value B_w agrees with one of the system's performance levels and is defined by M values ranging from 0 to M -1 ($w = 0,\ldots, M$ -1).

For example, set input attributes $\{A_1, A_2, A_3, A_4\}$ and output attribute B for the success of laparoscopic surgery are indicated in Table IV according to FDT terminology. Each attribute is defined as: $A_1=\{A_{1,0}, A_{1,1}\}$, $A_2=\{A_{2,0}, A_{2,1}\}$, $A_3=\{A_{3,0}, A_{3,1}, A_{3,2}\}$, $A_4=\{A_{4,0}, A_{4,1}, A_{4,2}\}$ and $B = \{B_0, B_1, B_2\}$.

TABLE III.
CORRELATION OF THE TERMINOLOGIES OF FDT AND RELIABILITY ANALYSIS

FDT	System reliability
Number of input attributes: n	Number of system components: n
Attribute A_i ($i = 1, \ldots, n$)	System component x_i ($i = 1, \ldots, n$)
Attribute A_i values: $\{A_{i,0},\ldots,A_{i,j},\ldots,A_{i,m_i-1}\}$	The i-th system component state: $\{0, \ldots, m_i-1\}$
Output attribute B	System performance level $\phi(\boldsymbol{x})$
Values of output attribute B: $\{B_0, \ldots, B_{M-1}\}$	Values of the system's performance levels: $\{0, \ldots, M-1\}$
Decision table	Structure function

A fuzzy set A with respect to an universe U is characterized by a membership function $\mu_A : U \to [0,1]$, assign a A-membership degree, $\mu_A(u)$, to each element u in U. $\mu_A(u)$ gives us an estimation of u belonging to A. The cardinality measure of the fuzzy set A is defined by $M(A) = \Sigma_{u\in U}\, \mu_A(u)$, which is the measure of the size of A.

For $u\in U$, $\mu_A(u) = 1$ means that u is definitely a member of A and $\mu_A(u) = 0$ means that u is definitely not a member of A, while $0 < \mu_A(u) < 1$ means that u is partially a member of A. If either $\mu_A(u) = 0$ or $\mu_A(u) = 1$ for all $u \in U$, A is a crisp set. The set of input attributes A is crisp for which $\mu_A(u) = 0$ or $\mu_A(u) = 1$. The values of input and output attributes are defined by the membership function. They are obtained from the monitoring data (Table II) according to the correlation shown in Table IV. Indicated attributes values are used as a training test to construct the FDT.

In this paper, we adopt the FDT induction principle to construct the structure function based on cumulative information estimates [24, 25]. The cumulative information estimates allow defining the criterion of expanded attribute selection to induct FDT with different properties. These estimates are calculated by measures of entropy and information. Entropy and information have been introduced to the information theory as a probabilistic approach. The application of these measures assumes that the sum of possibilities of all values of every attribute equals 1 [26, 27, 28]. Note that the likelihood of attribute's value in terms of FDT induction is measured as confidence degree or degree of truth in this value.

TABLE IV.
ATTRIBUTES VALUES OF SYSTEM FOR LAPAROSCOPIC SURGERY PROCEDURE

Structure function	Attribute	Attribute values	Description of attribute values
The first component state, x_1	A_1	$A_{1,0}$	Device failure
		$A_{1,1}$	Devise working
The second component state, x_2	A_2	$A_{2,0}$	Error of anesthesiologist
		$A_{2,1}$	Anesthesiologist work without complication
The third component state, x_3	A_3	$A_{3,0}$	Error of surgeon
		$A_{3,1}$	Sufficient work of surgeon
		$A_{3,2}$	Perfect work of surgeon
The forth component state, x_4	A_4	$A_{4,0}$	Error of nurse
		$A_{4,1}$	Sufficient work of nurse
		$A_{4,2}$	Perfect work of nurse
System performance level $\phi(\boldsymbol{x})$	B	B_0	Fatal medical error
		B_1	Some imperfection
		B_2	Surgery without any complication

Figure 2 depicts the FDT for the laparoscopic surgery example. This FDT has 4 levels and includes all input attributes. This implies that all input attributes are considered to be significant for this system. The attribute A_3 is most significant as it agrees with the surgeon's work. Therefore, A_3 is associated with the FDT root (top node). This attribute can have the values $A_{3,0}$, $A_{3,1}$, and $A_{3,2}$, which are associated with branches of the FDT. Each branch agrees with a block of the output attribute values, and the confidences of every value of the output attributes are indicated. This block is a leaf if one output attribute has a sufficient level of confidence. In the other case, the FDT provides the analysis of the next input attribute. In this example, the value $A_{3,1}$ and $A_{3,2}$ of the attribute A_3 supposes the analysis of the attribute A_1 and A_2, respectively. The sufficient value of the output attribute is defined by the user between 0 to 1. For instance in the laparoscopic surgery example, this sufficiency of the FDT has been set to 0.750 (Fig. 2).

The FDT can be transformed into classification (decision) rules. A new sample e may be classified into different classes with different confidence. Let the FDT have R leaves $L = \{l_1, \ldots, l_r, \ldots, l_R\}$, then each leaf $l_r\in L$ corresponds to one (r-th) classification rule. The condition part of the classification rule is a group of conditions that is represented in the form: "attribute is attribute's value". Such conditions are interconnected with an AND operator. The attributes are associated with the nodes in the path from the root to the leaf l_r.

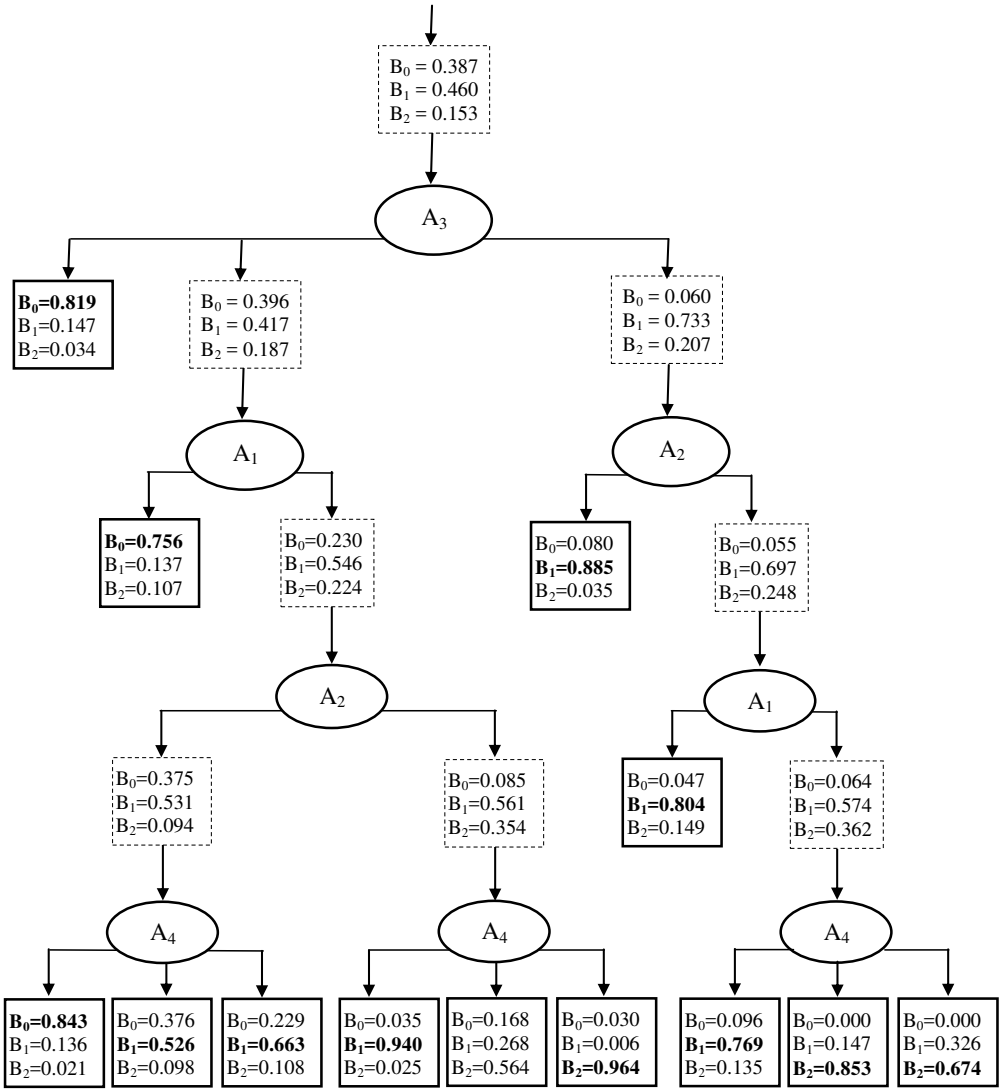


Fig. 2 The FDT for the construction of structure function to analyze laparoscopic surgery successful

The attribute's values are associated with the respective outgoing branches of the nodes in the path. Conclusions of the *r*-th rule are the values of class attribute B with their truthfulness vector $\mathbf{F}^r = [F_1{}^r;..., F_j{}^r;..., F_M{}^r]$ for each *r*-th leaf *l* and each *j*-th class B_j. Each value $F_j{}^r$ means the certainty degree (confidence) of the class B_j attached to the leaf node l_r.

Let's consider the path $P_r(e) = \{[A_{i_1 j_1}(e)]^r,\ldots,[A_{i_s j_s}(e)]^r,\ldots,[A_{i_S j_S}(e)]^r\}$ from the FDT root to the *r*-th leaf. This path $P_r(e)$ consist of *S* nodes which are associated with attributes $A_{i_1},.., A_{i_s},\ldots, A_{i_S}$ and respectively their *S* outgoing branches associated with the values $A_{i_1 j_1},.., A_{i_s j_s},\ldots , A_{i_S j_S}$. Then the *r*-th rule has the following form:

$$\text{IF } (A_{i_1} \text{ is } A_{i_1,j_1}) \text{ and } \ldots \text{and } (A_{i_S} \text{ is } A_{i_S,j_S}) \quad \text{THEN B (with truthfulness } \mathbf{F}^r) \tag{3}$$

The classification rules for the laparoscopic surgery example are shown in Fig.3.

C. Construction of the structure function based on the FDT

As it was shown in Section III, the structure function agrees with the decision table. The decision table can be constructed based on the FDT or classification (decision) rules or a [25]. A decision table indicates all possible values of input attributes and agrees with the structure function. Therefore, cal the decision table is calculated for all possible values of the component's states and considered as the structure function (1).

For the laparoscopic example, assume that the state vector is $\boldsymbol{x}$ = (0 1 1 0). According to the FDT (Fig. 2), the system output attribute for all possible vector states can have the value 0, 1, and 2 with the confidence of 0.387, 0.460, and 0.153, respectively. However, these confidences are under the given threshold of 0.750. Therefore, the analysis of the state vectors is implemented based on the FDT. It starts with the attribute A_3 (Fig. 2). The value of this component state is x_3 = 1 and the branch for the attribute value $A_{3,1}$ is

considered. The output attribute can have the value 0, 1, and 2 with the confidence of 0.396, 0.417, and 0.187, respectively. Again, all these confidences are below the threshold 0.750. Therefore, a decision about the value of the system's performance level is impossible and the analysis continues with the attribute A_1. The estimation of this attribute is implemented using a branch with attribute value $A_{1,0}$ because the specified state vector includes $x_1 = 0$. The branch of this value has the leaf-node. Therefore, the value of the output attribute is defined as the value with the maximal confidence, which is 0.756 for the value 0. Considering this, the system performance level for the specified state vector is $\phi(\boldsymbol{x}) = 0$. Therefore the fatal medical error is possible with confidence 0.756 if the level of the surgeon's work is sufficient and there is a malfunction of the laparoscopic robotic surgery machine. The analysis of other state vectors is similar and allows to obtain all possible values of the system performance level in the form of the structure function.

r=1 : IF A_3 is $A_{3,0}$ THEN B with $\mathbf{F}^1$ = [0.819; 0.147; 0.034];
r=2 : IF A_3 is $A_{3,1}$ *and* A_1 is $A_{1,0}$ THEN B with $\mathbf{F}^2$ = [0.756; 0.137; 0.107];
r=3 : IF A_3 is $A_{3,1}$ *and* A_1 is $A_{1,1}$ *and* A_2 is $A_{2,0}$ *and* A_4 is $A_{4,0}$ THEN B with $\mathbf{F}^3$ = [0.843; 0.136; 0.021];
r=4 : IF A_3 is $A_{3,1}$ *and* A_1 is $A_{1,1}$ *and* A_2 is $A_{2,0}$ *and* A_4 is $A_{4,1}$ THEN B with $\mathbf{F}^4$ = [0.376; 0.526; 0.098];
r=5 : IF A_3 is $A_{3,1}$ *and* A_1 is $A_{1,1}$ *and* A_2 is $A_{2,0}$ *and* A_4 is $A_{4,2}$ THEN B with $\mathbf{F}^5$ = [0.229; 0.663; 0.108];
r=6 : IF A_3 is $A_{3,1}$ *and* A_1 is $A_{1,1}$ *and* A_2 is $A_{2,1}$ *and* A_4 is $A_{4,0}$ THEN B with $\mathbf{F}^6$ = [0.035; 0.940; 0.025];
r=7 : IF A_3 is $A_{3,1}$ *and* A_1 is $A_{1,1}$ *and* A_2 is $A_{2,1}$ *and* A_4 is $A_{4,1}$ THEN B with $\mathbf{F}^7$ = [0.168; 0.268; 0.564];
r=8 : IF A_3 is $A_{3,1}$ *and* A_1 is $A_{1,1}$ *and* A_2 is $A_{2,1}$ *and* A_4 is $A_{4,2}$ THEN B with $\mathbf{F}^8$ = [0.030; 0.006; 0.964];
r=9 : IF A_3 is $A_{3,2}$ *and* A_2 is $A_{2,0}$ THEN B with $\mathbf{F}^9$ = [0.080; 0.885; 0.035];
r=10: IF A_3 is $A_{3,2}$ *and* A_2 is $A_{2,1}$ *and* A_1 is $A_{1,0}$ THEN B with $\mathbf{F}^{10}$ = [0.047; 0.804; 0.149];
r=11: IF A_3 is $A_{3,2}$ *and* A_2 is $A_{2,1}$ *and* A_1 is $A_{1,1}$ *and* A_4 is $A_{4,0}$ THEN B with $\mathbf{F}^{11}$ = [0.096; 0.769; 0.135];
r=12: IF A_3 is $A_{3,2}$ *and* A_2 is $A_{2,1}$ *and* A_1 is $A_{1,1}$ *and* A_4 is $A_{4,1}$ THEN B with $\mathbf{F}^{12}$ = [0.000; 0.147; 0.853];
r=13: IF A_3 is $A_{3,2}$ *and* A_2 is $A_{2,1}$ *and* A_1 is $A_{1,1}$ *and* A_4 is $A_{4,2}$ THEN B with $\mathbf{F}^{12}$ = [0.000; 0.326; 0.674];

Fig. 3 The classification rules for the construction of structure function for the analysis of successful laparoscopic surgery

It is important to note that this method of the structure function constructing based on FDTs permits to compute (restore) data missing from the monitoring. Therefore, probabilities of system performance can be calculated according to typical methods used in reliability engineering, i.e., based on the structure function. For example, the system availability (1) can be calculated for the system that is presented the laparoscopic surgery success based on the indicated values of the component state probabilities (Table V). The availability of this system according to (3) and data in Table V is calculated based on the structure function constructed through the FDT(Fig. 2):

$$A_0 = 0.098, \quad A_1 = 0.214, \quad A_2 = 0.688 \qquad (4)$$

TABLE V.
PROBABILITIES OF THE COMPONENTS STATES

System component description	Component's states probabilities		
	$p_{i,2}$	$p_{i,1}$	$p_{i,0}$
The laparoscopic robotic surgery machine functioning, x_1	—	0.98	0.02
The anesthesiologist's work, x_2	—	0.94	0.06
The surgeon's work, x_3	0.64	0.27	0.09
The nurse's work, x_4	0.47	0.35	0.18

The values of the availabilities in (4) imply that laparoscopic surgery can be with:

a) fatal medical error with probability 0.098,

b) sufficient result (some complications) with probabilities 0.214 and

c) perfect result 0.688 (without any complications).

Other measures can be computed by the structure function too. For example, importance measures for this system are defined according to the algorithms considered in [10, 30].

V. RESULTS

Our method has been investigated depending on the level specify of the initial data. We considered 3 benchmarks for the method evaluation [13, 15]:

- The system 1 has 3 performance levels, consist of 5 component and its structure function has 243 state vectors;
- The system 2 has 5 performance levels, consist of 4component and its structure function has 108 state vectors;
- The system 3 has 4 performance levels, consist of 5 component and its structure function has 521 state vectors;

The structure functions are known and defined for these systems. We have used this data to examine efficiency and accuracy of proposed method for the construction of the "new" structure function based on incompletely specified data. The incompleteness is modeled by random deleting of some state vectors and assigned performance level value. The range of deleted states is changed from 5% to 90%. The "new" structure function is constructed based these incompletely specified data and compare with initial structure function.

The constructed structure functions include individual or small groups of misclassified state vectors. Therefore, we have estimated this misclassification by an error rate. The constructed structure functions and initial completely and exact specified functions are compared and the error rate is calculated as the ratio of wrong values of the structure function to the dimension of unspecified part of the function. The experiments have been iterated 1000 times for every system. The error rate for every system and for different level of unspecified component states vector is shown in Fig.4. The error rate is depended on unspecified part of the initial data (state vectors) according results presented in

Fig.4. This error increases essentially, if the unspecified part is most than 80% for all investigated systems. We obtained an insignificant growth of the error rate if the unspecified part is less than 10%. Therefore according to this investigation we can declare recommended specify of initial data between 10% and 80%. This result is typical for FDT application [11, 12]. The specification of data less than 10% isn't sufficient for the correct construction of correlations between input and output attribute. And higher level of data specification (more 80%) restricts variations of FDT induction and causes misclassification values.

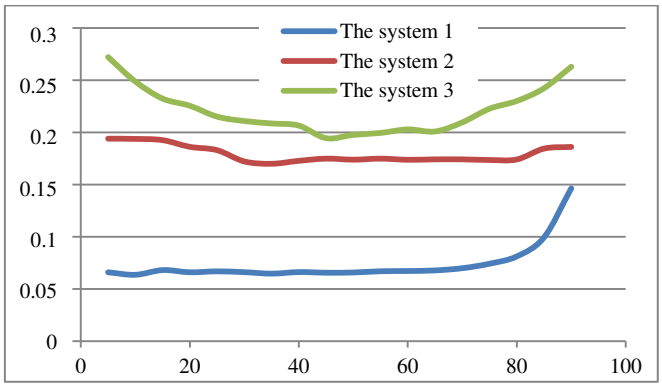


Fig. 4 The error rate for the construction of the structure function for three systems

VI. Discussion

The main contribution of this paper is that we have developed novel and original method for the construction of the structure function based on incompletely specified and ambiguous data. It directly supports MSS reliability estimation and can cope with uncertain data for the analysis of system reliability/availability. This is typical problem for reliability analysis of healthcare systems, where the data cannot be obtained for all possible situations.

The analysis of the error rate for the proposed method for the construction of the structure function based on FDT shows that this method has good efficiency. This method is acceptable for the incompletely data and the incompleteness of initial data can be indicated from 10% to 85%. The constructed structure function by the proposed method has less error rate than maximal error rate in interval of the incompleteness.

But error rate is not caused by range of incompleteness of initial data only. There are some factors that can influence to error rate. First of all it is specific properties of structure function. One of very important properties is monotony of structure function that is typical for coherent system [17, 30]. The structure function of the system 1 is monotonic and the error rate of this function is minimal. The structure function of the system 2 is monotonic too. Other property of structure function is "uniformity" that mean similar number of state vectors for all system performance levels. For example, the system 2 has maximal number of state vectors for one of performance level (46 of 108). Therefore range of incompleteness has small influence to error rate. Numbers of state vectors for every performance levels of the system 3 are similar. Therefore error rate increases for boundary range of incompleteness.

Beside the structure function properties, the system dimension (number of state vectors) influences to the error rate. This influence will be investigated in the future. Now, we can suppose that error rate decreases opposite to increasing of system dimension. The construction of the structure function of 50 state vectors is possible based on 20-30 state vectors (40-60% of defined state vectors). The structure function construction based on 5-10 state vectors (10-20%) is possible too. But level of accuracy depends on the quality of this set of state vectors. It will be essential to continue the verification and validation of our proposed method with data sets of different properties and sizes

Very important advantage of proposed method for constructing of the mathematical model of investigated system is possibility to ignore step of qualitative analysis that is typical in reliability analysis based on uncertain data. Methods for qualitative analysis (Failure mode and Effect Analysis (FMEA), Failure Mode Effects and Criticality Analysis (FMECA), Hazard Operability Study (HAZOP) etc.) are very good investigated in reliability engineering, but most of them include empirical evaluation of investigated system [8, 9, 13].Conclusion

Our comprehensive experiments have shown that reliability estimation of healthcare systems is possible with uncertain or incomplete data if the structure function is estimated using fuzzy decision trees. This opens a wide range of clinical applications for saver healthcare.

References

[1] Taylor E.F. "The reliability engineer in the health care system," in *Proc IEEE the 18th Annual Reliability & Maintainability Symposium*, USA, pp.245 – 248, 1972.

[2] Cohen T. "Medical and Information Technologies Converge," *IEEE Engineering in Medicine and Biology Magazine*, vol.23, no. 3, pp.59-65, 2004, http://dx.doi.org/10.1109/MEMB.2004.1317983

[3] Dhillon B.S. *Reliability Technology,Human Error, andQuality in Health Care*, CRC Press, 190 p., 2008.

[4] Taleb-Bendiab A, England D, at al. "A principled approach to the design of healthcare systems: Autonomy vs. governance," Reliability Engineering and System Safety, vol.91, no. 12, pp.1576–1585, 2006, http://dx.doi.org/10.1016/j.ress.2006.01.011

[5] Brall A., "Human Reliability Issues in Medical Care – A Customer Viewpoint," in *Proc. on Annual Reliability and Maintainability Symposium*, USA, pp. 46 – 50, 2006, http://dx.doi.org/10.1109/RAMS.2006.1677348

[6] Catelani M., Ciani L., Risaliti C., "Risk assessment in the use of medical devices: a proposal to evaluate the impact of the human factor," in *Proc. of IEEE International Symposium on Medical Measurements and Applications*, 2014, http://dx.doi.org/10.1109/MeMeA.2014.6860088

[7] Dhillon B.S., *Human Reliability and Error in Medicine*, World Scientific, 2003

[8] Deeter J., Rantanen E., "Human Reliability Analysis in Healthcare," in *Proc. of 2012 Symposium on Human Factor and Ergonomics in Health Care*, USA, pp.45-51, 2012, http://dx.doi.org/10.1518/HCS-2012.945289401.008

[9] Lyons M., Adams S., Woloshynowych M. and Vincent C., "Human reliability analysis in healthcare: A review of techniques," Int. J. of Risk and Safety in Medicine, vol. 16, no. 4, pp. 223–237, 2004

[10] Zaitseva, E., Kvassay, M., Levashenko, V., Kostolny, "New Methods for the Reliability Analysis of Healthcare System Based on Application of Multi-State System," in Applications of Computational Intelligence in Biomedical Technology, Eds.: Bris R., Majernik J., K.Pancerz, E.Zaitseva, Springer, pp. 229-251, 2016, http://dx.doi.org/10.1007/978-3-319-19147-8_14

[11] J. R. Quinlan, "Simplifying decision trees", Int. J. Man-Machine Studies, vol. 27, pp. 221-234, 1987, http://dx.doi.org/10.1016/S0020-7373(87)80053-6

[12] S. Mitra, K. M. Konwar, and S. K. Pal, "Fuzzy Decision Tree, Linguistic Rules and Fuzzy Knowledge-Based Network: Generation and Evaluation". Trans. on Syst., Man Cybernetics — Part C: Applications and Reviews, vol. 32, pp.328-339, 2002, http://dx.doi.org/10.1109/TSMCC.2002.806060

[13] Aven T., Zio E., Baraldi P., and Flage R., Uncertainty in Risk Assessment: The Representation and Treatment of Uncertainties by Probabilistic and Non-Probabilistic Methods, Wiley, 200 p., 2014, http://dx.doi.org/10.1002/9781118763032.ch07

[14] Tsipouras M. G., Exarchos T. P., and Fotiadis D. I., "A methodology for automated fuzzy model generation," Fuzzy Sets and Systems, vol. 159, no 23, pp. 3201-3220, 2008, http://dx.doi.org/10.1016/j.fss.2008.04.004

[15] B. Natvig, Multistate Systems Reliability Theory with Applications, Wiley, New York, 2011, http://dx.doi.org/10.1002/9780470977088

[16] E. Zio, "Reliability engineering: Old problems and new challenges", Reliability Engineering and System Safety, vol. 94, pp.125–141, 2009, http://dx.doi.org/10.1016/j.ress.2008.06.002

[17] T. Aven, and B. Heide, "On performance measures for multistate monotone system", Reliability Engineering and System Safety, vol. 41, pp.259–266, 1993.

[18] A. Lisnianski, and G. Levitin, Multi-state System Reliability. Assessment, Optimization and Applications. Singapore, SG: World Scientific, 2003.

[19] H.-Z.Huang, "Structural reliability analysis using fuzzy sets theory", Eksploatacja i Niezawodnosc – Maintenance and Reliability, vol.14, no. 4, pp.284–294, 2012.

[20] D. Ley, "Approximating process knowledge and process thinking: Acquiring workflow data by domain experts," in Proc. 2011 IEEE International Conference on Systems, Man, and Cybernetics, pp. 3274–3279, 2011, http://dx.doi.org/10.1109/ICSMC.2011.6084174

[21] D. Yu, W. S. Park, "Combination and evaluation of expert opinions characterized in terms of fuzzy probabilities", Annals of Nuclear Energy, vol. 27, no 8, pp. 713–726, 2000, http://dx.doi.org/10.1016/S0306-4549(00)82012-5

[22] Y. Wang, and L. Li, "Effects of Uncertainty in Both Component Reliability and Load Demand on Multistate System Reliability," IEEE Trans. on System, Man and Cybernetic – Part A: Systems and Humans, vol. 42, no. 4, pp. 958-969, 2012, http://dx.doi.org/10.1109/TSMCA.2011.2181163

[23] N. A. Stanton, and C. Baber, "Error by design: methods for predicting device usability," Design Studies, vol. 23, no. 4, pp.363-384, 2012, http://dx.doi.org/10.1016/S0142-694X(01)00032-1

[24] V. Levashenko, and E. Zaitseva, "Usage of New Information Estimations for Induction of Fuzzy Decision Trees," in Lecture Notes in Computer Science LCNS2412, Springer-Verlag, pp. 493-499, 2002, http://dx.doi.org/ 10.1007/3-540-45675-9_74

[25] V. Levashenko, E. Zaitseva, and S. Puuronen, "Fuzzy Classified based on Fuzzy Decision Tree," in Proc. EUROCON, Warsaw, Poland, pp.823–827, 2007, http://dx.doi.org/ 10.1109/91.227387

[26] R. Krishnapuram, and J. Keller, "A possibilistic approach to clustering," IEEE Trans. on Fuzzy Systems, vol.1, pp.98–110, 1993,

[27] R. Kruse, Ch. Doring, and M.-J. Lesot, "Fundamentals of Fuzzy Clustering," in Advances in Fuzzy Clustering and its Applications, eds. J.Valente de Oliveira and W. Pedrycz, Wiley, 434 p., 2007, http://dx.doi.org/ 10.1002/9780470061190.ch1

[28] H. Tanaka, L.T. Fan, F.S. Lai, and K. Toguchi, "Fault-tree analysis by fuzzy probability," IEEE Trans. on Reliability, vol. 32, pp. 453-457, 1983, http://dx.doi.org/10.1016/S0019-9958(65)90241-X

[29] H. R. Patel, A. Linares and J. V. Joseph., "Robotic and laparoscopic surgery: cost and training," Surg Oncol, vol.18, no.3, pp. 242-246, 2009, http://dx.doi.org/10.1016/j.suronc.2009.02.007

[30] M. Kvassay, E. Zaitseva, and V. Levashenko, "Minimal cut sets and direct partial logic derivatives in reliability analysis," in Proc. ESREL, Wroclaw, Poland, pp. 241-248, 2014, http://dx.doi.org/10.1201/b17399-37

1st International Workshop on AI aspects of Reasoning, Information, and Memory

THERE is general realization that computational models of human reasoning can be improved by integration of heterogeneous resources of information, e.g., multidimensional diagrams, images, language, syntax, semantics, memory. While the event targets promotion of integrated computational approaches, we invite contributions from any individual areas related to information, language, memory, reasoning.

TOPICS

We welcome submissions of papers on the following topics, without limiting to them, across approaches, methods, theories, and applications:

- Reasoning systems — theories and applications
- Proof systems and model checkers
- Theories of computation and information
- Interactive computation and reasoning
- Computation and reasoning with heterogeneous information
- Space and time in information, language, memory, and reasoning
- Partiality, underspecification, vagueness, and possibilities
- Detection of and reasoning with inconsistency
- Logic and language — approaches, theories, methods
- Computational morphology, syntax, semantics, and interfaces between these
- Constraint-based and type-theoretic approaches and grammars
- Logical approaches to multilingual processing
- Logical and computational foundations in machine learning and information retrieval
- Mathematics for linguistics and cognitive science
- Reasoning, information, and memory in computational neuroscience and life sciences
- Interdisciplinary approaches to information, language, memory, and reasoning

EVENT CHAIRS

- **Christiansen, Henning,** Roskilde University, Denmark
- **Jiménez López, María Dolores,** GRLMC, Universitat Rovira i Virgili, Spain
- **Loukanova, Roussanka,** Department of Mathematics, Stockholm University, Sweden

PROGRAM COMMITTEE

- **Andreasen, Troels,** Roskilde University, Denmark, Denmark
- **Angelov, Krasimir,** University of Gothenburg, Sweden
- **Becerra, Leonor,** Jean Monnet University
- **Grabowski, Adam,** Institute of Informatics, University of Bialystok, Bialystok, Poland
- **Kornilowicz, Artur,** Institute of Informatics, University of Bialystok, Poland, Poland
- **Kübler, Sandra,** Indiana University, United States
- **Moss, Larry,** Indiana University, United States
- **Nilsson, Jørgen Fischer,** Technical University of Denmark, Denmark
- **Parmentier, Yannick,** LIFO, Université d'Orléans, France
- **Ranta, Aarne,** University of Gothenburg, Sweden, Sweden
- **Schwarzweller, Christoph,** Institute of Informatics, University of Gdansk, Poland
- **Villadsen, Jørgen,** Technical University of Denmark, Denmark

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 343–352
DOI: 10.15439/2016F440
ACSIS, Vol. 8. ISSN 2300-5963

From Discourse Representation Structure to Event Semantics: A Simple Conversion?

Daniel Dakota and Sandra Kübler
Indiana University
Email: {ddakota, skuebler}@indiana.edu

***Abstract*—Many applications in Natural Language Processing require a semantic analysis of sentences in terms of truth-conditional representations, often with specific desiderata in terms of which information needs to be included in the semantic analysis. However, there are only very few tools that allow such an analysis. We investigate the representations of an automatic analysis pipeline of the C&C parser and Boxer to determine whether Boxer's analyses in form of Discourse Representation Structure can be successfully converted into a more surface oriented event semantic representation, which will serve as input for a fusion algorithm for fusing hard and soft information. We use a data set of synthetic counter intelligence messages for our investigation. We provide a basic pipeline for conversion and subsequently discuss areas in which ambiguities and differences between the semantic representations present challenges in the conversion process.**

I. Introduction

MANY applications in Natural Language Processing require a semantic analysis of sentences. However, automatic semantic analysis is a field in its infancy in Natural Language Processing. There is work on automatically analyzing semantic role labeling, as evidenced by two shared tasks at the Conference on Natural Language Learning [1], [2] and a special issue of the journal *Computational Linguistics* [3]. But for many downstream applications related to text understanding, semantic roles do not provide enough information. Our current work focuses on fusing soft and hard information, where soft information constitutes natural language. A fusion algorithm accepts information from different sources and provides an integrated, accurate, informative whole. While fusion algorithms for sensor data are advanced and reliable, efforts to include natural language are in their early stages [4]. When language is included, fusion often includes inference mechanisms [5]. In order to be able to integrate language information into a fusion approach, we need to provide the information in a variant of predicate logic, on which inference and fusion algorithms can work.

There are existing approaches to analyzing language into semantic representations based on different syntactic formalisms (cf. e.g., [6] for LFG and [7] for TAG). We focus here on truth-conditional semantics based on Combinatory Categorial Grammar (CCG) [8] since this grammar formalism provides the closest match to our needs in terms of the target predicate logic. CCG relies on combinatory logic, which is equivalent in expressive power to lambda calculus. One approach to parsing CCG is the C&C parser [9], [10], which can be used in combination with Boxer [11], [12], [13], a module that converts the CCG syntax to semantic representations in the form of Discourse Representation Structure (DRS) [14], [15]. Other CCG-based approaches attempt learning semantic representations from different sources directly (e.g. [16], [17]).

Our target semantic representation is a form of event semantics, which means that neither parser provides us with analyses that are usable directly. Thus, we present work on investigating a rule-based conversion from Discourse Representation Structure as provided by Boxer to our target event semantic representations.

The remainder of this paper is structured as follows: We first provide more details about the conversion task in section II, then we briefly introduce our target semantic representations in section III, focusing on those aspects and distinctions that we target in the conversion. We then describe the analysis pipeline in section IV and discuss cases that can be converted in a rule-based fashion in section V. Finally, we discuss linguistic phenomena that present challenges for a conversion in section VI and conclude with a discussion of approaches to handle those difficulties (section VII).

II. Task Overview

Our task is to perform a semantic analysis of sentences in order to use them in a data fusion model for fusing hard and soft information [18], [19]. The fusion model expects an analysis in terms of first order logic and can be extended to a Davidsonian model. In a Davidsonian model, semantics is non-propositional, and references are integrated into the semantic description. References between events are described using event variables.

Since an unlimited truth-conditional analysis of unrestricted sentences is a very challenging task and since we have a very specific task as downstream consumer of our annotations, we have decided to reduce the complexity of the task of semantic analysis by assuming an automatic syntactic simplification of the sentences to be analyzed. In contrast to standard approaches to sentence simplification, our syntactic simplification model (currently under development) will focus on specific syntactic phenomena and will simplify only sentences that display such phenomena. Simplification will be performed by a machine learning module trained on a small set of sentences displaying a specific phenomenon, based on a dependency parse. The simplified sentences will then be reparsed by the CCG parser.

978-83-60810-90-3/$25.00

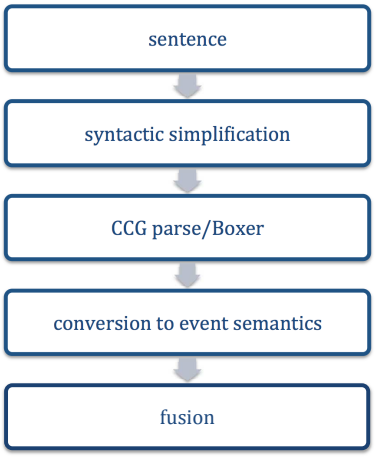


Fig. 1. Proposed Pipeline from the input sentence to information fusion.

The goal is a simplification without loss of information; we currently target coordination, reported speech, and passive sentences. Examples of simplifications are shown in examples (1)–(3). Note that in the case of reported speech, the reporting action is converted into an action and a certainty value (shown in square brackets), both of which can directly be used for information fusion.

(1) a. The man bought a book and a map.
b. The man bought a book. The man bought a map.

(2) a. The anonymous caller reported that an attack would happen next week.
b. An attack will happen next week. [action: report; certainty: 0.9]

(3) a. The man was given an object by a passer-by.
b. The passer-by gave the man an object.

As a consequence of the simplification step, we assume that the simplified sentences will be easier to parse by the CCG parser used in the semantic analysis. The final pipeline of the process is shown in Fig. 1.

III. The Annotated Data Set for Event Semantics

We use the SYNCOIN data set [20] for our experiments. SYNCOIN is a synthetically created set of counter insurgency scenarios in the form of collections of intercepted phone conversations and intelligence reports. The data set was designed to support approaches towards the fusion of hard and soft information.

We have access to truth-conditional semantic annotations of a set of sentences that constitute five "threads". These sentences were annotated within our project. The annotation scheme is based on Davidsonian event semantics [21], and the annotations are mostly rather surface oriented, in order to allow for robust automatic processing. The annotations are based on the following principles:

TABLE I
Annotation Scheme

Annotation	Description
x	named entity: person, group, or organization
y	a location
t	a time variable
e	an event variable
z	any other object

Since verbs report actions, they introduce event variables, and their syntactic arguments function as logical arguments in the following order: event, subject, direct object, and indirect object. We show an example sentence and its annotation in (4). Here, the verb *refuse* is analyzed as the 50th event in the thread, and it has two arguments, `x6` referring to *men* and `e2` referring to `attack`. Note that the annotation also states that the agent/subject of the attack are the men (`x6`).

(4) a. The men refuse to reveal operational details of the attack.
b. refuse(e50,x6,e51) ∧ men(x6) ∧ reveal(e51,x6,z14) ∧ details(z14,e44) ∧ operational(z14) ∧ attack(e14,x6)

There are two more points worth mentioning concerning the example: 1) Verbs are not the only concepts that introduce events. In the example, the nominalization *attack* is also represented as an event. 2) Both the men and the attack were mentioned previously in the text, which is indicated by shared variables. Thus, we integrate coreference information across sentence boundaries into the annotations.

The annotation scheme uses five types of variables, as shown in Table I. Nouns typically introduce variables of types `x` or `z` whereas the type `e` is typically introduced by verbs and nominalized verbs. Variable types `y` and `z` may be introduced by a wider range of parts of speech. Depending upon context, a word can potentially belong to more than one category.

All variables introduced by indefinite noun phrases and verbs are existentially quantified. However, existential quantifiers are not included as part of the semantic annotation. They are assumed to have scope over all sentences in the annotated document. Modifiers are introduced by adjectives, adverbs, and prepositional phrases. Adjectives typically introduce a new predicate, which is a property of objects as seen in (4) in the phrase *operational details*. Adverbs also introduce a new property. However, since they typically modify events, they have an event variable as argument, as seen in (5). Names are introduced by a predicate *named*. Possessive pronouns introduce a modifier which is annotated as a separate predicate as seen in (6). Units of measurements are introduced by a special predicate introducing the unit of measure and the numerical value, see the example in (7).

(5) a. The car drove fast.
b. car(z1) ∧ drove(e1,z1) ∧ fast(e1)

(6) a. John's car departed.

b. car(z1) ∧ possessive(x1,z1) ∧ named(x1,John) ∧ departed(e1,x1)

(7) a. 250 gallon tanks
b. tanks(z1) ∧ size_in_gallon(z1, 250)

Temporal phrases normally modify events, but can also introduce dates as seen in (8). Quantification is handled in a surface oriented way, as shown in (9). However, existential quantification over events is introduced by negation (see (10)) and questions.

(8) a. The man arrived on 01\05\10.
b. man(x1) ∧ arrived(e1,x1) ∧ on(e1,x1) ∧ date(t1,010510)

(9) a. Omar Khrayesh visits several bookstores in Adhamiya.
b. visit(e3,x1,y1) ∧ named(x1, Omar Khrayesh) ∧ bookstores(y1) ∧ several(y1) ∧ in(y1,y2) ∧ named(Adhamiya).

(10) a. Mallahati did not respond.
b. named(x1, Mallahati) ∧ not(∃e respond(e,x1))

IV. Analyzing Sentences Using DRS

A. Parsing

As discussed above, we utilize Combinatory Categorial Grammar [8] since it provides a clearly defined interface between syntax and truth-conditional semantics, making it ideal for our purposes. We use the C&C parser [9], [10] in combination with Boxer [11], [12], [13]. The parser provides pre-trained models based on CCGbank [22], sections 02-21 and MUC 7 [23]. We parse single sentences of SYNCOIN data into a CCG derivation. These derivations then serve as input for Boxer.

B. Boxer's DRS Analyses

Boxer is a module that uses the CCG derivations produced by the C&C parser to generate semantic representations in the form of Discourse Representation Structures (DRS), which are based on Discourse Representation Theory [15]. The theory assumes that hearers incrementally build a mental representation of a discourse, the DRS, which is a representational and non-compositional semantic representation. A DRS mainly models the referents of a discourse and the conditions that hold. A referent is an entity within the discourse while a condition is a predicate demonstrating properties of the respective referents.

A DRS consists of two parts, describing the referents and the conditions respectively. One important deviation from standard Discourse Representation Theory in Boxer is the use of a Neo-Davidsonian analysis of events and thematic roles [11]. This means that the representations provided by Boxer are close to our target event annotations, but that we need to abstract away from the semantic role representation and the more structured representation of referents.

DRS conditions are either basic or complex[1]. Complex conditions express phenomena such as implication or negation. In this paper, we focus on basic conditions, which express equalities, one-place relations, two-place relations, names, or time expressions. One-place relations are introduced by nouns, verbs, adjectives, and adverbs while two place relations are used to express verb roles and prepositions.

[1]A full account of Boxer's DRS representations can be found at http://svn.ask.it.usyd.edu.au/trac/candc/wiki/DRSs

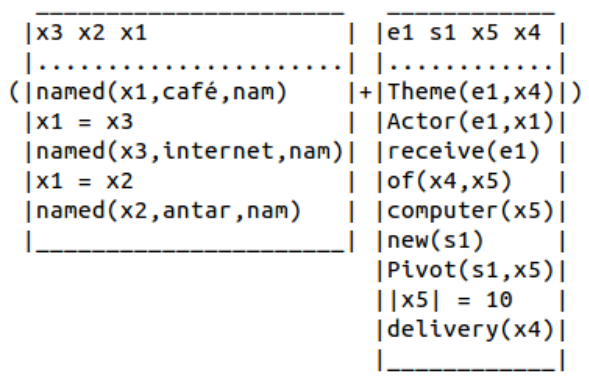


Fig. 2. Boxer DRS for sentence (11) in box format.

Boxer provides its analyses in two forms: a box representation and the same information in PROLOG style. The DRS structure of the SYNCOIN sentence (11) in box form is shown in Fig. 2. Each box, representing a single DRS, has the referents at the top, and the conditions within the box. The conditions of the second box show that the *café* is considered the actor and *delivery* the theme of the receiving event. The plus sign indicates the merging of the two basic DRS structures into a more complex one.

(11) The Antar Internet Café received a delivery of 10 new computers.

The second representation that Boxer offers is in PROLOG style, as shown in Fig. 3 (for example (11)). Here, the complex DRS is indicated by the merge operation. Since this representation is more easily processed automatically, we use it in our conversion to event semantics.

There are two primary types of information in the PROLOG representation, individual properties of the token and the semantic information of the tokens in the DRS. These two pieces of information are distinct, particularly as there are often tokens that do not have a semantic role in the DRS. Every token has an identification number indicating the position of the token in the sentence. This is accompanied by the original token, the part-of-speech (POS) tag of the token, the lemma, and information whether it is a named entity.

In regard to the semantic information in Fig. 3, we focus on two different types of conditions, predicates (pred) and relations (rel). Examining pred(x4,delivery,n,0)] more closely, a predicate consists of the referent (`x4`), the lemma (*delivery*), a general POS tag (n), and the word's named entity status (0 indicating false). A relation is represented as rel(e1,x1,'Actor',0) consisting of the event (`e1`), the first referent (`x1`), the thematic role of the referent (Actor), and the sense (0 indicating

```
id(1,1).
sem(1,[1001:[tok:'The',pos:'DT',lemma:the,namex:'O'],
1002:[tok:'Antar',pos:'NNP',lemma:'Antar',namex:'O'],
1003:[tok:'Internet',pos:'NNP',lemma:'Internet',namex:'O'],
1004:[tok:'Café',pos:'NNP',lemma:'Café',namex:'O'],
1005:[tok:received,pos:'VBD',lemma:receive,namex:'O'],
1006:[tok:a,pos:'DT',lemma:a,namex:'O'],
1007:[tok:delivery,pos:'NN',lemma:delivery,namex:'O'],
1008:[tok:of,pos:'IN',lemma:of,namex:'O'],
1009:[tok:'10',pos:'CD',lemma:'10',namex:'O'],
1010:[tok:new,pos:'JJ',lemma:new,namex:'O'],
1011:[tok:computers,pos:'NNS',lemma:computer,namex:'O'],
1012:[tok:'.',pos:'.',lemma:'.',namex:'O']],
merge(drs([[]:x3,[]:x2,[1001]:x1],
[[1004]:named(x1,café,nam,nam),[]:eq(x1,x3),
[1003]:named(x3,internet,nam,nam),[]:eq(x1,x2),
[1002]:named(x2,antar,nam,nam)]),drs([[]:e1,[]:s1,[]:x5,
[1006]:x4],[[]:rel(e1,x4,'Theme',0),[]:rel(e1,x1,'Actor',0),
[1005]:pred(e1,receive,v,0),[1008]:rel(x4,x5,of,0),
[1011]:pred(x5,computer,n,0),[1010]:pred(s1,new,a,0),
[]:rel(s1,x5,'Pivot',0),[1009]:card(x5,10,eq),
[1007]:pred(x4,delivery,n,0)])).
```

Fig. 3. Boxer DRS for sentence (11) in PROLOG representation.

none). The sense of the word will not be further addressed since it is not utilized in the conversion.

V. From DRS to Event Semantics

Our goal is to convert Boxer representations to our event semantic annotation scheme as accurately as possible. In the current paper, we first investigate the degree to which the required information is available from the Boxer analysis. A successful conversion can be hampered by several different issues: when information required in semantic representation is not present in Boxer's analyses, when Boxer is not consistent in representing a specific phenomenon, or when the analyses are incorrect. We present a basic methodology for the conversion, highlighting areas in which a successful conversion can be achieved with a high degree of success. We use Boxer given the following settings: PROLOG output format (but for readability, we present example in box format below), using DRS representations, using standard thematic proto-roles, and distinguishing variable types (into events, arguments, and modifiers).

A. Basic Conversion Principles

The PROLOG syntax allows for the ability to scan all aspects of the parse and develop checking mechanisms. The output is systematic in its structure, which allows an easy identification and isolation of referents and relations if they are produced by Boxer. In order to facilitate a closer examination of the conversion, we simplify example (11) to *café received a delivery*. The corresponding part of the DRS is presented in Fig. 4, the event semantic representation in (12).

```
1004:[tok:'Café',pos:'NNP',lemma:'Café',namex:'O'],
1005:[tok:received,pos:'VBD',lemma:receive,namex:'O'],
1007:[tok:delivery,pos:'NN',lemma:delivery,namex:'O'],
rel(e1,x4,'Theme',0)
rel(e1,x1,'Actor',0)
[1004]:named(x1,café,nam,nam)
[1005]:pred(e1,receive,v,0)
[1007]:pred(x4,delivery,n,0)])
```

Fig. 4. PROLOG representation for *café received a delivery*.

(12) a. café received a delivery
b. café(y1) $\wedge$ received(e1,y1,z1) $\wedge$ delivery(z1)

Relation representations vary in the information they represent, but they always provide information about the event, the referent, and the relation. We use regular expressions to extract all relations in the PROLOG representation, identifying the referents involved, and cross-reference them. In Fig. 4, there are two relations (Actor and Theme) which are captured and then cycled through. Once the referent has been identified, the corresponding predicate is extracted using additional regular expressions in order to extract relevant information: the referent, the lemma, and an indication of a general POS category.

We isolate the thematic role within each relation (in our case Actor and Theme). Based on the specific role, we need to employ specialized conversion strategies since a Pivot, for example, provides more complex information than an Agent. Once a relation type has been identified, the referents are extracted. For the case of Actor, we extract `e1` and `x1`. In this relation, the first argument, `e1`, refers to the first event introduced in the sentence. The second argument, `x1`, refers to an argument identified by the event. We then subsequently extract information on the corresponding predicate.

As mentioned in Section III, we introduce arguments of events in a specific order: the event, then the subject, then the direct object. This is demonstrated in (12), where the subject is represented by `y1` and the direct object by `z1`. In general, there are no discrepancies between the ordering of thematic roles and grammatical functions since we assume that passive sentences have been simplified to active ones.

B. Challenges

Our main challenge is to correctly identify and convert the PROLOG variables into our event semantic variables. As presented in Table I, there are five possible variables types in the event semantics. As the PROLOG representation for events is `e`, these events are easily transferable to our scheme. However, as explained in Section III, verbs are not the only type of concepts that introduce events, and there is no direct correlation between other event types in the Boxer representations and in the event semantics. For example, verbal nominalizations, such as *attack* in example (4) are treated as referents like any other nouns in Boxer. Thus, in order to handle such cases properly, we will need to use additional

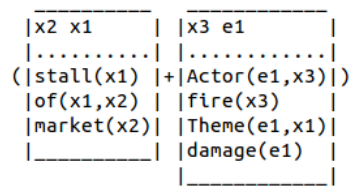


Fig. 5. Passive example.

semantic resources as well as event anaphora detection and resolution strategies.

In addition to event variables, we have four distinct argument variables. Returning to example in (12) and Fig. 4, the subject (*café*) is a named entity with an unknown tag in the PROLOG representation, thus the challenge is to correctly identify that the variable associated with the referent is of type `y` given that it is a location. The direct object (*delivery*) is assigned `z1` since it is the first non-named entity argument that is not a location. If the parser recognizes named entities, then they can be directly converted to corresponding variables (e.g., org $\rightarrow$ `x`). The difficulty lies in distinguishing between any two non-event labels for given tokens, particularly objects and locations, as although named entities are often detected by the parser while parsing the SYNCOIN data, they are frequently mislabeled, as many of these entities cannot be found in the training data of the parser. Thus, we must distinguish the possible named entities and their associated variables in our conversion (e.g. *Antar café* vs. *weapons cache*). This problem will be further addressed in section VI.

Additionally, the conversion requires multiple checks. One issue concerns the numbering of types of variable, i.e., how many events have been introduced, to correctly assign a number to a newly introduced variable. Another complexity arises from the fact that multiple variable types in the event semantic representations are represented by a single variable in Boxer; thus we need strategies to determine the resulting variable types via heuristics. For example, a referent of Boxer's type `x` can be of any type shown in Table I.

One discrepancy between the two semantic annotations concerns the thematic roles used by Boxer and the more surface oriented annotations in the event semantics that are based on syntactic arguments rather than thematic roles. The thematic roles generally correspond to syntactic arguments of a specific event and relation. For example, the role Actor is associated with the subject and Theme with a direct object. For many sentences, the order of the arguments is associated with the same sequence as our annotation scheme. However, there is an issue with passive sentences where there is no direct correspondence between roles and grammatical functions, as in Fig. 5 when parsing the passive sentence in (13).

(13) The market stalls were damaged by fire.

Here, strictly associating the Theme with a direct object position, and Actor with the subject, when introducing arguments of an event would result in an incorrect conversion. For such

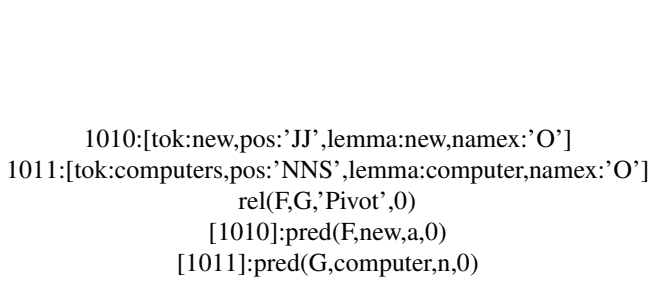


Fig. 6. PROLOG representation of the modification example in (14).

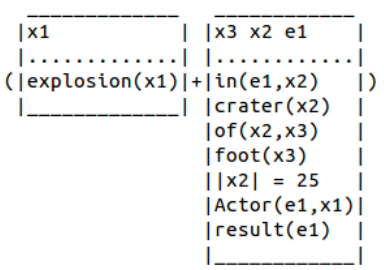


Fig. 7. DRS representation of the sentence in (15).

cases, we rely on sentence simplification to resolve the issue. Otherwise, we would have to go back to the CCG derivations and extract the syntactic arguments.

Another challenge arises from the representation of nominal modifiers. While the event semantic scheme opts for a surface oriented representation, converting these modifiers into predicates, Boxer uses Pivot roles, which capture direct relationships between referents and modifiers, such as between adjectives and nouns. Returning to the sentence depicted in Fig. 2, the adjective *new* modifies *computers* with the relevant PROLOG representation shown in Fig. 6. In this case, we need to extract the arguments of the Pivot and convert the modifier *new* into a predicate that directly applies to the argument `computers`. The result of this conversion is shown in (14).

(14) computers(z1) $\wedge$ new(z1)

Units of measurement are treated similarly. The DRS representation for the sentence in (15) is shown in Fig. 7. We can see that it contains both information that the referent is a number (specified by the cardinality information) and that it modifies the noun *crater* while the measurement *foot* is in an `of` relation to the number.

(15) a. The explosion resulted in a 25 foot crater.
b. crater(z1) $\wedge$ size_in_foot(z1,25)

Despite the differences described above, we can convert those concepts directly into event semantics due to the consistent and explicit relationship between nominals and their modifiers.

VI. Discrepancies

The basic conversion principles described above allow for an accurate transformation of many of the basic phenomena in sentences with a high degree of consistency. However, there are also discrepancies between the two representations that are less easily reconciled. We describe here the three

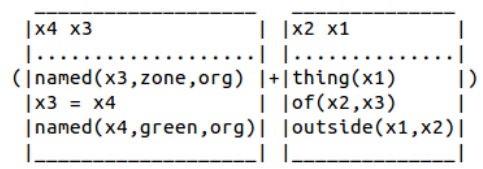


Fig. 8. DRS representation of *outside of the Green Zone*.

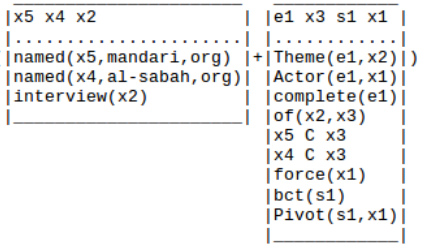


Fig. 9. DRS representation of *BCT forces*.

most important issues, namely the treatment of parsing errors, inconsistencies in the Boxer output, and coreference information.

A. Parsing Errors

While the C&C parser is an accurate parser, there are phenomena that are challenging for the parser, especially where Boxer does not have enough information to make correct decisions. To a certain degree, we anticipate such problems, such as coordination, which are challenging for any parser, and handle them in sentence simplification. Other phenomena, however, are more difficult to address.

1) Named Entity Referents: The C&C parser has a named entity recognizer [24], whose analyses are also used by Boxer. Named entities are categorized into various categories including geographical (geo), person (per), organization (org), and nam (unknown). This information provides the basis for determining the correct variable type in the event semantic representations (i.e., assigning `x` for a name or a person and `y` for a location). However, given the nature of the SYNCOIN data, there are many names that the parser mislabels, or occasionally fails to recognize. For example, *Green Zone* is correctly identified as a named entity but is consistently labeled as an organization rather than a location (see Fig. 8). In Fig. 9, the organization *BCT forces* is parsed as a referent *force* with a modifier (*BCT*) instead of being treated as a single multi-word named entity. We thus cannot always accept the named entity information provided for the correct categorization between `x` and `y` variables in the event semantics.

2) Recognizing Possessive Relationships and Compounds: Possessive relationship are represented with an `of` relation. This covers both cases where the possessor is nominal and pronominal. Fig. 10 shows the DRS representation for the sentence in (16). In this example, the pronoun *his* is converted into an `of` relation between *meeting* and a male referent.

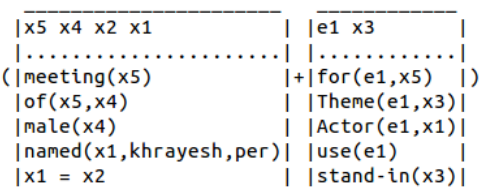


Fig. 10. DRS representation of the example in (16).

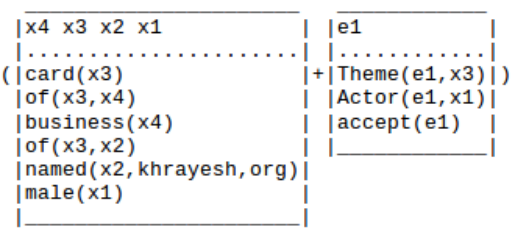


Fig. 11. DRS representation of the example in (17).

(16) a. Omar Khrayesh uses a stand-in for his meeting.
b. Omar_Khrayesh(x1) ∧ use(e1,x1,x2) ∧ stand_in_for(e2,x2,e3) ∧ meeting(e3,x3,x1)

However, the same relation is also used to represent noun-noun compounds. Fig. 11 shows the DRS analysis of the sentence in (17)[2]. Here, the compound *business card* is analyzed as *card of business*.

(17) a. He accepts Khrayesh's business card.
b. accept(e1,x1,z1) ∧ business_card(z1) ∧ named(x2,Khrayesh) ∧ possessive(x2,z1)

This parallel treatment of possessives and compounds in Boxer's DRS introduces the need to distinguish between the two usages since they differ in the event semantic representation: In the event semantics, the compound noun is retained (e.g., *business_card* in (17)) while the possessive pronoun is resolved into a possessive relation (e.g., *possessive*).

B. Inconsistencies

1) Adverbial Modifiers: In the DRS produced by Boxer, non-temporal modifiers are predominantly categorized in two different ways, as seen for the sentences in (18) and the DRS presented in Fig. 12 and Fig. 13 respectively.

(18) a. At approximately 1304hrs he appeared.
b. He is only interested in a prosperous Iraq.

For the first sentence, the adverb *approximately* is analyzed as being in a Pivot relation to the time. For the second

[2]Note that Boxer provides an option to analyze noun-noun compounds using prepositions [25]. However, this analysis introduces additional variation in the semantics-based relations between nouns, introducing an added level of complexity. For this reason, we choose not to utilize this option.

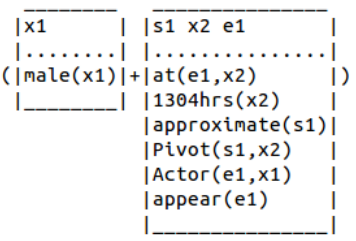


Fig. 12. Example of an adverbial pivot

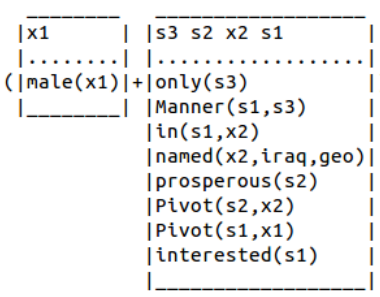


Fig. 13. Example of an adverb of manner

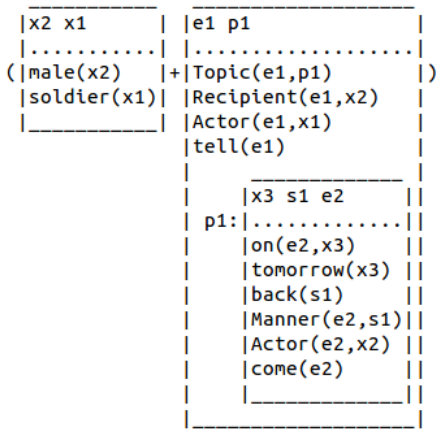


Fig. 14. Example of a temporal adverb

sentence, *only* as an adverb of manner. Both indicate that there is a type of relationship to a referent although they are being interpreted differently in terms of semantics. The pivot analysis overlaps with adjectival modifiers. The manner analysis overlaps with phrasal verbs, which will be addressed further in section VI-B3.

2) Temporal Modifiers: Time expressions are not consistently or accurately captured in every case [12]. This becomes evident in temporal modifiers. For example, the DRSs for the first two sentences in (19), depicted in Fig. 14 and Fig. 15, show that in the first case, the temporal adverb *tomorrow* is analyzed as being in an `on` relation to the event *come* while in the second case, the adverb *now* directly modifies the event *want*. Now, we could argue that the latter is a consequence of having fronted the adverb. However, if we use the adverb in

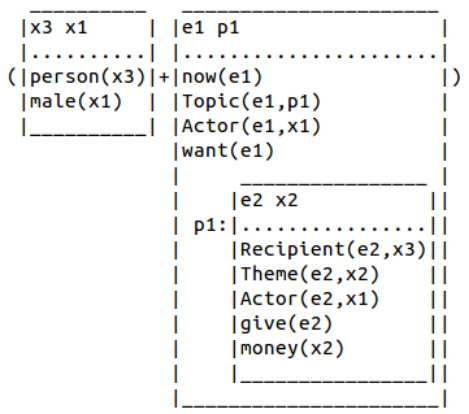


Fig. 15. Example of the temporal adverb fronted.

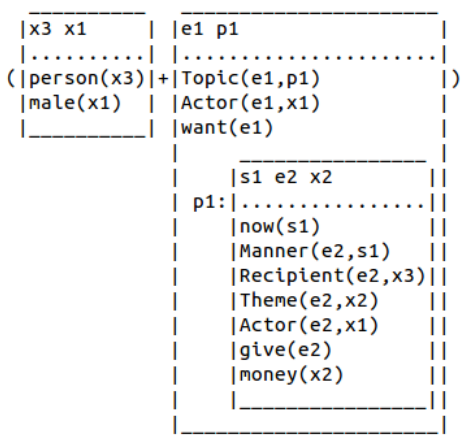


Fig. 16. Non-fronted example.

sentence final position, it receives the analysis in Fig. 16, in which it is analyzed as an adverb of manner, giving us a third analysis. Note that both versions of this sentence receive the same analysis in the event semantics, as shown in (19-d).

(19) a. The soldier told him to come back tomorrow.
b. Now he wants to give me money.
c. He wants to give me money now.
d. now(e1) $\wedge$ wants(e1,x1,give(e2,x1,x2,z1) $\wedge$ money(z1))

Furthermore, the representation of specific dates in the SYNCOIN data is not identified as a time referent by Boxer as seen in Fig. 17 for the sentence in (20).

(20) A meeting on 04/06/10 will work fine.

In the PROLOG representation, the date reference is marked as being a number, but not a time signature. Considering the importance of time in event semantics, the inconsistency in particular of temporal modifiers makes them a weak point in the conversion.

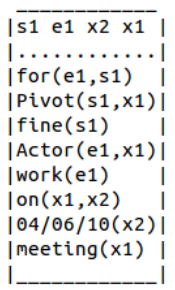


Fig. 17. Example of a date.

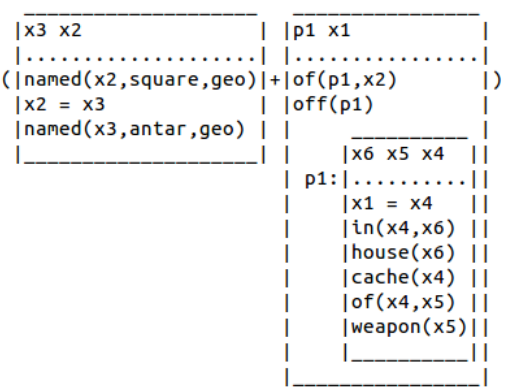


Fig. 18. Example of *off* as a preposition.

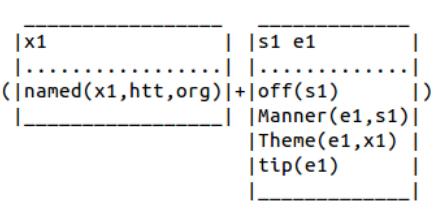


Fig. 19. Example of *off* in a phrasal verb.

3) Prepositions: In the event semantics, phrasal verbs are analyzed as multi-word expressions. Thus, in order to achieve a correct conversion from Boxer to event semantics, we need to be able to detect phrasal verbs as such. We show an example of a standard prepositional use of *off* and its use in a phrasal verb in (21). Their DRS analyses are shown in Fig. 18 and Fig. 19 respectively.

(21) a. There is a weapons cache in a house off of Antar Square.
b. HTT was tipped off.

In the DRS analyses, both usages of *off* are categorized similarly, once as a postmodifier, and once as a Manner relation, similar to adverbs. However, there is no indication that the latter is part of a phrasal verb. The corresponding analyses in event semantics are shown in (22).

(22) a. weapons_cache(y1) ∧ in(y1,y2) ∧ house(y2) ∧ off(y2,y3) ∧ named(y3,Antar Square)
b. HTT(x1) ∧ tip_off(e1,x1)

C. Coreference

Another major difference between Boxer's DRS and the event semantic representations is that the DRS is mostly an annotation on the sentential level while the event semantics also annotates discourse relations in form of coreference: Any coreferent entity in a text is referenced by the same variable. An example is shown in (23) where the *BCT patrol* is mentioned in two consecutive sentences, both times identified by variable `x2`, even though the surface representation is different.

(23) a. BCT patrol approached by man promising to reveal 2 additional weapons cache in Dour'a.
b. BCT reports little value in sites but pays man a small amount of cash
c. approach(e7,x4,x2) ∧ named(x2, BCT_patrol) ∧ man(x4) ∧ promise(e8,x4,e9) ∧ reveal(e9,y3) ∧ additional(y3) ∧ count(y3,2) ∧ weapons_cache(y3) ∧ in(y3,y10) ∧ named(y10,Dour'a))
d. named(x2,BCT) ∧ report(e10,x2,little_value(y3) ∧ sites(y3)) ∧ pay(e11,x2,z2) ∧ cash(z2) ∧ small_amount(z2)

The same also holds for event anaphora, for example in (24), where the *detonation* mentioned in the first sentence and the *attack* in the second sentence share the same event variable `e44`.

(24) a. Their description was passed to an Iraqi who subsequently apprehended them after a second failed attempt to detonate their satchel charge. . . .
b. The men detained for failed attack on 02/05/10 at the Soeudi Café, refuse to reveal operational details of the attack and deny being foreign insurgents.
c. pass(e41,x126,z10,x7) ∧ description(z10) ∧ possessive(x6,z10) ∧ iraqi(x7) ∧ apprehend(e42,x7,x6) ∧ after(e42,e43) ∧ attempt(e43,x6,e44) ∧ fail(e44) ∧ detonate(e44,x6,z11) ∧ satchel_charge(z11) ∧ possessive(x6,z11)
d. refuse(e50,x6,e51) ∧ men(x6) ∧ detain_for(e443,x129,x6,e44) ∧ fail(e44) ∧ attack(e44,x6) ∧ on(e44,t4) ∧ date(t4,020510) ∧ at(e44,y4) ∧ named(y4,Soeudi_Café) ∧ reveal(e51,x6,z14) ∧ details(z14,e44) ∧ operational(z14) ∧ deny(e52,x6,foreign(x6) ∧ insurgents(x6))

Boxer does have an option to perform coreference resolution using binding and accommodation theory to resolve the referents of pronouns and definite noun phrases. However, the

TABLE II
FREQUENCY OF PHENOMENA

Pivots	26
Units of measurement	3
Named entity referents	33
Possessive relationships / compounds	28
Adverbial modifiers	9
Temporal modifiers	6
Prepositions	7

module focuses on high precision, thus "definite descriptions and proper names are only linked to previous discourse referents if there is overlap in the DRS-conditions of the antecedent DRS ..." [26]. Additionally, it does not resolve event anaphora. Since, for our downstream application, the fusion algorithm, recall is extremely important, we will need a full coreference resolution integrated into our conversion procedure. This poses additional problems since especially event anaphora is an understudied process [27], [28].

D. Empirical Overview

We have looked at one of the threads in the SYNCOIN data in order to determine how frequent the phenomena are that we have discussed in the previous two sections. We have parsed those sentences using the C&C parser in combination with Boxer and then have inspected the resulting analyses manually to determine the frequency of the individual phenomena. The manual inspection was necessary since Boxer does not handle certain of those phenomena very well, as discussed in section VI. As a consequence of the necessary manual inspection, we chose the shortest thread, which consists of 21 sentences.

The distribution of phenomena is shown in table II. The numbers show that most of these phenomena occur with moderate frequency, on average in every third sentence. The exception are the named entities and the possessives, which occur on average more than once per sentence. These numbers show very clearly that the phenomena are frequent enough to necessitate a specialized treatment.

VII. FUTURE WORK

One of the major challenges in the conversion from Boxer DRS to event semantics is the underspecification and variance of specific phenomena in Boxer's DRS analyses. In order to ensure a fully automated high quality conversion, we will need to integrate tools and resources, along with machine learning algorithms to help resolve the ambiguities (e.g., to determine variable types). This includes the utilization of additional semantic resources such as PropBank [29] and full coreference to improve categorization. We will also explore the use of clustering to group frequently mislabeled words with words that most closely resemble their contextual behavior.

SYNCOIN data is full of infrequent words, particularly multi-word expressions of people and places, and given the nature of the data on which the C&C parser's models were trained, this makes it difficult to predict how an unknown word should be represented. Thus, we will investigate domain adaptation methods for all levels: parsing, named entity recognition, and DRS analysis. This is particularly necessary to resolve distinctions between proper names of people and proper names of places.

VIII. CONCLUSION

We have presented an investigation into the feasibility of converting from DRS to event semantics, demonstrating that it is a non-trivial task. We have started the conversion process using the PROLOG representation from Boxer to convert basic sentences via regular expressions that identify referents and relations to an event semantic representation. We have also highlighted areas that prove to be problematic in the conversion and require further exploration. We will expand the system beyond basic sentences and incorporate machine learning techniques and coreference resolution to increase the accuracy.

ACKNOWLEDGEMENT

This work is based on research supported by the U.S. Office of Naval Research (ONR) via grant #N00014-10-1-0140.

REFERENCES

[1] X. Carreras and L. Màrquez, "Introduction to the conll-2004 shared task: Semantic role labeling," in *Proceedings of the Eighth Conference on Computational Natural Language Learning (CoNLL-04)*, Boston, MA, 2004, pp. 89–97.

[2] X. Carreras and L. Màrquez, "Introduction to the CoNLL-2005 shared task: Semantic role labeling," in *Proceedings of the Ninth Conference on Computational Natural Language Learning (CoNLL-05)*, Ann Arbor, MI, 2005, pp. 152–164.

[3] L. Màrquez, X. Carreras, K. Litkowski, and S. Stevenson, "Semantic role labeling: An introduction to the special issue," *Computational Linguistics*, vol. 34, no. 2, pp. 145–159, 2008.

[4] T. Wickramarathne, K. Premaratne, M. Murthi, M. Scheutz, S. Kübler, and M. Pravia, "Belief theoretic methods for soft and hard data fusion," in *Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Prague, Czech Republic, 2011.

[5] R. Nunez, T. Wickramarathne, K. Premaratne, M. Murthi, S. Kübler, M. Scheutz, and M. Pravia, "Credibility assesment and inference for fusion of hard and soft information," in *Proceedings of the International Conference on Cross-Cultural Decision Making (HSBC FOCUS)*, San Francisco, CA, 2012.

[6] A. Cahill, M. McCarthy, J. van Genabith, and A. Way, "Quasi-logical forms from F-structures for the Penn Treebank," in *Proceedings of the Fifth International Workshop on Computational Semantics*, Tilburg, The Netherlands, 2003.

[7] C. Gardent and Y. Parmentier, "SemTAG: A platform for specifying Tree Adjoining Grammars and performing TAG-based semantic construction," in *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics*, Prague, Czech Republic, 2007, pp. 13–16.

[8] M. Steedman, *The Syntactic Process*. Cambridge, MA: MIT Press, 2001.

[9] S. Clark and J. Curran, "Formalism-independent parser evaluation with CCG and DepBank," in *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics (ACL)*, Prague, Czech Republic, 2007.

[10] ——, "Wide-coverage efficient statistical parsing with CCG and log-linear models," *Computational Linguistics*, vol. 33, no. 4, pp. 493—552, 2007.

[11] J. R. Curran, S. Clark, and J. Bos, "Linguistically motivated large-scale nlp with c&c and boxer," in *Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions*, ser. ACL '07. Prague, Czech Republic: ACL, 2007, pp. 33–36.

[12] J. Bos, "Wide-coverage semantic analysis with boxer," in *Semantics in Text Processing. STEP 2008 Conference Proceedings*, J. Bos and R. Delmonte, Eds. College Publications, 2008, pp. 277–286.

[13] ——, "Open-domain semantic parsing with boxer," in *Proceedings of the 20th Nordic Conference of Computational Linguistics (NODALIDA 2015)*, Vilnius, Lithuania, May 2015, pp. 301–304. [Online]. Available: http://www.let.rug.nl/bos/pubs/Bos2015NoDaLiDa.pdf

[14] H. Kamp, "A theory of truth and semantic representation," in *Formal Methods in the Study of Language*, J. Groenendijk, T. Janssen, and M. Stokhof, Eds. Amsterdam: Mathematical Centre, 1981, pp. 277–322.

[15] H. Kamp and U. Reyle, *From Discourse to Logic: An Introduction to Modeltheoretic Semantic of Natural Language, Formal Logic and DRT*. Dordrecht: Kluwer, 1993.

[16] L. Zettlemoyer and M. Collins, "Learning context-dependent mappings from sentences to logical form," in *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, Suntec, Singapore, 2009, pp. 976–984.

[17] Y. Artzi, K. Lee, and L. Zettlemoyer, "Broad-coverage CCG semantic parsing with AMR," in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, Lisbon, Portugal, September 2015, pp. 1699–1710.

[18] R. Dabarera, R. Nunez, K. Premaratne, and M. Murthi, "Dynamics of belief theoretic agent opinions under bounded confidence," in *International Conference on Information Fusion (FUSION)*, Salamanca, Spain, 2014.

[19] R. Nunez, M. Murthi, and K. Premaratne, "Efficient computation of DS-based uncertain logic operations and its application to hard and soft data fusion," in *International Conference on Information Fusion (FUSION)*, Salamanca, Spain, 2014.

[20] J. L. Graham, D. L. Hall, and J. Rimland, "A coin-inspired synthetic dataset for qualitative evaluation of hard and soft fusion systems," in *Proceedings of the 14th International Conference on Information Fusion (FUSION)*. Chicago, Illinois: IEEE, July 2011, pp. 1–8. [Online]. Available: https://www.researchgate.net/profile/David_Hall20/publication/261344900_A_COIN-inspired_synthetic_dataset_for_qualitative_evaluation_of_hard_and_soft_fusion_systems/links/542165450cf203f155c6693c.pdf

[21] D. Davidson, *Inquiries into Truth and Interpretation*. Oxford University Press, 1984.

[22] J. Hockenmaier and M. Steedman, "CCGbank: A corpus of CCG derivations and dependency structures extracted from the Penn Treebank," *Computational Linguistics*, vol. 33, no. 3, pp. 355–396, 2007.

[23] L. Hirschman and N. Chinchor, "MUC-7 coreference task definition," 1997, message Understanding Conference. [Online]. Available: http://www-nlpir.nist.gov/related_projects/muc/proceedings/co_task.html

[24] J. R. Curran and S. Clark, "Language independent ner using a maximum entropy tagger," in *Proceedings of CoNLL-03*, Edmonton, Canada, 2003, pp. 164–167.

[25] J. Bos and M. Nissim, "Uncovering noun-noun compound relations by gamification," in *Proceedings of the 20th Nordic Conference of Computational Linguistics (NODALIDA 2015)*, Vilnius, Lithuania, May 2015, pp. 251–255. [Online]. Available: http://www.let.rug.nl/bos/pubs/BosNissim2015NoDaLiDa.pdf

[26] J. Bos, "Implementing the binding and accommodation theory for anaphora resolution and presupposition projection," *Computational Linguistics*, vol. 29, no. 2, pp. 179–210, 2003.

[27] S. Pradhan, A. Moschitti, N. Xue, O. Uryupina, and Y. Zhang, "CoNLL-2012 shared task: Modeling multilingual unrestricted coreference in OntoNotes," in *Proceedings of the Sixteenth Conference on Computational Natural Language Learning (CoNLL 2012)*, Jeju, Korea, 2012.

[28] S. Pradhan, L. Ramshaw, M. Marcus, M. Palmer, R. Weischedel, and N. Xue, "CoNLL-2011 shared task: Modeling unrestricted coreference in OntoNotes," in *Proceedings of the Fifteenth Conference on Computational Natural Language Learning (CoNLL 2011)*, Portland, OR, 2011.

[29] M. Palmer, D. Gildea, and P. Kingsbury, "The proposition bank: An annotated corpus of semantic roles," *Computational Linguistics*, vol. 31, no. 1, pp. 71–106, 2005.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 353–362
DOI: 10.15439/2016F484
ACSIS, Vol. 8. ISSN 2300-5963

A connectionist approach to abductive problems: employing a learning algorithm

Andrzej Gajda, Adam Kups, Mariusz Urbański
Adam Mickiewicz University
in Poznań
ul. Szamarzewskiego 89AB, 60-568 Poznań, Poland
Email: {Andrzej.Gajda, adamku, Mariusz.Urbanski}@amu.edu.pl

***Abstract*—This paper presents preliminary results of an application of artificial neural networks and Backpropagation learning algorithm to solve logical abductive problems. To represent logic programs in the form of artificial neural networks CIL^{2}P approach proposed by Garcez et al. [3] is employed. Our abductive procedure makes use of translation of a logic program representing a knowledge base into a neural network, training of the neural network with an example representing an abductive goal and translation of the trained network back to the form of a logic program. An abductive hypothesis is represented as the symmetric difference between the initial logic program and the one obtained after training of the network. The first part of the paper introduces formal description of the tools used to model the abductive process, while the second part illustrates our contribution with results of a few computational experiments and discusses the ways of possible improvements of the proposed procedure.**

I. Introduction

ABDUCTION is a kind of reasoning which allows to *fill the gap* between a knowledge base Γ and a puzzling phenomenon ϕ, unattainable from Γ (cf. [6, 14]). We follow the algorithmic point of view, according to which an abductive hypothesis H "is legitimately dischargeable to the extent to which it makes it possible to prove (or compute) from a database a formula not provable (or computable) from it as it is currently structured" [2, p. 88]. A short characteristic of abduction interpreted along these lines can be found in [9]. Our goal in this paper is to use definite logic programs to formalise a class of abductive problems and to apply neural networks as a tool for abductive hypotheses generation.

Definite logic programs are characterised in the first order language [11]. However, in our approach we use only grounded definite logic programs [3]. Therefore, the formalisation of the abductive problems is restricted to the classical propositional logic.

The connectionist approach that we adopt makes use of positive partially recurrent one hidden-layer networks. This is sufficient to deal with definite logic programs as indicated in [3].

An abductive problem is stated for definite logic programs and the abductive hypotheses are generated by means of neural networks. Therefore, we use Connectionist Inductive Learning and Logic Programming System's (C-IL^2P) translation algorithm proposed by Garcez et al. [3] to translate logic programs into neural networks. The learning process of the neural network is aimed at changing it in such a way that the abductive goal is attained. Subsequently, the neural network is translated back into a logic program. Differences between the initial logic program and the one obtained from the trained neural network may be interpreted as abductive hypotheses generated in the learning process.

There are a few ways to model abduction using C-IL^2P. Two of them were described in [4] and they employ either a connectionist modal logic or an neural-symbolic system for abductive logic program. The approach that we want to present takes advantage of network learning process (using Backpropagation algorithm) and program-to-network and network-to-program translation algorithms. The main advantage of this approach in comparison with the two previous ones lies in its flexibility: abductive hypotheses do not have to be reduced to the form of propositional formulas (in particular conjunctions of atoms). At the present stage of our research we consider only abductive goals represented by atoms, but our approach allows for any form of an abductive goal (atoms or other types of formulas). Abduction here can be seen as a process of training of a network previously obtained from a logical program. The program represents a knowledge base and an abductive goal is represented by a training example. The reverse translation of the trained network allows to obtain abductive hypotheses. Using this approach abductive hypotheses can be represented by new logical formulas that are extending initial knowledge base. However, it is also possible that modifications or even removal of clauses from a knowledge base will represent abductive hypotheses (some similarity can be found in *contraction* and *revision* operations in belief revision theory, e.g. [5]).

In order to better comprehend the stated problem let us consider the following abductive problem. Suppose that we established a general medical facts[1]: "If people suffer from a fatigue and a fever, then they have a flu". "If people suffer from a fever, then they are fatigued". "If people have a flu, then they

Research reported in this paper were supported by the National Science Centre, Poland (DEC-2013/10/E/HS1/00172).

[1]Please, note that that the presented problem is only a simple toy-example, advanced medical decision support systems are for example presented in [13, 15]

are fatigued". Now, we would like to know what should we check if we suspect that someone has a flu. Formally, this problem can be described in the following way: let the knowledge base be a set of formulas $X = \{B \wedge C \rightarrow A, C \rightarrow B, A \rightarrow B\}$ and the abductive goal be a formula A, where A means a person has a flue, B means a person is fatigued and C means that a person has a fever. We shall go back to this example in every section.

II. Logic programs

Definite logic programs are usually formalised in the first order language. However, for the translation of definite logic program to neural network it is required for the program to be grounded. Therefore, grounded definite logic program consists only of predicates with constants as arguments. This is the reason for using a simpler language than the first order language. We characterize definite logic programs following [3] and similarly to [11].

We use language $\mathcal{L}$, which consists of the following elements:

- $\{a_1, a_2, \ldots\}$ — an infinite, countable set of propositional letters,
- $\leftarrow$ — a primitive connective,
- , — a comma.

Atomic formulas are denoted by propositional letters a_1, a_2,....

Definition 2.1: Let a_i for $0 \leq i \leq n$ be an atomic formula. *Horn clauses* are formulas of the form:

$$h_i = a_{i_0} \leftarrow a_{i_1}, \ldots, a_{i_n}.$$

The atomic formula a_{i_0} from the definition of the Horn clause is called the *head* of the Horn clause h_i and will be denoted as *head*(h_i). Similarly, the set of atomic formulas $\{a_{i_1}, \ldots, a_{i_n}\}$ forms the *body* of the Horn clause h_i and will be denoted as *body*(h_i). It is possible that the body of a Horn clause contains no atoms. Such Horn clauses will be called *facts*.

Atomic formulas and Horn clauses are the only well-formed formulas we use.

Definition 2.2: The *set of well-formed formulas* is defined in the following way:

$$\mathit{Form} = \{f \mid f = a_i \text{ or } f = h_j\}.$$

Definition 2.3: Let h_i for $1 \leq i \leq n$ be a Horn clause. A *definite logic program* is denoted by $\mathcal{P}$ and defined as follows:

$$\mathcal{P} = \{h_1, \ldots, h_n\}.$$

Definition 2.4: Let $\mathcal{P}$ be a definite logic program. We define the set of all atoms that occur in $\mathcal{P}$ in the following way:

$$B_\mathcal{P} = \{a_i \mid \text{for every } h_k \in \mathcal{P} \text{ and for every } a_j \in \mathit{body}(h_k)\text{: } a_i = \mathit{head}(h_k) \text{ or } a_i = a_j\}.$$

$B_\mathcal{P}$ is called Herbrand base of program $\mathcal{P}$.

We are now going to define *Least Herbrand model* of a definite logic program $\mathcal{P}$ which in turn will be used in the definition of logical consequence of a definite logic program $\mathcal{P}$.

Definition 2.5: A mapping v : *Form* $\mapsto$ {*true*,*false*} is a *valuation* defined as follows:

- for every atomic formula a_i: either $v(a_i) =$ *true* or $v(a_i) =$ *false*,
- for every Horn clause h_i: $v(h_i) =$ *true* iff $v(\mathit{head}(h_i)) =$ *true* or for at least one $a_j \in \mathit{body}(h_i)$: $v(a_j) =$ *false*.

Definition 2.6: Let $\mathcal{P}$ be a definite logic program, $B_\mathcal{P}$ a Herbrand base of $\mathcal{P}$ and v a valuation. An (Herbrand) *interpretation of a program* $\mathcal{P}$ w.r.t. v is a set $I_\mathcal{P}$ of all atoms in $B_\mathcal{P}$ that are mapped by v to *true*:

$$I_\mathcal{P}(v) = \{a_i \mid a_i \in B_\mathcal{P} \text{ and } v(a_i) = true\}.$$

Atoms that do not belong to the interpretation $I_\mathcal{P}$ are mapped to *false*.

Definition 2.7: Let $\mathcal{P}$ be definite logic program, h_i be a Horn clause that belongs to $\mathcal{P}$ and $I_\mathcal{P}$ a Herbrand interpretation of $\mathcal{P}$ w.r.t. valuation v. *Model of* $\mathcal{P}$ is defined as follows:

$$m_\mathcal{P} =_{df} I_\mathcal{P}(v) \text{ such that for every } h_i \in \mathcal{P}\text{: } v(h_i) = \mathit{true}.$$

It follows from the definition 2.7 that models of definite logic program $\mathcal{P}$ are of the form of sets of atoms. Therefore, we can establish a hierarchy over those models and use it to define the smallest model of $\mathcal{P}$.

Definition 2.8: Let S be a set. Function $c : S \rightarrow \mathbb{N}$ returns the number of elements in the set S.

Definition 2.9: Let $\mathcal{P}$ be a definite logic program and $m_\mathcal{P}$ a model of $\mathcal{P}$. By $m_\mathcal{P}^{min}$ we denote a *minimal model of* $\mathcal{P}$ and define it in the following way:

$$m_\mathcal{P}^{min} =_{df} m_\mathcal{P} \text{ such that for every } m'_\mathcal{P}\text{: } c(m_\mathcal{P}) \leq c(m'_\mathcal{P}).$$

Definition 2.10: Let $\mathcal{P}$ be a definite logic program and $m_\mathcal{P}$ a model of $\mathcal{P}$. By $M_\mathcal{P}$ we denote the *least Herbrand model of* $\mathcal{P}$ and defined it in the following way:

$$M_\mathcal{P} =_{df} m_\mathcal{P} \text{ such that for every } m'_\mathcal{P} \neq m_\mathcal{P}\text{: } c(m_\mathcal{P}) < c(m'_\mathcal{P}).$$

Now we are going to define Immediate Consequence Operator denoted as $T_\mathcal{P}$. It „provides the link between the declarative and procedural semantics of $\mathcal{P}$" [11, p. 37]. Related definitions concerning lattices are standard, therefore they are included in the Appendix.

Definition 2.11: Let $\mathcal{P}$ be a definite logic program and $I_\mathcal{P}$ an interpretation. The mapping $T_\mathcal{P} : 2^{B_\mathcal{P}} \mapsto 2^{B_\mathcal{P}}$ is defined as follows:

$$T_\mathcal{P}(I_\mathcal{P}) =_{df} \{\mathit{head}(h_i) \mid h_i \in \mathcal{P} \text{ and for all } a_j \in \mathit{body}(h_i)\text{: } a_j \in I_\mathcal{P}\}.$$

Proposition 2.1: Let $\mathcal{P}$ be a definite logic program. Then the mapping $T_\mathcal{P}$ is continuous.

Proposition 2.2: Let $\mathcal{P}$ be a definite logic program and $I_\mathcal{P}$ be an interpretation. Then $I_\mathcal{P}$ is a model for $\mathcal{P}$ iff $T_\mathcal{P}(I_\mathcal{P}) \subseteq I_\mathcal{P}$.

Theorem 2.3: Let $\mathcal{P}$ be a definite logic program. Then $M_\mathcal{P} = \mathit{lfp}(T_\mathcal{P}) = T_\mathcal{P} \uparrow \omega$.

Proofs of the propositions 2.1, 2.2 and theorem 2.3 are described in [11, p. 37–38].

Definition 2.12 (Logical consequence of $\mathcal{P}$): Let $\mathcal{P}$ be a definite logic program and $M_{\mathcal{P}}$ be a least Herbrand model of $\mathcal{P}$. Atom a_i is a logical consequence of $\mathcal{P}$ iff it belongs to $M_{\mathcal{P}}$:

$$\mathcal{P} \models a_i \text{ iff } a_i \in M_{\mathcal{P}}.$$

Definition 2.13: Let $\mathcal{P}$ be a definite logic program and $M_{\mathcal{P}}$ the least Herbrand model of $\mathcal{P}$. a_i is an atom and is called the *abductive goal* for $\mathcal{P}$ (denoted by $\mathcal{G}_{\mathcal{P}}$) if it fulfils the following criterion:

$$\mathcal{G}_{\mathcal{P}} =_{df} a_i \text{ such that } a_i \notin M_{\mathcal{P}}.$$

Coming back to the abductive problem given in the Introduction, observe that the knowledge base can expressed as a logic program $\mathcal{P} = \{A \leftarrow B, C; B \leftarrow C; B \leftarrow A\}$ and the least Herbrand model $M_{\mathcal{P}} = \emptyset$ (because $T_{\mathcal{P}} \uparrow (\emptyset) = \emptyset$) does not contain fact A, which is the abductive goal. This means that A is not entailed by $\mathcal{P}$.

III. Neural networks

Definition 3.1: We use the following language $\mathcal{L}^N$ to describe neural networks:

- $\{\mathfrak{i}_1, \ldots, \mathfrak{i}_n, \ldots\}$ — an infinite, countable set of symbols for input neurons,
- $\{\mathfrak{h}_1, \ldots, \mathfrak{h}_n, \ldots\}$ — an infinite, countable set of symbols for hidden layer neurons,
- $\{\mathfrak{o}_1, \ldots, \mathfrak{o}_n, \ldots\}$ — an infinite, countable set of symbols for output neurons,
- $\{\mathfrak{n}_1, \ldots, \mathfrak{n}_n, \ldots\}$ — an infinite, countable set of symbols for meta variables representing any neuron,
- $\{n_1, \ldots, n_n, \ldots\}$ — an infinite, countable set of symbols for *labels* of neurons which are not associated with an atom from $\mathcal{L}$,
- *label* — an identification symbol for a neuron,
- t — a label denoting *truth neuron*,
- tn — a label denoting hidden layer neuron reserved for truth neuron,
- $g(x)$, $h(x)$ — neuron activation functions,
- $x \in \mathbb{R}$ — a weighted sum of the input signals for the neuron,
- $A_{min} \in \mathbb{R}$ — a value computed by algorithm [3, p. 48–50],
- $A^f_{min} \in \mathbb{R}$ — A_{min} enlarge factor,
- $\theta_i \in \mathbb{R}$ — threshold of a neuron,
- $\theta_a \in \mathbb{R}$ — threshold of additional neurons,
- $\beta \in \mathbb{R}$ — steepness of neuron activation function (used only for bipolar semi-linear activation functions),
- $W \in \mathbb{R}$ — a weight computed by the algorithm [3, p. 48–50],
- $W^f \in \mathbb{R}$ — W enlarge factor,
- $r \in \mathbb{R}$ — a variable,
- $l \in \mathbb{N}$ — the number of additional hidden layer neurons per each neuron in the output layer.

Definition 3.2: Input neurons are tuples of the form:

$$\mathfrak{i}_i =_{df} \langle label, g(x), A^m_{min} \rangle$$

where:

- $label$ — a symbol of the represented atom form $\mathcal{P}$,
- $g(x) = x$,
- $A_{min} \in \mathbb{R}$ — if $g(x) \geq A_{min}$ then $label$ is mapped to *true*,
- the output signal of the neuron is equal to $g(x)$.

Definition 3.3: Truth neuron is an input neuron with the following properties:

- *label* $= t$,
- *g(x)* $= 1$.

Definition 3.4: Hidden layer neurons are tuples of the form:

$$\mathfrak{h}_i =_{df} \langle label, h(x), \theta_h, A^m_{min} \rangle$$

where:

- $label$ — n_i associated with clause h_i from $\mathcal{P}$,
- $h(x) = \frac{2}{1+e^{-\beta(x-\theta_h)}} - 1$,
- $\theta_h \in \mathbb{R}$,
- $A_{min} \in \mathbb{R}$ — if $h(x) \geq A_{min}$ then $label$ is mapped to *true*,
- the output signal of the neuron is equal to $h(x)$.

Definition 3.5: Output neurons are tuples of the form:

$$\mathfrak{o}_i =_{df} \langle label, h(x), \theta_o, A^m_{min} \rangle$$

where:

- $label$ — a symbol of the represented atom form $\mathcal{P}$,
- $h(x) = \frac{2}{1+e^{-\beta(x-\theta_o)}} - 1$,
- $\theta_o \in \mathbb{R}$,
- $A_{min} \in \mathbb{R}$ — if $g(x) \geq A_{min}$ then $label$ is mapped to *true*,
- the output signal of the neuron is equal to $h(x)$.

Definition 3.6: Let $\mathfrak{i}_1, \ldots, \mathfrak{i}_n$ be input neurons. By $\mathcal{N}_I$ we denote the set of all input neurons:

$$\mathcal{N}_I = \{\mathfrak{i}_1, \ldots, \mathfrak{i}_n\}.$$

Definition 3.7: Let $\mathfrak{h}_1, \ldots, \mathfrak{h}_n$ be input neurons. By $\mathcal{N}_H$ we denote the set of all hidden layer neurons:

$$\mathcal{N}_H = \{\mathfrak{h}_1, \ldots, \mathfrak{h}_n\}.$$

Definition 3.8: Let $\mathfrak{o}_1, \ldots, \mathfrak{o}_n$ be input neurons. By $\mathcal{N}_O$ we denote the set of all output neurons:

$$\mathcal{N}_O = \{\mathfrak{o}_1, \ldots, \mathfrak{o}_n\}.$$

Definition 3.9: Let $\mathcal{N}_I$, N_H and N_O be the set of all input, hidden and output neurons respectively. By $\mathcal{N}$ we denote the set of all neurons:

$$\mathcal{N} = \mathcal{N}_I \cup \mathcal{N}_H \cup \mathcal{N}_O.$$

Definition 3.10: Let $\mathfrak{i}_i$, $\mathfrak{i}_k$ and $\mathfrak{h}_j$, $\mathfrak{h}_l$ were input and hidden layer neurons respectively. By $\mathcal{C}_{i \to h}$ we denote the set of the connections from input to hidden layer neurons. The connection runs from the first neuron in the tuple to the second:

$$\mathcal{C}_{i \to h} = \{\langle \mathfrak{i}_i, \mathfrak{h}_j \rangle, \ldots, \langle \mathfrak{i}_k, \mathfrak{h}_l \rangle\}.$$

Definition 3.11: Let $\mathfrak{h}_i$, $\mathfrak{h}_k$ and $\mathfrak{o}_j$, $\mathfrak{o}_l$ were hidden layer and output neurons respectively. By $\mathcal{C}_{h \to o}$ we denote the set

of the connections from hidden layer to output neurons. The connection runs from the first neuron in the tuple to the second:

$$\mathcal{C}_{h\to o} = \{\langle \mathfrak{h}_i, \mathfrak{o}_j\rangle, \ldots, \langle \mathfrak{h}_k, \mathfrak{o}_l\rangle\}.$$

Definition 3.12: Let $\mathfrak{o}_i$ and $\mathfrak{i}_j$ were output and input neurons respectively. By $\mathcal{C}_r$ we denote the set of the recursive connections from output to input neurons. The connection runs from the first neuron in the tuple to the second:

$$\mathcal{C}_r = \{\langle \mathfrak{o}_i, \mathfrak{i}_j\rangle \mid \mathfrak{o}_i(label) = \mathfrak{i}_j(label)\}.$$

Definition 3.13: Let $\mathfrak{n}_i$, $\mathfrak{n}_k$, $\mathfrak{n}_j$, $\mathfrak{n}_l$ be neurons. By $\mathcal{C}_a$ we denote the set of the additional connections. The connection runs from the first neuron in the tuple to the second:

$$\mathcal{C}_a = \{\langle \mathfrak{n}_i, \mathfrak{n}_j\rangle, \ldots, \langle \mathfrak{n}_k, \mathfrak{n}_l\rangle\}.$$

Definition 3.14: $\mathcal{C}$ is the set of all connections in the network:

$$\mathcal{C} = \mathcal{C}_{i\to h} \cup \mathcal{C}_{h\to o} \cup \mathcal{C}_r \cup \mathcal{C}_a.$$

Definition 3.15: Let $\mathfrak{n}_i$, $\mathfrak{n}_j$ be neurons. The function $w : \mathcal{C} \to \mathbb{R}$ establishes the weight of the connection between two connected neurons:

- if $\langle \mathfrak{n}_i, \mathfrak{n}_j\rangle \in \mathcal{C}_r$ then $w(\langle \mathfrak{n}_i, \mathfrak{n}_j\rangle) = 1$,
- if $\langle \mathfrak{n}_i, \mathfrak{n}_j\rangle \in \mathcal{C}_a$ then $w(\langle \mathfrak{n}_i, \mathfrak{n}_j\rangle) \in [-r, 0) \cup (0, r]$,
- otherwise $w(\langle \mathfrak{n}_i, \mathfrak{n}_j\rangle) = W^m$.

Definition 3.16: Let $\mathfrak{n}_i$, $\mathfrak{n}_j$ be neurons. The set of all weights is denoted by $\mathcal{W}$. It consists of the tuple, where on the first and second place are connected neurons and on the third the weight of the connection:

$$\mathcal{W} = \{\langle \mathfrak{n}_i, \mathfrak{n}_j, w(\mathfrak{n}_i, \mathfrak{n}_j)\rangle \mid \langle \mathfrak{n}_i, \mathfrak{n}_j\rangle \in \mathcal{C}\}.$$

Definition 3.17: Neural network denoted by $\mathfrak{N}$ is a tuple:

$$\mathfrak{N} =_{df} \langle \mathcal{N}, \mathcal{C}, \mathcal{W}\rangle.$$

Definition 3.18: Let $\mathcal{P}$ be a definite logic program and $\mathcal{G}_\mathcal{P}$ an abductive goal. By $T_{\mathcal{P}\to\mathfrak{N}}$ we denote the translation from $\mathcal{P}$ to $\mathfrak{N}$ w.r.t. the set of predetermined factors $\{l, A^f_{min}, W^f, \beta, \theta_a, r\}$:

$$T_{\mathcal{P}\to\mathfrak{N}}(\langle \mathcal{P}, \{l, A^f_{min}, W^f, \beta, \theta_a, r\}\rangle) =_{df} \langle \mathcal{P}, \mathfrak{N}\rangle,$$

$\mathfrak{N}$ is obtained from $\mathcal{P}$ in the following way:

1) Calculate the following values by means of the algorithm [3, p. 48–50]:
   - A_{min},
   - W.
2) Calculate the following values:
   - $A^m_{min} = A_{min} + A^f_{min}$,
   - $W^m = W + W^f$.
3) For every atom $a_i \in B_\mathcal{P}$ an input neuron $\mathfrak{i}_i$ is added to the set of input neurons $\mathcal{N}_I$. Properties of each input neuron are the following:
   - *label* $= a_i$.
4) For every clause $h_i \in \mathcal{P}$, if *body*$(h_i) \neq \emptyset$ then a hidden layer neuron $\mathfrak{h}_i$ is added to the set of hidden layer neurons $\mathcal{N}_H$. Properties of each hidden layer neuron are the following:
   - *label* $= n_i$,
   - θ_o is computed as in the algorithm [3, p. 48–50].
5) For every atom in $a_i \in B_\mathcal{P}$ an output $\mathfrak{o}_i$ neuron is added to the set of output neurons $\mathcal{N}_O$. Properties of each output neuron are the following:
   - *label* $= a_i$,
   - if $a_i \in B^h_\mathcal{P}$ then θ_o is computed as in the algorithm [3, p. 48–50], otherwise $\theta_o = \theta_a$.
6) For $\mathcal{G}_\mathcal{P}$, *name* = '$g\mathcal{G}_\mathcal{P}$' (letter 'g' is added to the propositional letter assigned to atom $\mathcal{G}_\mathcal{P}$) and:
   - if $\mathcal{G}_\mathcal{P} \notin B_\mathcal{P}$ then:
     - an input neuron is added to the set of input neurons $\mathcal{N}_I$ where: *label* = *name*,
     - an output neuron is added to the set of output neurons $\mathcal{N}_O$ where: *label* = *name*, $\theta_o = \theta_a$,
   - otherwise:
     - *label* fields in neurons associated with $\mathcal{G}_\mathcal{P}$ in input and output set of neurons are changed to *name*.
7) Truth neuron *t* is added to the set of input neurons $\mathcal{N}_I$.
8) A hidden layer neuron is added to the set of hidden layer neurons $\mathcal{N}_H$ with the following properties:
   - *label* = *tn*,
   - $\theta_h = 0$.
9) For each neuron in the set of output neurons $\mathcal{N}_O$ add *l* additional neurons to the hidden layer with the properties (the overall number of the additional neurons is: $k = l \cdot c(\mathcal{N}_O)$):
   - *label* $= an_i$, where $i \in [1, k]$,
   - $\theta_h = 0$.
10) Generate the set of all neurons $\mathcal{N}$.
11) For every $h_i \in \mathcal{P}$:
   - for every $a_j \in$ *body*(h_i) add a tuple $\langle \mathfrak{i}_j, \mathfrak{h}_i\rangle$ to the set $\mathcal{C}_{i\to h}$, where $\mathfrak{i}_j(label) = a_j$,
   - for $a_k = head(h_i)$:
     - if *body*$(h_i) = \emptyset$ then add tuple $\langle \mathfrak{h}_t, \mathfrak{o}_k\rangle$ to the set $\mathcal{C}_{h\to o}$, where $\mathfrak{h}_t(label) =$ *tn* and $\mathfrak{o}_k(label) = a_k$,
     - otherwise add tuple $\langle \mathfrak{h}_i, \mathfrak{o}_k\rangle$ to the set $\mathcal{C}_{h\to o}$, where $\mathfrak{o}_k(label) = a_k$.
12) For every $\mathfrak{o}_i \in \mathcal{N}_O$ add tuples $\langle \mathfrak{h}_j, \mathfrak{o}_i\rangle$ to the set $\mathcal{C}_a$, where $\mathfrak{h}_j(label) = an_k$. Every output neuron $\mathfrak{o}_i$ should be connected with l additional hidden neurons that are assigned to it[2].
13) For every $\mathfrak{o}_i \in \mathcal{N}_O$: add tuples $\langle \mathfrak{h}_t, \mathfrak{o}_i\rangle$ to the set $\mathcal{C}_a$ if :
   - $\mathfrak{h}_t(label) =$ *tn* and
   - $\langle \mathfrak{h}_t, \mathfrak{o}_i\rangle \notin \mathcal{C}_{h\to o}$.

[2]For example: we have 2 output neurons and we set the number of additional neurons per output neuron $l = 2$. In this case we establish connections between the first two additional hidden layer neurons with the first output neuron, and the other two additional hidden layer neurons with the second output neuron.

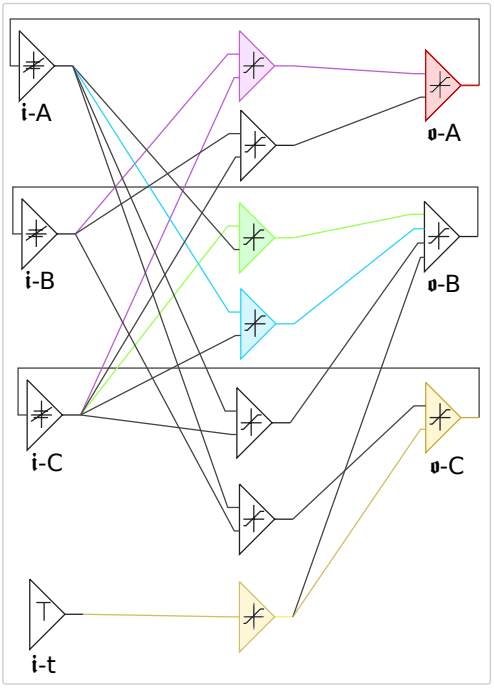


Fig. 1. A diagram of a neural network generated for the program in Example 1. The input neurons are labelled i-X, where X is a letter referring to an atom in the body of a clause, and output neurons are labelled o-Y, where Y refers to an atom in the head of a clause. The rest of the neurons are hidden layer units which represent possible clauses. Red color represents an abductive goal (fact A), purple color represents formula $A \leftarrow B, C$, green color represents formula $B \leftarrow C$, blue color represents formula $B \leftarrow A$. Yellow color represents the abductive hypothesis obtained for the problem (a fact C).

14) For every $\mathfrak{i}_i \in \mathcal{N}_I$, for every $\mathfrak{h}_j \in \mathcal{N}_H$: add tuple $\langle \mathfrak{i}_i, \mathfrak{h}_j \rangle$ to the set $\mathcal{C}_a$ if:
    - $\mathfrak{h}_j(\mathit{label}) \neq \mathit{tn}$ or
    - $\langle \mathfrak{h}_j, \mathfrak{o}_z \rangle \notin \mathcal{C}_{h \to o}$, where $\mathfrak{i}_i(\mathit{label}) = \mathfrak{o}_z(\mathit{label})$.
15) Generate the set of recursive connections $\mathcal{C}_r$.
16) Generate the set of all connections $\mathcal{C}$.
17) Generate the set of all weights $\mathcal{W}$.
18) Generate the neural network $\mathfrak{N}$.

The result of the application of the above procedure to the logic program obtained from the exemplary problem is depicted in Fig. 1.

Definition 3.19: Let $\mathfrak{N}$ be a neural network. By $T_{\mathfrak{N} \to \mathcal{P}}$ we denote the translation from $\mathfrak{N}$ to $\mathcal{P}$:

$$T_{\mathfrak{N} \to \mathcal{P}}(\mathfrak{N}) =_{df} \langle \mathfrak{N}, \mathcal{P} \rangle,$$

where $\mathcal{P}$ is obtained from $\mathfrak{N}$ by using the pedagogical approach described in [3, Chapter 5].

Using the naive version of pedagogical approach we simply map input neurons into output neurons in the following way: for each output neuron we check all combinations of input values for all input neurons relevant for this output neuron. In the case of the activation of the considered output neuron we associate the atom represented by this neuron with the following clause h_i: $\mathit{head}(h_i)$ is the considered atom and the body consists of the atoms represented by input neurons which were set to 1.0 and the negated atoms represented by input neurons that were set to -1.0. For example for output neuron with $\mathit{label} = C$ with only one relevant input neuron with $\mathit{label} = A$, if it is the case that for the considered input neuron $g(x) = 1.0$ and for the output $h(x) > A_{min}$ then we obtain the clause $C \leftarrow A$; analogically, if it is the case that for the considered input neuron $g(x) = -1.0$ and for the output $h(x) > A_{min}$ then we obtain the clause $C \leftarrow nA$ (where nA means negated A). The set of clauses extracted from a neural network is minimized by means of the the Quine-McCluskey algorithm [7, 12].

The results of learning and translation of the running example are presented in Section VI-A.

IV. Abduction

Traditionally abduction can be seen as a process of searching for explanations of a problem during which *additional* information is obtained (see [1, p. 74] on abductive explanatory characterization styles). Formally speaking, whenever we have some knowledge base K and abductive goal A , the abduction leads to generation of some additional formulas $h_1, \ldots, h_n$, that are not present in K, but which together with formulas in K enable derivation of A. In the approach presented in this paper, this does not necessarily has to be the case. It is possible that after a training of the network which represents logical program/knowledge base with an example that represents an abductive goal, translation of the trained network back to the form of logical clauses will change the initial program in other ways than just extending it. It may happen that some formulas will be modified (e.g. by removing or adding some atoms in the body of some clauses) or they can be even removed from the initial program (it is probably a matter of discussion whether it is a desirable phenomenon or not). In our approach each such modification can be seen as an abductive hypothesis, hence the term *hypothesis* gain somewhat dynamic character. This approach is more flexible than the traditional one and allows more interesting conceptual applications for the abduction, in particular accomodating substantial revisions of the initial knowledge base (see abductive schematics in [2, p. 47]). The general scheme of the proposed procedure is depicted in Fig. 2.

The abductive procedure begins in the left upper corner of the scheme presented in Fig. 2. The knowledge base is represented by the definite logic program $\mathcal{P}$ and there is a fact ϕ which cannot be derived from the knowledge base. The fact ϕ is of the form of an atom a_i and the abductive problem is represented by the fact that $a_i \notin M_{\mathcal{P}}$.

The first step of the abductive procedure is the translation of the $\mathcal{P}$ to a neural network $\mathfrak{N}$. The first step of the translation is obtained by means of the algorithm developed by Garcez in [3, p. 49]. The difference is that we add all atoms from the $B_{\mathcal{P}}$ along with the a_i to the input ($\mathcal{N}_I$) and output ($\mathcal{N}_O$) layer of the $\mathfrak{N}$ (steps 3 and 5 in $T_{\mathcal{P} \to \mathfrak{N}}$). In case of absence of the *facts* in $\mathcal{P}$, we also add a truth neuron t (which gives always 1 on the output) to $\mathcal{N}_I$ (step 7 in $T_{\mathcal{P} \to \mathfrak{N}}$). The hidden layer

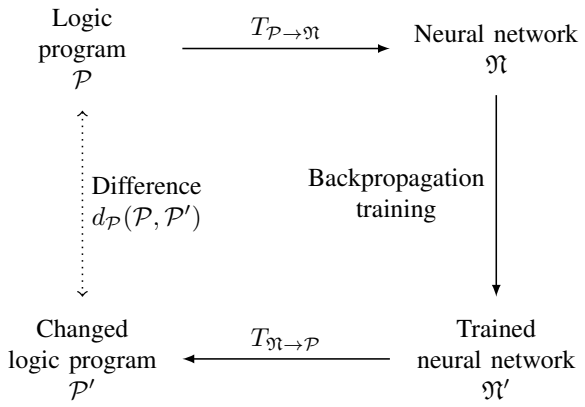


Fig. 2. A scheme of the abductive procedure

($\mathcal{N}_H$) is modified in the following way: we add *tn* neuron, which is reserved only for *t* neuron from $\mathcal{N}_I$, and a number of additional atoms per clause and a_i (steps 8 and 9 in $T_{\mathcal{P}\to\mathfrak{N}}$).

The set of all connections from input to hidden layer ($\mathcal{C}_{i\to h}$) is enriched by the additional connections running from every atom in $\mathcal{N}_I$ to $\mathcal{N}_H$ with the exception of *t* and *tn* neurons. There is a unique connection between *t* and *tn*. Additional connections do not double the connections obtained by the previous step. We also forbid additional connections from input to the hidden layer neurons that are connected with an output neuron with the same label as the neuron from input layer. In other words, additional connections cannot allow to formulate formulas of the form $A \leftarrow A$. There are also additional connections added to the set of connections from hidden to output layer ($\mathcal{C}_{h\to o}$). Those connections run from additional neurons assigned to the particular clause from $\mathcal{P}$ to the neuron which represents the head of the concerned clause. The neuron *tn* from the hidden layer connects with every neuron from output layer, except for the neuron which represents the fact ϕ denoted by a_i. The function w gives every additional connection a random value from the range $[-r, r]$ (with the exception of 0).

After the whole neural network $\mathfrak{N}$ is completed, the training begins with the use of the standard backpropagation algorithm. The training set consists of only one element: a table with all input atoms set to -1 and all output atoms set to 1. Error calculation is performed after $\mathfrak{N}$ achieves the stable state.

Trained neural network $\mathfrak{N}'$ is then translated back to logic program $\mathcal{P}'$. We define the difference between $\mathcal{P}$ and $\mathcal{P}'$ as $d_{\mathcal{P}}(\mathcal{P}, \mathcal{P}')$.

Definition 4.1: Let $\mathcal{P}$ and $\mathcal{P}'$ were a definite logic programs. The difference between $\mathcal{P}$ and $\mathcal{P}'$ is denoted by $d_{\mathcal{P}}(\mathcal{P}, \mathcal{P}')$:

$$d_{\mathcal{P}}(\mathcal{P}, \mathcal{P}') = (\mathcal{P}' \setminus (\mathcal{P} \cap \mathcal{P}')) \cup (\mathcal{P} \setminus (\mathcal{P} \cap \mathcal{P}')).$$

The change of the logic program defined as a symmetric difference is interpreted as a set of abductive hypotheses.

Definition 4.2: Let $\mathcal{P}$ be a definite logic program and $\mathfrak{N}$ a neural network obtained from $\mathcal{P}$ by the translation $T_{\mathcal{P}\to\mathfrak{N}}$. Let us further assume that $\mathfrak{N}'$ is obtained from $\mathfrak{N}$ by backpropagarion training described above. After translation of the $\mathfrak{N}'$ to $\mathcal{P}'$ by translation $T_{\mathfrak{N}\to\mathcal{P}}$ the set of abductive hypotheses $H_{\mathcal{P}}$ can be defined in the following way:

$$H_{\mathcal{P}} = d_{\mathcal{P}}(\mathcal{P}, \mathcal{P}').$$

V. Implementation

To implement the ideas given above, we have decided to use Framsticks software [10] — a versatile tool which among its many merits, gives the possibility to perform computational experiments concerning artificial neural networks. Framsticks platform is equipped with an advanced scripting language that easily enables any kind of experiment. This, together with an advanced network simulator, makes it a suitable tool for the research presented in our article. Apart from that, the software was already used in computational experiments concerning logical abduction [8, 9].

The whole implementation can be seen as a general framework consisting of scripts representing operations described earlier in the text. Hence, the whole abduction experiment can be represented as the knowledge flow between different scripts which is schematically depicted in Fig. 2.

The first important part implemented is the algorithm of translation of the default logic programs into one-hidden layer neural networks described in Sect. IV.

Next, Backpropagation algorithm with momentum has been programmed as it was not available in Framsticks software that is mainly focused on evolutionary optimisation. Finally, the algorithm translating (trained) neural networks into logic programs, described in Sect. IV has been also implemented. The whole learning procedure, however, has been modified (in comparison to the standard application of Backpropagation algorithm) and adapted to the needs of the presented research. The training set for any abductive problem consists of only one training example which in its input part contains only -1.0 values. This represents the situation where the operator $T_{\mathcal{P}}$ starts from the bottom of the lattice of all possible interpretations of $\mathcal{P}$. The values in the output part are now selected arbitrarily (apart from the value related to the neuron representing an abductive goal, which is always set to 1.0), but as mentioned later in Sect. VII solving this issue is one of the immediate future tasks. The scheme of learning is also modified as each change of the weights is based on the error calculation performed after the network achieves stable state (after several cycles of signal propagation). The network-to-program translation is implemented using the brute force approach (sometimes called *pedagogical*, which is of exponential complexity with respect to the number of input neurons), that is inefficient, especially for more complicated problems. However, as complexity-reducing approaches may be associated with lack of completeness and soundness of the translation [3, Chapter 5] and a general view on the abductive process is needed at the moment, we have decided to temporarily pay the price of the computational cost. The future plans include reduction of complexity of the used methods. Yet another issue related to the network-to-program

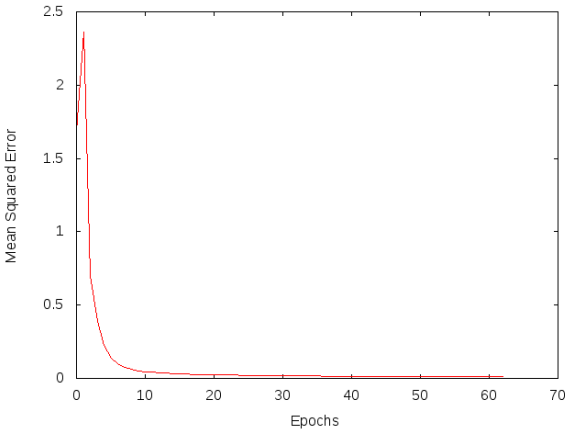


Fig. 3. A learning error chart for the Example 1.

translation is that it demands several stages of processing of the obtained clauses. In the present approach the first stage of translation may result in obtaining redundant information (e.g. two formulas of the form $A \leftarrow B, \neg C$ and $A \leftarrow B, C$ can be obtained, which should be reduced to $A \leftarrow B$).

VI. Results

This section presents preliminary results for two toy-example problems. The first example is presented in more detailed way to demonstrate how the procedure works and it concerns the problem provided in the introduction.

A. Example 1

The knowledge base for the Example 1 is $\mathcal{P} = \{A \leftarrow B, C; B \leftarrow C; B \leftarrow A\}$ and the abductive goal is a fact A. According to the procedure described in Sect. IV, the network obtained after translation of $\mathcal{P}$ is extended to contain additional hidden-layer nodes and connections to enable generation of new concepts in the knowledge base. To perform learning we prepared a training example in which every input neuron sends a signal of value -1.0 and every output value should be equal 1.0.

The learning procedure is a modified version of a Back-propagation algorithm, as mentioned earlier and in Sect. IV to let the network compute the fix point of the program.

The learning error is presented in Fig. 3. As it can be seen the error was minimized rather quickly (after a bit more than 60 epochs), which is not surprising as the training set contained only one example.

The translation of the network back to the form of a logical program resulted in the following set of formulas $\mathcal{P}_t = \{A \leftarrow B, C; B \leftarrow nC, A, t; B \leftarrow C, nA, t; B \leftarrow C, A, t; C \leftarrow nA, nB, t; C \leftarrow nA, B, t; C \leftarrow A, nB, t; C \leftarrow A, B, t\}$, where t stand for truth and n is a negation (which at this stage may be interpreted as negation as a failure — however it is removed during reduction of the obtained logic program). Further reduction of the obtained set of formulas resulted in the set $\mathcal{P}_{tr} = \{A \leftarrow B, C; B \leftarrow C; B \leftarrow A; C\}$. This means that the initial knowledge base was extended by a fact C, which is an abductive hypothesis for the problem.

B. Example 2

Example 2 demonstrates an effect of increase of the number of hidden neurons (which represent the potential concepts that can be learned by a network, see Sec. IV) and the role of the initial state of the network (represented by randomly initiated weights of the additional connections). In order to preliminarily examine these effects we performed the abductive procedure several times with other parameters fixed. The knowledge base is $\mathcal{P} = \{A \leftarrow B; A \leftarrow C; D \leftarrow E; C \leftarrow E; B \leftarrow C\}$ and an abductive goal is a fact A. The training example in the output part contained only 1.0 values. The most frequent program obtained from a network containing *one* additional hidden neuron per each output neuron is $\mathcal{P}_{t1} = \{A \leftarrow B; A \leftarrow C; D \leftarrow E; C \leftarrow E; B \leftarrow C; E\}$ — which means that the abductive hypotheses is E. The most frequent definite logic[3] program obtained from the trained networks with 17 additional hidden-layer neurons per output neuron is $\mathcal{P}_{t2} = \{A \leftarrow B; A \leftarrow C; D; C; B \leftarrow C; E\}$. This means that the abductive hypothesis is: E and change of formulas $D \leftarrow E$ and $C \leftarrow E$ into respectively D and C. This example illustrates two phenomenons. Firstly, the knowledge base can be modified during the learning process: for example clause $D \leftarrow E$ has been strengthened into a fact D. Secondly, the number of hidden-layer neurons can influence the obtained solutions for abductive problems. These two observations in turn mean, that a size of hidden layer represents a "reasoning potential" of a network and may possibly be used later to control the quality of obtained hypotheses. In the presented exemplary result, a greater number of the hidden neurons resulted in the lowered quality of the obtained hypothesis: this is because, D does not help with the derivation of A and the hypothesis contains redundancy as A can be derived from both E and C. Moreover, C can be derived from E. This lowers the internal minimality of the hypothesis, understood as non-derivability of one of the subformula of the hypothesis from another. Note, however, that the increased number of neurons is not completely unpromising, as the best solution (obtained most frequently for networks with the lower number of hidden neurons) also appeared in the set of possible solutions, but much less often.

Interestingly, further analyses revealed one more interesting issue related to the increased number of hidden neurons and connections. Apparently, in the larger networks the abductive processes became more sensitive to the initial state of the connections (with randomly initialized weights). The repetitions of the abductive experiments resulted in establishing different abductive hypotheses and this diversity was bigger than in the case of the network with lower number of neurons. The knowledge bases obtained from the subsequently trained networks with the larger number of neurons were represented i.a. by the

[3]In this work we purposely left out of analysis other kinds of logic programs, however the most frequent program obtained with these set of parameters was actually a general logic program.

following sets of formulas: $\{A \leftarrow B; A \leftarrow C; D \leftarrow E; C \leftarrow E; B; E\}$, $\{A \leftarrow B; A \leftarrow C; D \rightarrow E; C; B; E\}$. These few results preliminarily demonstrate that random initialization of the network which is capable of acquiring new concepts (due to the additional hidden neurons and connections), may influence the degree of modification of the initial knowledge and the quality of the abductive hypotheses.

VII. Summary and future work

In this paper we presented an attempt at synthesizing formal description of abductive problems with connectionist methodology based on work of Garcez et al. [3]. Contrary to the received approaches to the problem of logical abduction, we decided to employ Backpropagation algorithm to search for abductive hypotheses. The results of such experiments show that training of an artificial neural network can indeed be a method of filling a gap between a knowledge base and an abductive goal.

The neural networks obtained according to the presented procedure can be further used to solve real life problems as classifiers (just like traditional artificial neural networks are usually used) – examples of such applications of $C\text{-}IL^2P$ are presented in [3, Chapter 4]. Apart from that, the trained networks can be used as massively parallel deduction systems to solve logical problems. Yet another advantage of the presented approach, which at the same time distinguishes it from often discussed, pure logical accounts, is that it offers broad and flexible definition of the abductive hypothesis. The obtained hypotheses can either be additional formulas extending the initial knowledge base or some modifications done to the knowledge base. Finally, such methodology can be used for modelling abduction as a cognitive process, by combining connectionist structures, resembling in the limited sense the human brain, with rigours of the formal logic. The flexibility of the concept of the abductive hypothesis is likely to increase the accuracy of such modelling.

One of the interesting observations obtained at this preliminary stage of the currently presented research is the joint influence of the size of the trained network and its initial state on the obtained abductive hypotheses. While such observations may seem a little vague, more advanced computational experiments and quantitative analyses are currently under way. Among them, interesting research task is examination of the influence of the different parameters of the networks (i.e. biases of the additional neurons, initial weights of the additional connections, etc.) and parameters of training process (e.g. a form of the training example, value of learning rate or momentum) on the resulting knowledge. Research on the influence of interaction of these parameters with characteristics of the considered problems (e.g. a number of clauses in the knowledge base, a number of propositional variables, some more sophisticated structural traits) on the resulting knowledge (understood as a logic program obtained from the trained neural network) is also possible. This kind of research demands development of some measures of similarity between different knowledge bases, which is an interesting issue in itself and gives possibility of performing quantitative multi-dimensional analyses of the results of abductive process.

In the nearest future a number of improvements to the existing implementations are also in order. Among them, the most pending ones are: a decrease of computational complexity of the network-to-program implementation and intelligent automatic translation of an abductive goal into the form of a training example. This encompasses formulation of the training example in such a way that knowledge base is properly represented.

The further plans concern extension of our methodology to include more types of logic programs in order to introduce a negation, default or classical, into the researched abductive problems. The introduction of a negation will require even more sophisticated approach to generation of training examples containing representations of abductive goals. Additionally, as it is straightforward for our methodology, we would like to introduce the abductive goals in the form of formulas more complex than single atoms.

To solve the issue concerning generation of training examples a computational approach may be employed. The computational experiments may concern analysis of the impact of the desirable output values on the obtained modifications of the knowledge base. The potential results may shed light on how to construct an initial training example to properly reflect an abductive goal alongside desired concepts from the knowledge base.

Finally, a well-established abduction research requires employment of some quality-control mechanisms to obtain efficient abductive hypotheses. This goal is partially achieved in the presented approach heuristically, by removal of some unwanted connections, careful generation of the training examples and (possibly) by imposing restrictions on the size of the trained network. However, employment of some more sophisticated quality criteria is still a pending issue. An interesting approach concerning application of multi-criteria dominance relation [8, 9] to evaluate already generated abductive hypotheses gives the possibility to perform a comparative study with the methodology presented in this paper.

Appendix

Definition 7.1: A binary relation R on a set is called a *partial order* when it is reflexive, transitive and antisymmetric.

We will denote a relation which is a partial order by $\preceq$. A set S with a partial order $\preceq$ is called a *partially ordered set*.

Definition 7.2: Let S be a set with a partial order $\preceq$ and $x \in S$. We define the following:

- x is a *minimum* of S if for all elements $y \in S$: $x \preceq y$.
- x is a *maximum* of S if for all elements $y \in S$: $y \preceq x$.

Minimum and maximum of a set S are unique, if they exist, and will be denoted as *inf*(S) and *sup*(S) respectively.

Definition 7.3: Let S be a set with a partial order $\preceq$ and $R \subseteq S$. We define the following:

- An element $x \in S$ is an *upper bound* of R if for all elements $y \in R$: $y \preceq x$.

- An element $x \in S$ is an *lower bound* of R if for all elements $y \in R$: $x \preceq y$.
- An element $x \in S$ is *least upper bound* of R if x is an upper bound of R and $x \preceq z$ for all upper bounds z of R.
- An element $x \in S$ is *greatest lower bound* of R if x is an lower bound of R and $z \preceq x$ for all lower bounds z of R.

The least upper bound and the greatest lower bound of a set R are unique, if they exist, and will be denoted as $lub(R)$ and $glb(R)$ respectively.

Definition 7.4: Let S be a partially ordered set and $R \subseteq S$. We call S a *complete lattice* if $lub(R)$ and $glb(R)$ exist for every $R \subseteq S$.

Definition 7.5: Let S be a complete lattice and $R \subseteq S$. We call R *directed* if every finite subset of R has an upper bound in R.

Definition 7.6: Let S be a complete lattice, x and y be elements of S, and T: $S \to S$ be a mapping. The following holds:

- T is *monotonic* if $T(x) \preceq T(y)$, where $x \preceq y$.
- T is *continuous* if for every directed subset R of S: $T(lub(R)) = lub(T(R))$.

The collection of all Herbrand interpretations of a definite logic program P, which is $2^{B_\mathcal{P}}$, forms a complete lattice under the partial order of set inclusion. The top and the bottom element of $2^{B_\mathcal{P}}$ is $B_\mathcal{P}$ and $\emptyset$ respectively. There can be described a continuous and monotonic mapping from $2^{B_\mathcal{P}}$ to $2^{B_\mathcal{P}}$ which will serve in finding the least Herbrand model of P. The procedure is based on fixpoints of mappings on the set $2^{B_\mathcal{P}}$.

Definition 7.7: Let S be a complete lattice and $T : S \to S$ be a mapping. An element $x \in S$ is the *least fixpoint* of T if x is a fixpoint of T (*i.e.* $T(x) = x$) and for all fixpoints y of T: $x \preceq y$. An element $x \in S$ is the *greatest fixpoint* of T if x is a fixpoint of T and for all fixpoints y of T: $y \preceq x$.

The least fixpoint of T and the greatest fixpoint of T will be denoted as $lfp(T)$ and $gfp(T)$ respectively.

Proposition 7.1: Let S be a complete lattice and $T : S \to S$ be monotonic. T has $lfp(T)$ and $gfp(T)$.

The proof of the theorem 7.1 is described in [11, p. 28].

Now we want to define *ordinal powers of* T. The definition is based on properties of ordinal numbers described in [11, p. 28–29] or [3, p. 26].

Definition 7.8: Let S be a complete lattice and $T : S \to S$ be monotonic. Then we define:
$T \uparrow 0 = inf(S)$;
$T \uparrow \alpha = T(T \uparrow (\alpha - 1))$, if α is a successor ordinal;
$T \uparrow \alpha = lub(T \uparrow \beta \mid \beta \prec \alpha)$, if α is a limit ordinal;
$T \downarrow 0 = sup(S)$;
$T \downarrow \alpha = T(T \downarrow (\alpha - 1))$, if α is a successor ordinal;
$T \downarrow \alpha = glb(T \downarrow \beta \mid \beta \prec \alpha)$, if α is a limit ordinal.

Proposition 7.2: Let S be a complete lattice and $T : S \to S$ be continuous. Then $lfp(T) = T \uparrow \omega$.

The ω in theorem 7.2 denotes the first infinite ordinal. Proof of the theorem 7.2 is described in details in [11, p. 30].

REFERENCES

[1] Atocha Aliseda. *Abductive Reasoning. Logical Investigations into Discovery and Explanation*. Springer, Dordrecht, 2006. doi: 10.1007/1-4020-3907-7.

[2] Dov M. Gabbay and John Woods. *The Reach of Abduction. Insight and Trial*. Elsevier, 2005. doi: 10.1016/S1874-5075(05)80034-8.

[3] Artur S. d'Avila Garcez, Krysia Broda, and Dov M. Gabbay. *Neural-symbolic learning systems: foundations and applications*. Springer Science & Business Media, 2002. doi: 10.1007/978-1-4471-0211-3.

[4] Artur S d'Avila Garcez, Dov M Gabbay, Oliver Ray, and John Woods. Abductive reasoning in neural-symbolic systems. *Topoi*, 26(1):37–49, 2007. doi: 10.1007/s11245-006-9005-5.

[5] Peter Gärdenfors. *Belief Revision*. Tracts in Theoretical Computer Science 29. Cambridge University Press, 2003. doi: 10.1017/CBO9780511526664.

[6] Jaakko Hintikka. Abduction — inference, conjecture, or an answer to a question? In *Socratic Epistemology. Explorations of Knowledge-Seeking by Questioning*, pages 38–60. Cambridge University Press, 2007. doi: 10.1017/CBO9780511619298.003.

[7] Tarun Kumar Jain, Dharmender Singh Kushwaha, and Arun Kumar Misra. Optimization of the quine-mccluskey method for the minimization of the boolean expressions. In *Fourth International Conference on Autonomic and Autonomous Systems (ICAS'08)*, pages 165–168. IEEE, 2008. doi: 10.1109/ICAS.2008.11.

[8] M. Komosinski, A. Kups, and M. Urbański. Multi-criteria evaluation of abductive hypotheses: towards efficient optimization in proof theory. In *Proceedings of the 18th International Conference on Soft Computing*, pages 320–325, Brno, Czech Republic, 2012.

[9] M. Komosinski, A. Kups, D. Leszczyńska-Jasion, and M. Urbański. Identifying efficient abductive hypotheses using multi-criteria dominance relation. *ACM Transactions on Computational Logic*, 15(4), 2014. doi: 10.1145/2629669.

[10] Maciej Komosinski and Szymon Ulatowski. Framsticks web site, 2016. http://www.framsticks.com.

[11] John Wylie Lloyd. Foundations of logic programming. 1993. doi: 10.1007/978-3-642-83189-8.

[12] Edward J McCluskey. Minimization of boolean functions. *Bell system technical Journal*, 35(6):1417–1444, 1956. doi: 10.1002/j.1538-7305.1956.tb03835.x.

[13] Noel Pérez, Miguel Angel Guevara, Augusto Silva, Isabel Ramos, and Joana Loureiro. Improving the performance of machine learning classifiers for breast cancer diagnosis based on feature selection. In M. Paprzycki M. Ganzha, L. Maciaszek, editor, *Proceedings of the 2014 Federated Conference on Computer Science and Information Systems*, volume 2 of *Annals of Computer Science and Information Systems*, pages 209–217. IEEE, 2014. doi: 10.15439/2014F249. URL http://dx.doi.org/

10.15439/2014F249.

[14] P. Thagard. Abductive inference: From philosophical analysis to neural mechanisms. In A. Feeney and E. Heit, editors, *Inductive reasoning: Cognitive, mathematical, and neuroscientific approaches*, pages 226–247. Cambridge University Press, Cambridge, 2007. doi: 10.1017/cbo9780511619304.010.

[15] Agnieszka Wosiak and Danuta Zakrzewska. On integrating clustering and statistical analysis for supporting cardiovascular disease diagnosis. In M. Ganzha, L. Maciaszek, and M. Paprzycki, editors, *Proceedings of the 2015 Federated Conference on Computer Science and Information Systems*, volume 5 of *Annals of Computer Science and Information Systems*, pages 303–310. IEEE, 2015. doi: 10.15439/2015F151. URL http://dx.doi.org/10.15439/2015F151.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 363–371
DOI: 10.15439/2016F520
ACSIS, Vol. 8. ISSN 2300-5963

On Algebraic Hierarchies in Mathematical Repository of Mizar

Adam Grabowski, Artur Korniłowicz
Institute of Informatics
University of Białystok
ul. Ciołkowskiego 1 M, 15-245 Białystok, Poland
Email: {adam, arturk}@math.uwb.edu.pl

Christoph Schwarzweller
Department of Computer Science
University of Gdańsk
Wita Stwosza 57, 80-952 Gdańsk, Poland
Email: schwarzw@inf.ug.edu.pl

***Abstract*—Mathematics, especially algebra, uses plenty of structures: groups, rings, integral domains, fields, vector spaces to name a few of the most basic ones. Classes of structures are closely connected – usually by inclusion – naturally leading to hierarchies that has been reproduced in different forms in different mathematical repositories. In this paper we give a brief overview of some existing algebraic hierarchies and report on the latest developments in the Mizar computerized proof assistant system. In particular we present a detailed algebraic hierarchy that has been defined in Mizar and discuss extensions of the hierarchy towards more involved domains. Taking fully formal approach into account we meet new difficulties comparing with its informal mathematical framework.**

I. Introduction

SINCE its development at the beginning of the 20th century abstract algebra has spread over to various branches of mathematics. One reason is the highly reusable results produced, not least due to the hierarchical structure of algebraic domains. This kind of reuse, of course, is also highly desirable in mathematical proof assistants. Consequently one naturally finds various systems in which algebraic hierarchies similar to abstract algebra have been constructed. However, most of them served to facilitate the formalization of a particular theorem or a particular application:

Though not in a proof assistant, but in a computer algebra system the – to the best of our knowledge – first algebraic hierarchy was constructed in **Axiom** [24]. Started back in 1978 – the first release under the name Axiom took place in 1991 – this was the first system in which types were connected to mathematical domains: Algebraic domains have types in their own right – called categories – that can be used to form hierarchies. So, for example, it is possible to define an operation `Fraction` that takes an argument of type `IntegralDomain` and returns its field of fractions. The algebraic hierarchy of **Nuprl** [23] was developed to support computational abstract algebra. In **Coq** [9] more than one algebraic hierarchy exists, we name two of them: One [11] was constructed as part of the FTA project to prove the fundamental theorem of algebra, another one was used in the formalization of the Feit-Thompson Theorem [12]. In the **HOL/Isabelle** Archive of Formal Proofs [22] one finds a number of proof libraries devoted to algebraic domains. Lately in **ACL2** [1] an algebraic hierarchy has been built in order to support reasoning about Common Lisp programs [21].

The Mizar system [5], [17], [31] provides a methodology to model algebraic domains based on attributed types [4]. Using so-called cluster registrations one can express (and prove) logical implications between attributes, in this way extending subtyping of attributed types. This allows not only to model algebraic domains in a generic way, but also to draw connections between – also already existing – algebraic domains. We claim that this approach is suitable to develop algebraic hierarchies that a) are generic in the sense that notations and theorems introduced in a class of algebraic domains are automatically available in subclasses b) are easily extensible by both algebraic domains and additional notations c) can automate a great deal of the natural switching between algebraic domains mathematicians are used to and d) are highly convenient for open repositories with lots of authors.

To support this claim we present in Section II a detailed hierarchy of rings up to fields, containing such algebraic domains such as unique factorization domains (UFDs), principal ideal domains (PIDs), and others. In Section III we show how homomorphisms can be incorporated into this hierarchy and how properties of homomorphisms can be used to automatically infer properties about the underlying algebraic domains. Finally, in Section IV, we discuss how to extend the hierarchy towards more involved domains such as polynomial rings and ordered fields. At the end we draw some conclusions.

II. An Intrinsic Hierarchy of Rings

In Mizar, algebraic domains are built based on structure definitions giving the signature – carriers and operations – of the domain. Informally, a ring is understood as *an algebraic structure consisting of a set of elements equipped with binary operations $+$ and $\cdot$ satisfying three sets of axioms*. More formally, usually this leads to understanding mathematical structures as ordered tuples, and in the case of a ring we have

$$\langle R, +, \cdot \rangle.$$

This could make potential troubles if we try to define rings through simpler notions, namely groups, which are usually

$$\langle G, + \rangle \quad \text{or} \quad \langle G, \cdot \rangle$$

978-83-60810-90-3/$25.00

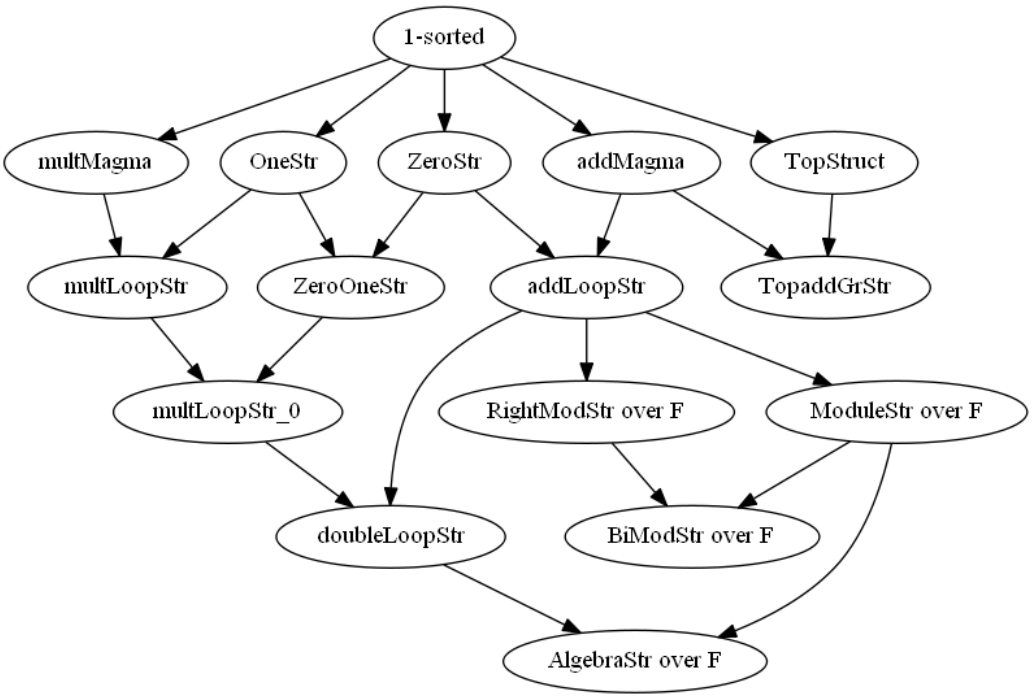


Figure 1. Net of basic algebraic structures in the Mizar Mathematical Library

(in additive or multiplicative notation, respectively). Of course then, ordinary concatenation of tuples does not work properly.

In the Mizar system, structures were implemented as partial functions with the syntax as below.

```
struct (Predecessor_List) Structure_Name
  (# selector_1 -> type_1,
     selector_2 -> type_2,
               ...
     selector_n -> type_n #);
```

This could lead to the tree, or rather a forest of 157 structures, as there are primitive structures other than `1-sorted`. However, as multiple predecessors are allowed, we should look at the diagram of interconnections as at a net. A part of such structure, dealing with basic algebraic signatures, is shown at Figure 1.

So, for example, central item in this hierarchy is

```
definition
  struct (addLoopStr, multLoopStr_0)
    doubleLoopStr
  (# carrier -> set,
     addF, multF -> BinOp of the carrier,
     OneF, ZeroF -> Element of the carrier #);
end;
```

which gives the signature of rings and fields (another one is `ModuleStr over F` which gives raise to the theory of vector spaces). Note that `doubleLoopStr` inherits from both `addLoopStr` and `multLoopStr_0`, that is it joins the signatures of additive and multiplicative groups. Particular properties such as commutativity or the existence of inverse elements are described by attribute definitions (see [35]). As a consequence, attributes defined for these become available and need not to be stated again. A ring is now just a `doubleLoopStr` with the appropriate collection of attributes:

```
definition
  mode Ring is Abelian add-associative
    right_zeroed right_complementable
    associative well-unital distributive
      non empty doubleLoopStr;
end;
```

Observe that because the Axiom of Choice is hardcoded in the Mizar checker, the collection of attributes clustered in the above definition of type should be shown to exist for at least one object; otherwise (with the illustrative example of *infinite empty set*) this should be contradictory. This is called the paradigm of non-emptiness of types in Mizar.

More interesting are different subclasses of rings that form a hierarchy according to their additional properties, e.g.

$$rings \supseteq commutative\ rings \supseteq integral\ domains \supseteq$$

$$\supseteq GCD\ domains \supseteq UFDs \supseteq PIDs \supseteq$$

$$\supseteq Euclidean\ domains \supseteq fields$$

to mention the most common ones. Each such subclass is easily characterized by adding an attribute describing its defining property, for example

```
definition
  let L be non empty doubleLoopStr;
```

```
  attr L is PID means
    for I being Ideal of L holds I is principal;
end;
```

Note that the definition does not use integral domains – in fact not even rings, but just their signature. The property of an ideal being principal just not relies on other properties such as commutativity or the absence of zero divisors (at least when defining the property; later on more properties may be necessary to show that this one is fulfilled in a particular domain – see [37]). The above hierarchy can now be established by observing that one defining property implies another – just like in mathematical textbooks:

```
registration
  cluster Euclidean -> PID for comRing;
end;
```

This way of defining the hierarchy has two major advantages. Firstly, a proof of the implication has to be given. This may be obvious, but we like to emphasize at this point, that proofs are an indispensable part of a repository. Note also, that the registration is about commutative rings, not about integral domains. Analogous to the definition of the attribute `PID` this points out, that Euclid's property implies that every ideal of the ring is principal even in presence of zero divisors – although both domains are usually defined as integral domains with the appropriate additional property.

Secondly, and more important here, cluster registrations extend automation of proving in Mizar, that is after the above registration the theorem like

```
theorem
  for R being Euclidean domRing holds
    R is PID domRing;
```

becomes obvious. Here `domRing` denotes integral domain, where in the attribute `domRing-like` the commutativity is not taken into account. Such granularity allows for better reuse of knowledge. As a secondary consequence all notations – definitions, predicates and also theorems – established for the subclass now are available for the superclass, too. In practice, this means that notations are generic: There is no need to define, for example, greatest common divisors in Euclidean domains. They can be already introduced in GCD domains (see [20]) and are therefore available in Euclidean domains, once Euclidean domains have been incorporated into the hierarchy.

The proofs necessary to built the above-mentioned hierarchy of rings have been carried out in a number of Mizar articles [2], [28], [30], [36]. Together they establish an environment in which arguing about different kinds of rings – and switching between them – can be performed in a way very similar to the usual mathematical processing.

First of all, the hierarchy can easily be extended when necessary or convenient: For example Noetherian rings are integral domains (rings) in which every ideal is finitely generated. Thus every PID is a Noetherian ring. The corresponding part of the hierarchy looks as follows:

```
definition
  let L be non empty doubleLoopStr;
  attr L is Noetherian means
    for I being Ideal of L holds
      I is finitely_generated;
end;

registration
  cluster PID -> Noetherian
        for non empty doubleLoopStr;
end;
```

Furthermore, concrete domains such as the ring of integers or the field of real numbers can be integrated in a straightforward way: a concrete domain is an instance of an abstract domain and can be introduced by just defining its concrete carriers and operations. The ring of integers, for example, is then given by

```
definition
  func INT.Ring -> doubleLoopStr equals
    doubleLoopStr(#INT,addint,multint,In(1,INT),
                  In(0,INT)#);
end;
```

and the following registrations then show that `INT.Ring` is both an integral and a Euclidean domain, hence connect `INT.Ring` with the hierarchy.

```
registration
  cluster INT.Ring -> non degenerated
    add-associative right_zeroed
    right_complementable distributive
    commutative associative Abelian
    domRing-like;
end;

registration
  cluster INT.Ring -> Euclidean;
end;
```

With these registrations all notations – definitions, predicates and theorems – established for the abstract domains become available for `INT.Ring`, too. Note that from this in particular follows – without any further proof, because this has been proven inside the hierarchy – that the ring of integers is both UFD and Noetherian, that is

```
theorem
  INT.Ring is UFD domRing;

theorem
  INT.Ring is Noetherian domRing;
```

are obvious. Moreover, it even does not matter which domain – Noetherian rings or the ring of integers – is added to the hierarchy first. In both cases the above theorems are obvious for the Mizar checker. Note however that in order to make this automation working, it is convenient to have a bunch of useful examples of concrete mathematical structures (just to assure that at least one object with desired properties exists).

At the end of this section it should be noted that the formal proof that UFDs are GCD domains has not been completed yet, but see [27] where the fundamental theorem of arithmetic – giving a blueprint for the proof – has been formalized.

III. RING HOMOMORPHISMS

When working with algebraic domains (ring-) homomorphisms are indispensable. Therefore homomorphisms are an essential part of an algebraic hierarchy. Homomorphisms are essentially mappings between rings with additional properties, that hence can again be defined by adding attributes describing these properties:

```
definition
  let R be Ring,
      S be R-homomorphic Ring;
  mode Homomorphism of R,S is
     additive multiplicative unity-preserving
     Function of R,S;
end;
```

The attribute `homomorphic` for `S` is necessary here, because Mizar does not allow empty modes: For each pair of parameters `R` and `S` it has to be proved that there exists a homomorphism from `R` into `S`. Therefore the definition of homomorphisms can take into account only such rings `S`, for which such a homomorphism indeed exists. This is ensured by adding the attribute `homomorphic`, which has the ring `R` as a parameter:

```
definition
  let R,S be Ring;
  attr S is R-homomorphic means
    ex f being Function of R,S
     st f is additive multiplicative
     unity-preserving;
end;
```

Note that together with the hierarchy presented in Section II this definition provides homomorphisms for all kinds of rings up to fields. By the way, homomorphisms are functions preserving the unity and the zero, but the latter one can be deduced (and really is in this framework) automatically, hence it is not explicitly given in this collection of attributes. Therefore additional properties of homomorphisms for more advanced rings can now be easily incorporated, for example that homomorphisms between fields are actually monomorphisms:

```
registration
  let F be Field, E be F-homomorphic Field;
  cluster -> monomorphism
           for Homomorphism of F,E;
end;
```

The property of being monomorphic is then automatically added when later working with homomorphisms of fields. The same holds naturally for properties of the image of homomorphisms:

```
registration
  let F be comRing, E be F-homomorphic Ring,
      f be Homomorphism of F,E;
  cluster Image f -> commutative;
end;
```

says that the image of a commutative ring is a commutative ring. So there is no need to distinguish homomorphisms between different kind of rings. In fact – as the last registration shows – it is even sufficient to claim that the homomorphism's codomain is an ordinary ring.

For a small example illustrating these techniques consider now rings R and S and a homomorphism $f : R \longrightarrow S$. The first isomorphism theorem then states that $R/(\ker f) \cong$ Image f. Quotient rings have been defined in [26]. The image of f is here understood as the subring of S with carrier range f, the operations of Image f are then just restrictions of the ones of S. This gives

```
definition
  let R be Ring, S be R-homomorphic Ring,
      f be Homomorphism of R,S;
  func Image f -> Ring means
    the carrier of it = rng f &
    the addF of it =
                (the addF of S) || (rng f) &
    the multF of it =
                (the multF of S) || (rng f) &
    the OneF of it = 1.S &
    the ZeroF of it = 0.S;
end;
```

Now the homomorphism $h : R/(\ker f) \longrightarrow$ Image$f$ given by $[a] \mapsto f(a)$ for $a \in R$ can easily be defined and shown to be bijective [28], so

```
theorem
  for R being Ring, S being R-homomorphic Ring,
      f being Homomorphism of R,S holds
    R / (ker f), Image f are_isomorphic;
```

Note that if R is a field we get that $R/(\ker f)$ is a field also: If R is a field, so is Image f, and hence its isomorphic copy $R/(\ker f)$. This argument can now be automated by observing that homomorphic images of fields are fields – as in the case of commutative rings from above – and reformulating the isomorphism theorem as a registration using the attribute `isomorphic` that is defined analogously to `homomorphic`.

```
registration
  let F be Field, E be F-homomorphic Ring,
      f be Homomorphism of F,E;
  cluster Image f -> almost_left_invertible;
end;

registration
  let R be Ring, S be R-homomorphic Ring,
      f be Homomorphism of R,S;
  cluster R / (ker f) -> (Image f)-isomorphic;
end;
```

These two registrations hence automate the argument above and therefore the following theorems are now obvious, that is are accepted by the Mizar checker without further proof.

```
theorem
  for F being Field, R being F-homomorphic Ring,
      f being Homomorphism of F,R holds
    Image f is Field;

theorem
  for F being Field, R being F-homomorphic Ring,
      f being Homomorphism of F,R holds
    F/(ker f) is Field;
```

IV. Extending the Hierarchy

A. Polynomial Rings

New domains are not always built solely by adding new properties, but may contain other (abstract) domains as parameters. The standard example here are vector spaces or modules that are built over a field or a ring, respectively. They, however, define new classes of algebraic domains.

More interesting are polynomial rings $R[X]$ realizing an operator within the class of rings. The standard definition of polynomial rings is well-known: Polynomials over R are sequences over R or functions $p : \mathbb{N} \longrightarrow R$, on which addition and multiplication are defined appropriately (see [29]):

```
definition
  let R be Ring;
  func Polynom-Ring R -> non empty doubleLoopStr
    equals
  doubleLoopStr (# Polys R, addpoly R,
    multpoly R, 1_.R, 0_.R #);
end;
```

Using registrations `Polynom-Ring R` can now be incorporated as usual into the hierarchy by showing that the carrier – the set of polynomials – fulfills the necessary properties, for example

```
registration
  let R be Ring;
  cluster Polynom-Ring R ->
        add-associative right_zeroed
        right_complementable;
end;
```

In fact it is not necessary for R to be a ring to prove each individual property – even when defining $R[X]$. In this registration, for example, distributivity and that polynomial addition forms a group are sufficient [29].

However, the hierarchy is able to deal with more involved properties of $R[X]$ also. For example, if R is without zero divisors, so is $R[X]$, which is described by the following registration.

```
registration
  let R be domRing;
  cluster Polynom-Ring R -> domRing-like;
end;
```

In this way additional properties of $R[X]$ are added depending on properties of R. When working with the hierarchy Mizar now automatically adds such properties to $R[X]$, if R fulfills the conditions of the registration.

In fact, the parameter R can even be a field F – based on the hierarchy of Section II the Mizar checker infers that F is a ring, so the notation `Polynom-Ring F` exists and one can formulate

```
registration
  let F be Field;
  cluster Polynom-Ring F -> Euclidean;
end;
```

So we automatically get that $F[X]$ is a PID and that gcds for polynomials over a field exist. Note also that `F_Real` is the field of real numbers; therefore real polynomials are now just given by `Polynom-Ring F_Real`.

In the context of polynomials another notation also becomes interesting: the notion of subring – $\subseteq$ for short – giving relations such as $\mathbb{Z} \subseteq \mathbb{Q}$, $\mathbb{Q} \subseteq \mathbb{R}$ or $\mathbb{Z}[X] \subseteq \mathbb{R}[X]$ – and for polynomial rings one often reads $\mathrm{R} \subseteq R[X]$ (see e.g. [39]). Now, the notation of a subring is easily defined by

```
definition
  let R be Ring;
  mode Subring of R -> Ring means
    the carrier of it c= the carrier of R &
    the addF of it =
        (the addF of R) || the carrier of it &
    the multF of it =
        (the multF of R) || the carrier of it &
    1.it = 1.R &
    0.it = 0.R;
end;
```

Then theorems for the above relations can easily be shown, for example

```
theorem
  INT.Ring is Subring of F_Real;

theorem
  Polynom-Ring INT.Ring is
    Subring of Polynom-Ring F_Real;
```

The property $R \subseteq R[X]$, however, cannot be shown; it is just not true: an element $a \in R$ is not a polynomial, so the carrier of R is not included in the carrier of $R[X]$ as it contains sequences or functions, that is ordinary set inclusion between carriers does not work here. The solution is found in the literature [39]:

> *We regard $F \subset F[X]$ by identifying the element $a \in F$ with the constant polynomial $a \in F[X]$.*

(More precisely, the identification $i : R \longrightarrow R[X]$, $a \mapsto a(x)$ is a monomorphism, and therefore allows to embed R into $R[X]$.) To formally reconstruct this identification in a repository one now has to construct a new ring R' with the corresponding carrier

$$(R[X] \backslash \{p \in R[X] : p \text{ is constant}\}) \cup R$$

and adapted addition and multiplication. This is both tedious and technical, but, what is more important, R' does not solve the problem, either: Though now one has $R \subseteq R'$, of course, R' is not exactly the polynomial ring $R[X]$ in the above sense, but only an isomorphic copy of it.

Modifying the definition of subring in the sense that a ring R is a subring of R' if R can be embedded into R' – that would allow to prove that R is a subring of $R[X]$ – is too liberal: It destroys the simplicity and elegance of the subring notation. As a consequence in mathematical repositories this kind of using the definition of subring can be modelled only at the level of morphisms, e.g. via theorems such as

```
theorem
  for R being Ring
   ex R' being Ring st R c= R' &
```

```
      R',Polynom-Ring R are_isomorphic;

theorem
  for R being Ring
   ex R' being Subring of Polynom-Ring R
   st R,R' are_isomorphic;
```

B. Ordered Fields

There are situations in which the extension of an algebraic domain can be realized in more than one way. The standard example here is the neutral element e, that is, for example added to semigroups in order to construct monoids. e can be introduced solely as an adjective in an attribute definition then claiming the existence of e – or as an additional part of the underlying structure then just claiming $a * e = a$ for all a in the carrier, where e is now the element given by the new part of the structure. Usually the second alternative is used here, because this allows for an equational definition of e's properties.

A similar situation occurs when the additional properties to be defined do not concern the domain at hand itself, but are described based on additional notations. A typical example are ordered domains, here the newly added properties concern a relation over the domain. A standard definition, for example, is:

> *An ordered field is a pair $(F, \leq)$, where F is a field and $\leq$ is a (total) relation being compatible with the field operations.*

So, ordered structures can be easily built using the second alternative from above by just adding an additional part for the relation to the underlying structure definition.

```
definition
  struct (doubleLoopStr) ordereddoubleLoopStr
    (# carrier -> set,
       addF, multF -> BinOp of the carrier,
       OneF, ZeroF -> Element of the carrier,
       OrdF -> Order of the carrier #);
end;
```

Then, based on an attribute `compatible_with` describing compatibility of the order with the field operations, one defines the mode `orderedField`. This allows to formalize and prove theorems such as follows:

```
theorem
  for F being orderedField holds 0.F <= 1.F;

theorem
  for F being orderedField,
      a being Element of F holds
    0.F <= a|^2;

theorem
  for F being orderedField holds -1.F <= 0.F;
```

Here `a <= b` denotes `[a,b] ∈ the OrdF of F`, if `a` and `b` are of type `Element of F`.

Also concrete domains, such as for example these over the set of all real numbers, fit into this approach. With `<=_R` being the usual order relation over the real numbers, the following definition establishes the real numbers as an ordered field.

```
definition
  func OF_Real -> orderedField equals
    (# REAL, addreal, multreal, In(1,REAL),
       In(0,REAL), <=_R #);
end;
```

In this case, however, introducing concrete domains this way turns out to be too restrictive: When fixing the field – of an ordered field – one immediately also has to fix the ordering. This results in inconveniences when further developing the theory. For real numbers, for example, there exists one ordering only. To formalize this within the above approach one needs to say that for two ordered fields, in both of which the field happens to be the real numbers, the orderings coincide:

```
theorem
  for F,E being orderedField
    st the doubleLoopStr of F = F_Real &
    the doubleLoopStr of E = F_Real
  holds the OrdF of F = the OrdF of E;
```

This is too clumsy to work with – and to be part of a contemporary repository. Of course, this does not bother mathematicians. If convenient they just fix the field and leave the ordering(s) as a parameter:

> *Let F be the field of real numbers. Let $\leq$ and $\leq'$ be orderings of F. Then $\leq = \leq'$.*

At this point it should be mentioned that apart of the abstract hierarchy shown at Fig. 1, another one, with the set of all real numbers fixed in certain places, is available in the MML. Such net of notions (see Fig. 2) is concentrated around the structure as follows:

```
definition
  struct (addLoopStr) RLSStruct
  (# carrier -> set,
       ZeroF -> Element of the carrier,
        addF -> BinOp of the carrier,
        Mult -> Function of
      [:REAL, the carrier :], the carrier #);
end;
```

and is still kept in the Mizar library for backward compatibility reasons. This was useful and handy some ten years ago for Mizar developers, but the approach was reimplemented. The mechanism of the identification of ordinary operations on reals with corresponding abstract field operations was discussed in detail in our paper [18]. There we described the usefulness of automatic consideration of core equalities via `identify` construction, which does not force the mathematician to add them explicitly to the proof.

For a reasonable formalization one also needs such a flexible way of talking about a domain and its (possible) orderings. Therefore one has to resign from the above, natural approach: The solution is not to extend the structure by another part, but to define the existence of an ordering externally – as an additional property. Then an ordered field is just a field for which (at least) one ordering exists:

```
definition
  let F be Field;
```

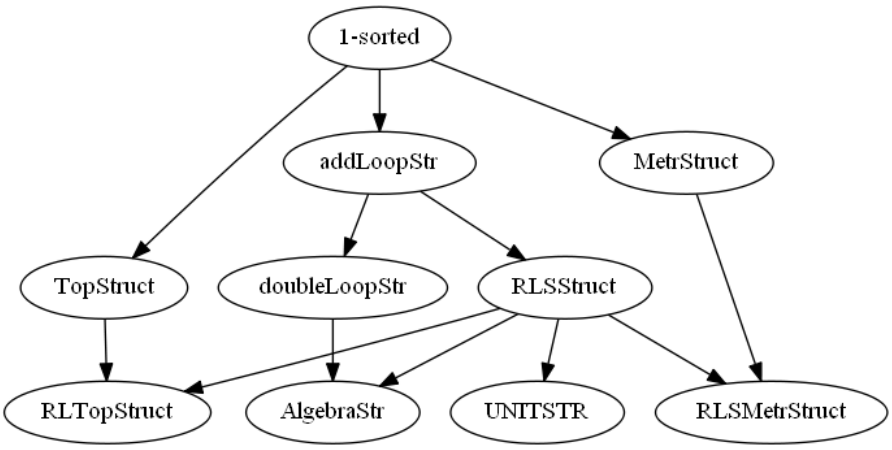


Figure 2. The correspondence of real-valued structures in the Mizar Mathematical Library

```
  attr F is ordered means
    ex O being Order of the carrier of F
      st O is compatible_with F;
end;
```

Now an ordered field is just an ordinary field, a concrete ordering has only to be fixed in the proof that the field can be ordered. This gives the above-mentioned flexibility: One can state that the real numbers can be ordered – by using the natural ordering <=_R. After that – now knowing that an ordering for a concrete or abstract domain exists – one can introduce one or more of them whenever necessary or convenient.

```
registration
  cluster F_Real -> ordered;
end;

theorem
  for O,P being Ordering of F_Real holds O = P;

theorem
  for F being ordered Field
  for O,P being Ordering of F
    st O c= P holds O = P;
```

Of course the theorems stated above for the first approach remain valid. However, they now look a little different, because the ordering has become a parameter of the theorems – as it is not fixed in the structure, here the ordered field, anymore.

```
theorem
  for F being ordered Field
  for P being Ordering of F
    holds 0.F <=_P 1.F;

theorem
  for F being ordered Field
  for P being Ordering of F,
  for a being Element of F
    holds 0.F <=_P a|^2;

theorem
  for F being ordered Field
  for P being Ordering of F
    holds -1.F <=_P 0.F;
```

Summarizing, it turns out, that for ordered fields adding a new part to the structure – the solution usually preferred – is inferior to describing the property of being ordered solely by an adjective, though this means that an existential quantifier occurs in the attribute definition.

V. Conclusion

We have illustrated how in Mizar a deep algebraic hierarchy has been built that to a great deal resembles the natural changing between algebraic domains known from mathematics. The key technique is the application of Mizar's attributed types: The adjectives defined by attributes enable the natural extension of existing algebraic domains by adding new properties. Furthermore, implications between adjectives can be formulated in the form of cluster registrations. These not only prove that, for example, an Euclidean domain is a UFD, but also automate inferring the implication. Summarizing the presented approach supports building algebraic hierarchies that are easily extended and refined when necessary.

Mizar structures together with *locales* – a similar concept implemented in Isabelle [3] can be a reasonable improvement in writing mathematical proofs and their automated discovery. Interesting categorical motivation of such an approach is presented in [7]. Modules can be treated globally for all proof assistants, and a kind of interface allowing for information interchange is proposed as MMT – a module system for mathematical theories with scalable formalism [34]. All axiomatic theories can be viewed from the metalevel, using the concept of *realms* – which could also enable reasonings via consolidating knowledge about theories.

The hierarchy of structures available in the Mizar Mathematical Library was described in [35] and was enriched during the formalization of the proof of Fundamental Theorem of Algebra [29]. However in 2007 a big refinement (a revision [19]) took place, and parts of the net of structures together with

corresponding attributes were a subject for refactoring. In that time the Library Committee of the Association of Mizar Users (with the two first authors of the current paper as its members) wrote a library item [30] which is now a backbone for all described constructions. This field of algebra is continuously formalized, and recent Mizar article [38] contains results about selected properties of polynomial ring.

In fact this methodology is not restricted to algebraic domains and algebraic hierarchies. In Mizar one finds similar hierarchies concerning a number of mathematical structures such as, for example, posets, lattices, topologies, topological groups, topological lattices, and graphs. In the Mizar repository there is a series of articles devoted to algebraic structures using multiplicative notation. Recently, Coghetto [8] made a monographic Mizar article which was a modification of already accepted ones with an addition as a basic binary operation (instead of a multiplication). Even if the work was not too much time-consuming, some of the library techniques can be further improved (e.g. in the direction of generic structures – because `addMagma` and `multMagma` could be potentially more unified, but this would require implementational work of the Mizar checker).

The direction of enhancing computer reasoning tools, which is quite popular and efficient, is to use external specialized software. In the case of Mizar some experiments were proposed by Naumowicz [32] with SAT solvers in order to improve the efficiency of boolean calculations. Of course, this could be taken into account in the area of algebraic domains. As another example, Grabowski succesfully used Prover9 for reasonings about lattices [14]. This is another field of research where efficient treatment of algebraic hierarchies is very important, and recent formalization of Stone algebras as generalization of Boolean algebras [15] shows the usefulness of the Mizar system.

REFERENCES

[1] The ACL2 Sedan Theorem Prover; available at http://acl2s.ccs.neu.edu/acl2s/doc/

[2] J. Backer, P. Rudnicki, and C. Schwarzweller, *Ring Ideals*; Formalized Mathematics, vol. 9(3), pp. 565–582, 2001.

[3] C. Ballarin, *Interpretation of Locales in Isabelle: Theories and Proof Contexts;* in: 5th International Conference on Mathematical Knowledge Management, MKM 2006, Lecture Notes in Computer Science, 4108, pp. 31–43, 2006. http://dx.doi.org/10.1007/11812289_4

[4] G. Bancerek, *On the Structure of Mizar Types*, Electronic Notes in Theoretical Computer Science, vol. 85 (7), Elsevier, 2003. http://dx.doi.org/10.1016/S1571-0661(04)80758-8

[5] G. Bancerek, C. Byliński, A. Grabowski, A. Korniłowicz, R. Matuszewski, A. Naumowicz, K. Pąk, and J. Urban, *Mizar: State-of-the-art and Beyond*, in: M. Kerber, J. Carette, C. Kaliszyk, F. Rabe, and V. Sorge (eds.), International Conference on Intelligent Computer Mathematics – Proceedings, Lecture Notes in Computer Science 9150, pp. 261–279, 2015. http://dx.doi.org/10.1007/978-3-319-20615-8_17

[6] J. Carette, W.M. Farmer, and M. Kohlhase, *Realms: A Structure for Consolidating Knowledge about Mathematical Theories,* in: S. Watt et al. (eds.), International Conference on Intelligent Computer Mathematics – Proceedings, Lecture Notes in Computer Science 8543, pp. 252–266, 2014. http://dx.doi.org/10.1007/978-3-319-08434-3_19

[7] J. Carette and R. O'Connor, *Theory Presentation Combinators,* in: J. Jeuring et al. (eds.), International Conference on Intelligent Computer Mathematics – Proceedings, Lecture Notes in Computer Science 7362, pp. 202–215, 2013. http://dx.doi.org/10.1007/978-3-642-31374-5_14

[8] R. Coghetto, *Groups – Additive Notation*; Formalized Mathematics, vol. 23(2), pp. 127–160, 2015. http://dx.doi.org/10.1515/forma-2015-0013

[9] The Coq Proof Assistant; available at http://coq.inria.fr.

[10] Y. Futa, H. Okazaki, and Y. Shidama, *Torsion Part of $\mathbb{Z}$-module*; Formalized Mathematics, vol. 23(4), pp. 297–307, 2015. http://dx.doi.org/10.1515/forma-2015-0024

[11] H. Geuvers, R. Pollack, F. Wiedijk, and J. Zwanenburg, *A Constructive Algebraic Hierarchy in Coq*; Journal of Symbolic Computation, vol. 34(4), pp. 271–286, 2002. http://dx.doi.org/10.1006/jsco.2002.0552

[12] G. Gonthier et al., *A Machine-Checked Proof of the Odd Order Theorem*; in: S. Blazy, C. Paulin-Mohring, D. Pichardie (eds.), Proceedings of the 4th International Conference on Interactive Theorem Proving, Lecture Notes in Computer Science 7998, pp. 163–179, 2013. http://dx.doi.org/10.1007/978-3-642-39634-2_14

[13] A. Grabowski, *Efficient Rough Set Theory Merging*; Fundamenta Informaticae, 135(4), pp. 371–385, 2014. http://dx.doi.org/10.3233/FI-2014-1129

[14] A. Grabowski, *Mechanizing Complemented Lattices Within Mizar Type System*; Journal of Automated Reasoning, vol. 55(3), pp. 211–221, 2015. http://dx.doi.org/10.1007/s10817-015-9333-5

[15] A. Grabowski, *Stone Lattices*; Formalized Mathematics, vol. 23(4), pp. 387–396, 2015. http://dx.doi.org/10.2478/forma-2015-0031

[16] A. Grabowski, A. Korniłowicz, and A. Naumowicz, *Mizar in a Nutshell*; Journal of Formalized Reasoning, 3(2), pp. 153–245, 2010. http://dx.doi.org/10.6092/issn.1972-5787/1980

[17] A. Grabowski, A. Korniłowicz, and A. Naumowicz, *Four decades of Mizar*; Journal of Automated Reasoning, vol. 55(3), pp. 191–198, 2015. http://dx.doi.org/10.1007/s10817-015-9345-1

[18] A. Grabowski, A. Korniłowicz, and C. Schwarzweller, *Equality in computer proof-assistants*; in Proceedings of 2015 Federated Conference on Computer Science and Information Systems, FedCSIS 2015, M. Ganzha, L. Maciaszek, M. Paprzycki (eds.), IEEE, pp. 45–54, 2015. http://dx.doi.org/10.15439/2015F229

[19] A. Grabowski and C. Schwarzweller, *Revisions as an essential tool to maintain mathematical repositories*; in: 14th Symposium on Towards Mechanized Mathematical Assistants, Calculemus'07/MKM'07, Lecture Notes in Computer Science, pp. 235–249, 2007 http://dx.doi.org/10.1007/978-3-540-73086-6_20

[20] A. Grabowski and C. Schwarzweller, *Towards Standard Environments for Formalizing Mathematics*; in: Proceedings of the 6th Podlasie Conference on Mathematics, A. Gomolinska, A. Grabowski, M. Hryniewicka, M. Kacprzak, E.Schmeidel (eds.), Białystok, Poland, 2014.

[21] J. Heras, F.J. Martín-Mateos, and V. Pascual, *Modelling Algebraic Structures and Morphisms in ACL2*; Applicable Algebra in Engineering, Communication and Computing, vol. 26(3), pp. 277–303, 2015. http://dx.doi.org/10.1007/s00200-015-0252-9

[22] The Isabelle Proof Assistant; available at isabelle.in.tum.de.

[23] P.B. Jackson, *Enhancing the Nuprl Proof Development System and Applying it to Computational Abstract Algebra*; PhD thesis, Cornell University, 1995.

[24] R.D. Jenks and R. Sutor, *AXIOM – The Scientific Computation System*; Springer Verlag, 1992. http://dx.doi.org/10.1007/978-1-4612-2940-7

[25] A. Korniłowicz, *Definitional Expansions in Mizar*; Journal of Automated Reasoning, vol. 55(3), pp. 257–268, 2015. http://dx.doi.org/10.1007/s10817-015-9331-7

[26] A. Korniłowicz, *Quotient Rings*; Formalized Mathematics, vol. 13(4), pp. 573–576, 2005.

[27] A. Korniłowicz and P. Rudnicki, *The Fundamental Theorem of Arithmetic*; Formalized Mathematics, vol. 12(2), pp. 179–186, 2004.

[28] A. Korniłowicz and C. Schwarzweller, *The First Isomorphism Theorem and Other Properties of Rings*; Formalized Mathematics, vol. 22(4), pp. 291–302, 2014. http://dx.doi.org/10.2478/forma-2014-0029

[29] R. Milewski, *The Ring of Polynomials*; Formalized Mathematics, vol. 9(2), pp. 339–346, 2001.

[30] The Mizar Library Committee, *Basic Algebraic Structures*; 2007. MML Id: `ALGSTR_0`, available at http://mizar.org/version/current/html/algst_0.html

[31] The Mizar Home Page; available at http://mizar.org.

[32] A. Naumowicz, *Automating Boolean Set Operations in Mizar Proof Checking with the Aid of an External SAT Solver*; Journal of Automated Reasoning, vol. 55(3), pp. 285–294, 2015. http://dx.doi.org/10.1007/s10817-015-9332-6

[33] Pąk K.: Methods of Lemma Extraction in Natural Deduction Proofs, *Journal of Automated Reasoning,* vol. 50(2), pp. 217–228, 2013. http://dx.doi.org/10.1007/s10817-012-9267-0
[34] F. Rabe and M. Kohlhase, *A Scalable Module System,* Information & Computation, 230, pp. 1–54, 2013. http://dx.doi.org/10.1016/j.ic.2013.06.001
[35] P. Rudnicki, A. Trybulec, and C. Schwarzweller, *Commutative Algebra in the Mizar System*; Journal of Symbolic Computation, vol. 32(1/2), pp. 143–169, 2001. http://dx.doi.org/10.1006/jsco.2001.0456
[36] C. Schwarzweller, *The Ring of Integers, Euclidean Rings and Modulo Integers*; Formalized Mathematics, vol. 8(1), pp. 29–34, 1999.
[37] C. Schwarzweller, *Designing Mathematical Libraries based on Requirements for Theorems*; Annals of Mathematics and Artificial Intelligence, vol. 38(1–3), pp. 193–209, 2003. http://dx.doi.org/10.1023/A:1022924032739
[38] C. Schwarzweller, A. Korniłowicz, and A. Rowinska-Schwarzweller, *Some Algebraic Properties of Polynomial Ring*; Formalized Mathematics, vol. 24(3), 2016. http://dx.doi.org/10.1515/forma-2016-0019
[39] S.H. Weintraub, *Galois Theory*; 2nd edition, Springer Verlag, 2009.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 373–381
DOI: 10.15439/2016F290
ACSIS, Vol. 8. ISSN 2300-5963

Tarski's Geometry Modelled in Mizar Computerized Proof Assistant

Adam Grabowski
Institute of Informatics
Department of Mathematics and Informatics
University of Białystok,
ul. Konstantego Ciołkowskiego 1 M, 15-245 Białystok, Poland
Email: adam@math.uwb.edu.pl

***Abstract*—In the paper, we discuss the formal approach to Tarski geometry axioms modelled with the help of the Mizar computerized proof assistant system. Although our basic development was inspired by Julien Narboux's Coq pseudo-code and is dated back to 2014, there are significant steps in the formalization of geometry done in the last decade of the previous century. Taking this into account, we will propose the reuse of existing results within this new framework (including Hilbert's axiomatic approach), with the ultimate future goal to encode the textbook *Metamathematische Methoden in der Geometrie* by Schwabhäuser, Szmielew and Tarski. We try however to go much further from the use of simple predicates in the direction of the use of structures with their inheritance, attributes as a tool of more human-friendly namespaces for axioms, and registrations of clusters to obtain more automation (with the possible use of external equational theorem provers like Otter/Prover9).**

I. Introduction

FOR YEARS, foundations of geometry attracted a lot of interest of resarchers from various areas of mathematics. From the very beginnings, human thought was stimulated by geometrical objects, to take Thales as the prominent example of an ancient philosopher. Classical geometry involved illustrative examples and construction problems instead of a building strong axiomatic basis. However, from the modern viewpoint of automated theorem-provers, diagrams can deliver some really tough problems. Here an important example is the possibility of ruler-and-compass construction: impossibility of trisecting the angle and doubling the cube (as two out of four problems of antiquity), where the treatment of constructible numbers is way more efficient from the formal point of view.

Euclid and his *Elements* are often recalled as one of the first successful uses of an axiomatic method in mathematics, and such an approach can be formalized efficiently with the use of computer proof-assistants. Then, changing even simple notions with obvious (at least at the very first sight) properties, as parallel postulate (or Playfair axiom), gave rise to various geometries (e.g., Bolyai-Lobachevskian hyperbolic geometry). The same work can be modelled with machine formalizations, using various sets of axioms (creating *types*). Now, apart from the discussion whether the non-emptiness of types in Mizar is real difficulty (because from informal point of view one can consider an object with any properties, even mentioning their coherence), and how much more can be attained if the reimplementation of the Mizar type system will be done in the foreseeable future, we have to cope with the limitations of the existing type structure. On the other hand, type checking allows some errors to be caught early – when making mathematical definitions. In this context the requirement of constructing at least one object of the desired type is quite natural, as it prevents contradictory types. Similarly, the appropriate model had to be constructed either to assure that proposed axioms are correct (which is not very hard as they can be parsed by ordinary mathematician even straight from its corresponding Mizar source code), or (and this is probably even more important) to bind the fresh formal apparatus with the existing Mizar developments. Some of them are written in a language which is not as expressive as contemporary Mizar language is; in the time of the beginnings of the Mizar Mathematical Library (MML) as a tight collection of Mizar articles covering various branches of mathematics, geometry was an area which was developed quite dynamically.

The language of the Mizar system was influential for other systems for formalization of mathematics, e.g. `miz3` is a proof interface built on top of HOL Light interactive theorem prover, with the declarative language compatible with the Mizar language [47]. Recently, William Richter used `miz3` tool to formalize Tarski geometry axioms, with the ultimate aim to incorporate it in HOL Light, but of course part of his pseudo-code could be also treated both as a case study and as a good starting point for further work.

The choice of the topic is not accidental – recent code available in Coq [7] and the use of automated equational provers caught an eye of researchers and, as a by-product, some results, which shed some new light on the axiomatization of geometry, were published. One of the bright milestones was also the publication of the new issue of the classical textbook *Metamathematische Methoden in der Geometrie* by Wolfram Schwabhäuser, Wanda Szmielew and Alfred Tarski [40] (to which we refer by the acronym SST) with the foreword of Michael Beeson.

In this paper, we are not focused on any of geometric challenges known in the community, as proving Hilbert axioms from Tarski [7], or formalizing full SST in Mizar, although it can be definitely good starting point as in [38] we prove some Hilbert's axioms. What we were trying to do was to

978-83-60810-90-3/$25.00

increase of the integrity of (the geometrical part of) the MML as pointed out in the paper of Piotr Rudnicki and Andrzej Trybulec [39] and this could be a kind of partial realization of their postulates. This was done mainly via the mechanism of *revisions* of the repository – stepwise refinement of items already included in MML, done not necessarily by authors themselves [19]. An alternative approach – focusing on computations instead of proofs, and work in the analytic framework of Euclidean spaces $\mathbb{R}^n$ is also well-represented in the repository of Mizar texts, with the recent examples of Morley trisector theorem or Ceva theorem. The process of formalizing geometry within Mizar Mathematical Library started years before first Coq geometry formalizations; furthermore, constructive logic behind the Coq proof assistant naturally forces program extraction from proofs and intuitionistic setting of the reasoning. In Mizar the stress is put on three main issues:

- writing readable proofs using classical logic,
- possibility of cooperation with external theorem provers, and
- knowledge reuse (increasing possible connections between various developments, called *integrity of a repository*).

The outline of the paper is as follows. In Section II, we describe the history of formalizing geometry in Mizar, Section III presents the discussion on abstract and concrete mathematics; the next one outlines some basic constructions needed to understand our work, that is concrete translation of chosen properties straight from Tarski's axioms (A1)–(A7). In Section V and VI we give some insight on knowledge reuse and on its readability, respectively, while in the last part we describe related work, then we draw some concluding remarks and propose some future work.

II. Mizaring Affine Geometry

In Tarski's system of axioms [44] the only primitive geometrical notions are points, the ternary relation B of "soft betweenness" and quaternary relation $\equiv$ of "equidistance" or "congruence of segments". The axioms are reflexivity, transitivity, and identity axioms for equidistance; the axiom of segment construction; reflexivity, symmetry, inner and outer transitivity axioms for betweenness; the axiom of continuity, and some others. The original set consisted of 20 axioms for two-dimensional Euclidean geometry and was constructed in 1926–27, submitted for publication in 1940, and finally appeared in 1967 in a limited number of copies. There are many modifications of this system, and Gupta's work in this area [20] offers an important simplification. The strict betweenness was studied even before: it gave rise to "betweenness geometry" by Veblen in 1904.

Another notable axiomatization, proposed by Hilbert [21] in 1899, has three sorts: planes, points, and lines, and three relations: betweenness, containment, and congruence. In this sense it is a little bit more complex than Tarski's (but not necessarily in terms of numbers of axioms as it has also 20 of these). The two approaches establish a geometrical framework, in which theorems can be proven logically (remember Euclid's *Elements* proofs are mainly pictures or graphical constructions and rely heavily on the intuition). This allows to use a computerized theorem prover in order to find the proof or proof checker to check the theory for its correctness. In the paper, we deal with the proof checker Mizar, based on classical first order logic and Tarski-Grothendieck set theory (a variant of Zermelo-Fraenkel). Using this tool we describe, how Tarski's theory was built formally.

There are two significant connections of Andrzej Trybulec, the founder of the Mizar project, with persons involved in the area of Tarski's geometry. The first one and very influencial was the cooperation with Lesław W. Szczerba, the author of [42], in the early 80s of the previous century. Szczerba, who also submitted to the Mizar Mathematical Library twice, was at that time a head of the Institute of Mathematics in Białystok, Poland; the place the Mizar system was mainly developed (and that was also the affiliation of Trybulec). These contacts resulted in the research on the theory of interpretation and semantic foundations of logic in the sense of Epstein. At that time, Trybulec himself was not very active in formalizing geometry. The other connection was that after finishing his study in mathematics at the University of Warsaw under the guidance of Karol Borsuk (famous Polish topologist), Andrzej Trybulec took the position of an assistant in the Chair of Geometry, where Wanda Szmielew was a professor.

Although the very first approach to formal geometries we can consider the work of Wojciech A. Trybulec `INCSP_1` [45], at the state of its writing it was not tightly connected with the rest of the formal appproach to geometry in Mizar. Krzysztof Prażmowski (who is currently the head of the Institute of Mathematics, University of Białystok, Poland) created quite an active research group of fifteen people in the field of automated deduction in geometry, with the extensive use of the Mizar system. Main authors of these contributions were Krzysztof Prażmowski, Henryk Oryszczyszyn, and Wojciech Leończuk, all from former University of Warsaw, Białystok Branch. Together with the other authors, e.g., Kusak, Skaba, Muzalewski, and others, they wrote 43 Mizar articles on geometry (5 were removed later in the process of revisions). At the very beginning, there were many independent paths of formal development of various geometries (which expressed in various Mizar structures: directed vs. undirected parallelity relation, orthogonality, etc.).

Statistical data from Table I are not very impressive with respect to the current state of MML. The number of authors is quite big (six percent of all the authors of MML), hence the style of geometry in Mizar is not very uniform. The writing style can be measured by the percentage of zipped kBytes, which is higher than for the whole MML. This specific ratio (zipped instead of unpacked bytes) is usually taken when computing the so-called de Bruijn factor, showing how much additional code we should write comparing with informal proof to be fully understandable by computers. Furthermore, hierarchy of geometric objects is based on eight various structures, so it raises communication issues between various

TABLE I
STATISTICAL DATA ON THE FORMAL DEVELOPMENT OF GEOMETRY WITHIN MML

	Geometry	MML	Percentage
files	38	1,254	3.0
authors	15	255	6.0
kBytes	1.981	93.324	2.1
zipped kBytes	406	17,642	2.3
definitions	294	11,606	2.5
theorems	1,319	56,547	2.3
attributes	135	3,079	4.4
clusters	113	13,071	0.9
structures	8	157	5.0

approaches. The last article was PROJPL_1 dated back to 1994, and the series was not very actively developed (with the exception of the paper of the combinatorial Grassmannians COMBGRAS). But in 1990 Mizar articles on geometry were about 30% of the whole Mizar repository (out of 140 files). Of course, after that time revisions of this specific area were quite active: one of the main streams (done also by the current author) was to get separate axioms of selected properties instead of a large block for *the mode* (i.e., the constructor of the type in Mizar). As we can read from the table, the use of clusters (relatively new feature of Mizar language) is still very low comparing with the rest of MML, so there is a lot of work to be done. Affine approach to geometry was less important as there was another big challenge which for fifteen years stimulated geometry (in analytical setting, however): the proof of the Jordan Curve Theorem [23], which will be recalled in Section V.

As notable affine geometrical facts already formalized in Mizar we can enumerate:

- Hessenberg's theorem – HESSENBE;
- Desargues theorem (present in "Top 100 mathematical theorems") – ANPROJ_2;
- Pasch configuration axioms – PASCH;
- Fano projective spaces – PROJRED1;
- Desarguesian projective planes – PROJRED2;
- Pappus, Minor, Major and Trapezium Desargues axioms – AFF_2;
- Minkowskian geometry – ANALORT.

The "Top 100 mathematical theorems" list of important theorems in mathematics, proposed by Paul and Jack Abad at the end of provious century, is the popular collection of challenges for contemporary computerized proof-assistant systems; the list maintained by Freek Wiedijk described in [48] is available at http://www.cs.ru.nl/~freek/100/. Currently, the Mizar system holds the third position with 64 items formalized so far.

In our opinion, the unifiying approach in Tarski's spirit could be quite useful to bind all of these geometrical results together. Among another significant facts in geometry we can point out Morley trisector theorem [8], Ceva and Menelaus theorem [41]. These facts however deal with Euclidean plane, so the proofs are in the area of analytic geometry; they were developed more than ten years later than the foundations of geometry in Mizar, when Euclidean spaces were more thoroughly explored in MML.

TABLE II
THREE MAIN SECTIONS OF THE MIZAR MATHEMATICAL LIBRARY

Part of MML	Files	%	kBytes	%
classical part	323	25.8	20,013	21.8
abstract part	866	69.0	67,996	74.0
SCM	65	5.2	3,837	4.2
Total	1254	100.0	91,846	100.0

III. THE CHOICE OF FORMAL FRAMEWORK: CLASSICAL VS. ABSTRACT MATHEMATICS

The distinction between classical and abstract mathematics (i.e. the one based on ordinary axioms of set theory, and all those using the notion of a structure, respectively) is important from the viewpoint of the organization of the Mizar repository. We had to choose between two paths:

- it is possible to formulate Tarski's axioms without the use of a structure, and also set theory could be meaningless for that framework, only the classical logic with Mizar predicates is enough (equality plays a special role in the system based on set theory [18]);
- the use of Mizar structures forces us paradoxically to use basics of set theory – defining a signature of Tarski's plane needed to give a type of congruence of segments and betweenness relation, which was set-theoretic (Mizar language is typed, and in the earlier case one should also give a type at least to points, but it can be defined as Mizar object).

The latter was also chosen by us as the whole geometry in MML is written in abstract style (as the majority of MML, as you can read from Table II) as structures in Mizar are present for a long time. Even if in ordinary mathematical tradition they are considered as ordered tuples, in the implementation in Mizar they are treated rather as partial functions, with selectors as arguments, and ordinary inheritance mechanism (with polymorphic enabled, which will be extensively used in our formalizations). The details of the implementation can be found in [16], we give here only general syntax of declaring Mizar structure:

```
struct (Predecessor_List) Structure_Name
  (# selector_1 -> type_1,
     selector_2 -> type_2,
                ...
     selector_n -> type_n #);
```

To every argument, its type should be declared, which corresponds to ordinary definition of the signature of an algebra.

In the contemporary MML, the basic Mizar article devoted to structures is [26]; it is dated as for 1995, earlier than the first article from MML on structures [45], i.e. 1989, because it was created much later as a result of revision. Its main step was introducing common predecessor of all structures – 1-sorted, and the name of the selector carrier was

chosen. Wojciech A. Trybulec proposed in [45] the formalization of Hilbert axioms (hence we have three sorts: points, lines and planes). Interestingly, the original name `Points` remained untouched (although many similar approaches were later unified).

```
definition
  struct IncProjStr
     (# Points, Lines -> non empty set,
           Inc -> Relation of
             the Points, the Lines #);
end;

definition
  struct (IncProjStr) IncStruct
   (# Points, Lines, Planes -> non empty set,
   Inc -> Relation of the Points, the Lines,
   Inc2 -> Relation of the Points, the Planes,
   Inc3 -> Relation of the Lines, the Planes #);
end;

definition
  let S be IncProjStr;
  mode POINT of S is
         Element of the Points of S;
  mode LINE of S is
         Element of the Lines of S;
end;
```

Introducing two different structures to host only points and lines in the first case was a result of revision allowing to use simpler structures to cope with planar geometry. Note however, that this disagrees with Tarski's single sort universe (of points); and this is quite a basic structure in MML, `1-sorted`, as we already mentioned. Based on this, other descendant objects can be defined:

```
definition
  struct (1-sorted) AffinStruct
    (# carrier -> set,
       CONGR -> Relation of
         [:the carrier,the carrier:] #);
end;
```

This was the very first approach, and in fact the article with MML identifier `ANALOAF` offered quite simple structure; here, `CONGR` is a relation on the Cartesian square of the carrier, and it is used mainly in the context of parallelity predicate. Merging it with another relation, the orthogonality, MML offers a variety of affine geometries. The outline of this universe is shown in Figure 1. It can be observed that the hierarchy is not very deep, as many approaches were done originally in parallel, without reusing notions or theorems between various paths of development. Furthermore, incidence structures are not descendants of `1-sorted`, which should be definitely corrected in the future.

```
definition
  struct (AffinStruct, OrtStr) ParOrtStr
              (# carrier -> set,
    CONGR, orthogonality -> Relation of
      [:the carrier,the carrier:] #);
end;
```

In the original form of [45], the number of axioms was introduced in the form of a single big formula introducing `mode IncSpace`. What is interesting, after the revision, and evolution of the Mizar system, the file is 10 kB smaller.

```
definition
  struct (1-sorted) TarskiPlane
    (# carrier -> set,
    Betweenness -> Relation of
      [:the carrier, the carrier:], the carrier,
    Equidistance -> Relation of
      [:the carrier, the carrier:],
        [:the carrier, the carrier:] #);
end;
```

Observe that the choice of the arity of the relation is really meaningless here: the betweenness relation can be treated as a ternary relation; but the choice of this concrete model was quite arbitrary as the difference between dealing with ternary relations and relations between ordered pairs and elements will not cause any major problems later (we use mainly predicates).

```
definition
  let S be TarskiPlane;
  mode POINT of S is Element of S;
end;

definition
  let S be TarskiPlane;
  let a, b, c be POINT of S;
  pred between a,b,c means
:: GTARSKI1:def 1
  [[a,b],c] in the Betweenness of S;
end;

definition
  let S be TarskiPlane;
  let a, b, c, d be POINT of S;
  pred a,b equiv c,d means
:: GTARSKI1:def 2
  [[a,b],[c,d]] in the Equidistance of S;
end;
```

One can mention possibly misleading type definitions for the betweenness relation (it can be considered as three-argument relation than as it is now, i.e. relation between an ordered pair and an element), and also – pretty technical – predicate `between a,b,c` for arbitrary points a, b, c. Of course, the notation could be changed easily into well-readable candidate `b is_between a,c` (maybe with the need for replacing arguments), but we have chosen this notation to be closer to SST. Brackets can be used as `between (a,b,c)`, but they can also be omitted, there is no strict obligation for using them.

IV. The Original Self-Contained Approach

In the very first version of Mizar formalization of Tarski's axioms done by William Richter with the help of `miz3`, we can find the remark:

> *In Mizar it isn't possible to define such a type (or model) without proving that some model exists. Trybulec's existence proofs runs over 450 lines. So we define a predicate 'S Tarski-model' which means that the plane S satisfies the axioms A1–A7, and*

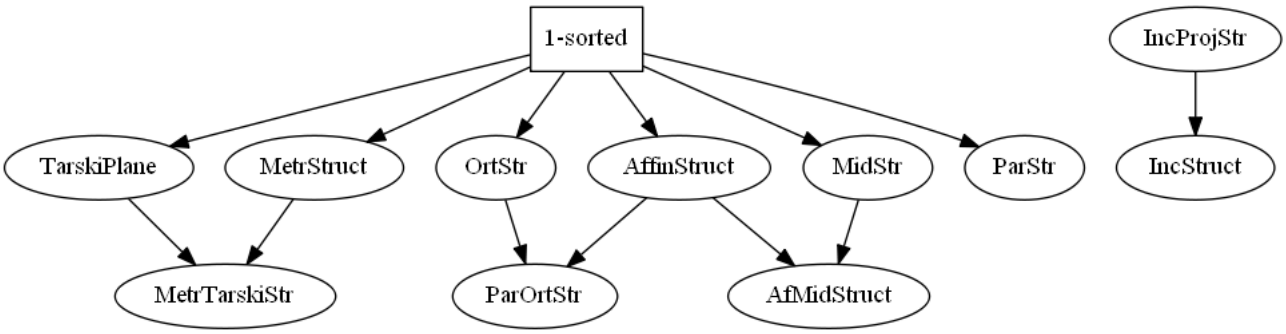


Fig. 1. Geometrical structures in the Mizar Mathematical Library

then prove trivial theorems A1–A7 which say that if S Tarski-model, then S satisfies an axiom A1–A7. The extra clutter involving the predicate Tarski-model, and the label TarskiModel which stands for the statement 'S Tarski-model' could be avoided by loading all our results into one gigantic theorem. Our approach seems preferable.

This motivation essentially caused the lack of the appropriate Mizar type. Indeed, Trybulec's existence proof gets even more lines (over 500), but it should be taken into account that [45] was one of the very first Mizar articles submitted to MML (numbered #25) and auxiliary set of handy lemmas or models was really modest at that time.

One can consider having predicates, but attributes instead of them seems to be better idea: we can have modular building of a complex structure, all other can be reused; hence in our Mizar article we focus on pure betweenness-equidistance part, not really mentioning the question of dimensions. We prove 44 theorems (properties of the predicates), with the Gupta's proof of Hilbert's I1 axiom (that two distinct points determine a line).

```
theorem ::: I1:
  a <> b & x <> y &
    a on_line x,y & b on_line x,y
      implies x,y equal_line a,b
   proof
     assume
H1:  a <> b & x <> y; then
P2:  b,a equal_line a,b by LineEqA1;
     assume
H2:  a on_line x,y & b on_line x,y;
     per cases;
     suppose x = b; then
       x,y equal_line b,a by H1, H2, I1part2;
       hence thesis by P2;
     end;
     suppose
       x <> b; then
P4:    x,y equal_line x,b by H2, I1part2; then
       x,b equal_line a,b
         by H1, I1part2Reverse, H2;
       hence thesis by P4;
     end;
  end;
```

TABLE III
THE STATISTICS OF THE CONTENT OF [38]

Items	Numbers
attributes	10
lines	1522
kBytes	50
theorems	47

Final Gupta's proof of I1 (GTARSKI1:46) above uses additional auxiliary predicates, e.g. `on_line`, which shortens the notation and provides future connections with Hilbert's axioms. Of course, at the end we are left with the proofs that the set of formulas (obtained by the so-called *definitional expansions* [24]), chosen formula can be deducted, but our attribute-steered approach seems to be better, not in terms of the efficiency of proving, but readability and usefulness for human mathematician.

Our Mizar versions of Tarski's axioms have descriptive names, and follow the ones from SST (using $\equiv$ for congruence of segments and B for betweenness relation):

- `CongruenceSymmetry` (A1): $\forall_{a,b}\ ab \equiv ba$,
- `CongruenceEquivalenceRelation` (A2): $\forall_{a,b,p,q,r,s}\ ab \equiv pq \wedge ab \equiv rs \Rightarrow pq \equiv rs$,
- `CongruenceIdentity` (A3): $\forall_{a,b,c}\ ab \equiv cc \Rightarrow a = b$,
- `SegmentConstruction` (A4): $\forall_{a,q,b,c} \exists_x\ B(q,a,x) \wedge ax \equiv bc$,
- `BetweennessIdentity` (A6): $\forall_{a,b}\ B(a,b,a) \Rightarrow a = b$,
- and `Pasch` (A7): $\forall_{a,b,p,q,z}\ \ B(a,p,z) \wedge B(b,q,z) \ \Rightarrow\ \exists_x\ \ B(p,x,b) \wedge B(q,x,a)$.

One attribute has the form which substantially differs from SST version: in order to shorten the notation, we introduced technical predicate

```
definition
  let S be TarskiPlane;
  let a, b, c, x, y, z be POINT of S;
  pred a,b,c cong x,y,z means
:: GTARSKI1:def 3
    a,b equiv x,y &
```

```
    a,c equiv x,z &
    b,c equiv y,z;
end;
```

denoting essentially SSS predicate for triangles. Using this notion, SST axiom (A5) could be encoded as follows:

```
definition let S be TarskiPlane;
  attr S is satisfying_SAS means
:: GTARSKI1:def 9
     for a, b, c, x, a1, b1, c1, x1
       being POINT of S holds
        a <> b & a,b,c cong a1,b1,c1 &
        between a,b,x & between a1,b1,x1 &
        b,x equiv b1,x1 implies
          c,x equiv c1,x1;
end;
```

that is,

$$\forall_{a,b,c,x,a',b',c',x'}\ (a \neq b \wedge ab \equiv a'b' \wedge bc \equiv b'c' \wedge ac \equiv a'c' \wedge$$
$$\wedge\ B(a,b,x)\ \wedge\ B(a',b',x')\ \wedge\ bx \equiv b'x')\ \Rightarrow\ cx \equiv c'x'.$$

It should be noted that the sequence of internal Mizar utilities can be run on the code, and hence proofs can be modified or even shortened. The remarkable case are the three assumptions from auxiliary lemmas added in the original `miz3` code, which was automatically removed as unnecessary. The usual running cycle of auxiliary programs like `relprem`, `relinfer`, `chklab`, and `inacc` is as follows: first, unnecessary premises are marked, then irrelevant (obvious for the Mizar checker) proof steps are marked as unused and eliminated, and finally the whole structure of the proof can be pretty-printed. Such "cleaning" cycle [16] caused the reduction of proofs of properties by more than 300 lines, not affecting readability that much. The net of connections between definitions, lemmas, and theorems obtained in this way can be further studied in order to get better refactoring of knowledge via dedicated techniques [17].

Having separate attributes for distinct axioms had already shown its usefulness in various geometrical settings. It could also allow later for defining equivalent axiom systems for Tarski geometry (and due to mechanism of clusters this equivalence will be obvious for the checker, once proven).

V. Introducing Metric Structure

It is well known fact that every metric space can be equipped with the natural topology. This informally obvious mathematical property brings some unexpected difficulties when dealing with structures if automated proof-assistants play a role. Namely then, if one considers topological spaces in a quite natural way, that is $\langle U, \tau \rangle$, and metric space as $\langle U, d \rangle$, respectively, one can rather naturally merge both structures into common $\langle U, \tau, d \rangle$.

During the formalization of the Jordan Curve Theorem in Mizar [23], however, another approach was chosen (and pushed consequently until the successful finale): `Euclid 2` denoted metric space concerned with the Euclidean plane, and then special functor converting any metric space into the topoogical space was applied to obtain `TOP-REAL 2` (of course the conversion can be made for arbitrary natural number n, not necessarily 2, but Jordan curves deal with two-dimensional case). The basic signature for metric spaces are `MetrStruct`, where distance is a function defined on the Cartesian square of the carrier with the real values. Then, metrics (or pseudo-, quasi-, semimetrics, etc.) can be defined in terms of attributes [22], that is properties of the distance function.

```
definition
  struct (MetrStruct,TarskiPlane) MetrTarskiStr
   (# carrier -> set,
     distance -> Function of
            [:the carrier, the carrier:], REAL,
     Betweenness -> Relation of
            [:the carrier, the carrier:],
              the carrier,
     Equidistance -> Relation of
            [:the carrier, the carrier:],
              [:the carrier, the carrier:] #);
end;
```

Then we have two worlds merged: affine, where we have two Tarski's relations, and Euclidean, where in terms of distance function, we can have betweenness relation and the the measure for segments. We argue that, regardless of all the complications caused by merging structures [12] (which increases the number of selectors, hence the chain of notions gets more complicated), such approach – not converting between two contexts, but rather to make reasoning in the world which is successor of both – allows for more flexible reuse of knowledge from two original areas.

```
definition let M be MetrTarskiStr;
  attr M is naturally_generated means
:: GTARSKI1:def 15
    (for a, b, c being POINT of M holds
      between a,b,c iff b is_Between a,c) &
    (for a, b, c, d being POINT of M holds
      a,b equiv c,d iff
        dist (a,b) = dist (c,d));
end;
```

In merged structure, we want to have two segments congruent, if they are equal in terms of distance, and betweenness relation uses the predicate `is_Between`, also given in terms of the sum of distances. To construct a model of the Tarski's space, such "one gigantic theorem" mentioned by Bill Richter in the quotation from the previous section can be somehow avoided. In Trybulec's paper [45], the massive construction of the appropriate space satisfying all incidence axioms was hidden under the existence proof of the following Mizar type:

```
registration
  cluster strict IncSpace-like for IncStruct;
end;
```

Recalling the previous discussion on the structure merging, the construction of such a space from scratch (essentially more or less modified copy-and-paste work on the proof from [45]) can be avoided with the help of some useful tricks, investigating knowledge already present in MML with its possible reuse. For example, based on the geometry on the real

line, appropriate geometrical structure from metric structure can be just an extension with properly defined distance.

```
definition
  func TarskiSpace -> MetrTarskiStr equals
:: GTARSKI1:def 22
    the naturally_generated TarskiExtension
      of RealSpace;
  coherence;
end;
```

Then, we can show that in such extension the metric is well defined, i.e. this space is reflexive, symmetric, and discerning. Furthermore, if we take into account "geometrical part", the following axioms can be proven true:

```
registration
  cluster TarskiSpace ->
    satisfying_CongruenceSymmetry
    satisfying_CongruenceEquivalenceRelation
    satisfying_CongruenceIdentity
    satisfying_SegmentConstruction
    satisfying_BetweennessIdentity;
```

VI. Human-oriented Representation of Knowledge in *Formalized Mathematics*

Of course, detailed proofs of all the work are already present in the Mizar Mathematical Library and were published in the special issue of *Formalized Mathematics* devoted to 25 years of MML. We tried to find subtle compromise between strict mathematical notation and a fluent text looking quite as good as that written by human – as the author has the opportunity of chosing the translation formats for newly introduced definitions (in XML stylesheet); the example of two automatically translated theorems are given on Figure 2 (versions straight from MML and FM).

```
theorem :: GTARSKI1:31 ::: GuptaEasy:
  a <> b & between a,b,c & between a,b,d &
    b <> c & b <> d implies not between c,b,d;

theorem :: GTARSKI1:32
  a,b,c cong a9,b9,c9 &
    between a,x,c & between a9,x9,c9 &
      c,x equiv c9,x9 implies b,x equiv b9,x9;
```

VII. Related Work

Although the history of the development of the axiom system for geometry by Tarski is not very clear from the beginnings (as the early works by Tarski seem to be postponed by the World War II), and even if the ultimate source of information is SST which was badly printed (the book was recently reissued with the foreword of Michael Beeson), now it attracted a lot of focus from automated deduction systems.

The remarkable item here is of course Julien Narboux's formal development of Tarski's system with the use of Coq [33]. We are planning to include some interesting results from GeoCoq in the Mizar system; Nakasho's et al. MML symbol reference system [32][1] offers quite user-friendly interface for browsing appropriate content. Of course, GeoCoq provides ready-to-use list of notions, which are connected only with the Tarski's system – that makes the browsing more convenient. Of course, after rescaling all definitions and theorems can be browsed within MML providing a look in the style of GeoCoq project.

Similar web-browsing system giving practical insight into Tarski's geometry is offered by the aforementioned Michael Beeson's *Tarski Formalization Project*. It offers linkages with underlying items from SST; the correctness relies on the Otter prover. Here the readability is not the main issue, however a few open questions was answered. Tim Makarios used Isabelle to formalize elliptic geometry in order to provide independence results for the parallel postulate [28]. Accidentally, this is one the items from the "Top 100 mathematical theorems", which is also challenging for us. It can be treated as a further development of axiomatical approach to geometry in Isabelle started by Meikle [31] (in Hilbert's *Foundations of Geometry* [21] style).

As we are concerned with the integrity issues for the Mizar Mathematical Library, we do not want to make basics from scratch, especially if the strong fundamentals are already done. There a need for further careful study of how much can be done in the direction of unveiling the knowledge included in the repositories of the Mizar system. Due to mechanism of revisions it is not the case that this work is just useless as it is very old. Mizar geometry just used the other way.

There are many simplifications of the original system by Tarski; e.g. by Tarski himself and his collaborators, Gupta's [20], or Makarios [28]. Also showing the correspondence between various axiom systems (with Hilbert's at the very beginning) for geometry is quite influencial and here provers can be quite useful. Of course, one of the well-advertised areas is the topic of axiom system for various equationally-defined classes of algebras, with the leading problem by Robbins and its solution by William McCune obtained in 1996 with the help of EQP/Otter equational theorem prover. But the question remains, how to cope with the space between obviousness for the human and for the machine, in the time before *QED Singularity*, as Michael Beeson (at QED+20 workshop in Vienna, July 18, 2014 [4]) called this time when formal proofs will be the norm in mathematics.

VIII. Conclusions and Future Work

In the paper, we have shown how already quite well established repository can be enhanced to cope with new capabilities of the Mizar system. We mention here a new approach to structures, including their merging, extensive use of attributes, and implemented automatization of definitional expansions. Also the issues of the integration with external provers should be taken into account.

We have created in [38] (and recent `GTARSKI2` [9]) complete formal axiomatization of Tarski's geometry which, at least in our opinion, has the advantage of higher readability for ordinary mathematicians than, e.g., Coq or Prover9 proof objects. In the same it is tightly connected with another

[1]The system can be browsed at http://webmizar.cs.shinshu-u.ac.jp/mmlfe/current/ with the official version of the Mizar system.

(31) If $A \neq B$ and B lies between A and C and B lies between A and D and $B \neq C$ and $B \neq D$, then B does not lie between C and D. The theorem is a consequence of (30), (21), and (18).

(32) Suppose $\triangle ABC \cong \triangle A'B'C'$ and X lies between A and C and X' lies between A' and C' and $\overline{CX} \cong \overline{C'X'}$. Then $\overline{BX} \cong \overline{B'X'}$. The theorem is a consequence of (5), (11), (8), (6), (12), (25), (7), (16), and (19).

Fig. 2. Screenshot of an excerpt from *Formalized Mathematics*

axiomatization of Euclidean plane, due to Hilbert, already available in MML. We have shown also that the real Euclidean plane satisfies all Tarski's axioms.

Of course, one of our aims is to do similar work to that of Michael Beeson's – the use of Prover9 to formalize practically everything from SST. But taking into account that Prover9 proof object can be relatively easily converted into working (and correct) Mizar proofs, it is important to construct appropriate background in order to have something which can offer a kind of help for a mathematician by the following requirements:

- it will reflect all theorems from SST;
- it will offer relatively high readability;
- the proofs can help to catch the idea of a proof, not necessarily will be just *l'art pour l'art*.

Translations from OTTER (at that time, currently Prover9) were helpful in the process of building various complemented lattices: Sheffer-stroke based and those needed in the Robbins algebras [13]. Also encodings of various short axiomatizations for Boolean algebras were easier with that tool; also in the area of rough sets [14] some results were obtained in that way [11]. On the other hand, fuzzy sets, which are much closer to set theory than rough sets, can be also nicely formalized [15].

We are also interested in building a Mizar model of elliptic geometry; Japanese team did some introductory work in [10], but it is not quite feasible now. This could help in showing that parallel postulate is independent from the other ones, which will solve another item from "Top 100 mathematical theorems", which is also an important issue. Further research on logical reasoning systems (as in the case of Tarski's system formulable in first-order logic with identity, set theory is not required) formalized by means of automated proof-assistants can be fruitful not only in the area of pure or applied mathematics, but also, e.g., in the area of argumentation theory, including legal expert systems [49].

We should definitely also have in mind the usefulness of the approach in the context of didactic use of the Mizar system. In order to achieve this goal, the existing state of geometry in MML should be made definitely more coherent, without the need of the creation of new special encyclopaedic articles which could serve better for students.

REFERENCES

[1] J. Alama, M. Kohlhase, L. Mamane, A. Naumowicz, P. Rudnicki, and J. Urban: Licensing the Mizar Mathematical Library, Proceedings of Mathematical Knowledge Management 2011, *Lecture Notes in Artificial Intelligence,* 6824, pp. 149–163 (2011) http://dx.doi.org/10.1007/978-3-642-22673-1_11

[2] J. Avigad, E. Dean, and J. Mumma: A formal system for Euclid's Elements, *Review of Symbolic Logic,* 2 pp. 700–768 (2009) http://dx.doi.org/10.1017/S1755020309990098

[3] G. Bancerek, Cz. Byliński, A. Grabowski, A. Korniłowicz, R. Matuszewski, A. Naumowicz, K. Pąk, and J. Urban: Mizar: state-of-the-art and beyond, Intelligent Computer Mathematics, *Lecture Notes in Computer Science* 9150, pp. 261–279 (2015) http://dx.doi.org/10.1007/978-3-319-20615-8_17

[4] M. Beeson: Mixing computations and proofs, *Journal of Formalized Reasoning,* 9(1), pp. 71–99 (2016) http://dx.doi.org/10.6092/issn.1972-5787/4552

[5] M. Beeson and L. Wos: OTTER proofs in Tarskian geometry, in Proceedings of IJCAR 2014, *Lecture Notes in Computer Science,* 8562, pp. 495–510 (2014) http://dx.doi.org/10.1007/978-3-319-08587-6_38

[6] K. Borsuk and W. Szmielew: *Foundations of Geometry,* North-Holland (1960)

[7] G. Braun and J. Narboux: From Tarski to Hilbert, in *Automated Deduction in Geometry,* T. Ida and J. Fleuriot (eds.) Proceedings of ADG 2012 (2012) http://dx.doi.org/10.1007/978-3-642-40672-0_7

[8] R. Coghetto: Morley trisector theorem, *Formalized Mathematics,* 23(2), pp. 75–79 (2015) http://dx.doi.org/10.1515/forma-2015-0007

[9] R. Coghetto and A. Grabowski: Tarski geometry axioms – part II, *Formalized Mathematics*, 24(2) (2016). http://dx.doi.org/10.1515/forma-2016-0012.

[10] Y. Futa, H. Okazaki, D. Mizushima, and Y. Shidama: Operations of points on elliptic curve in projective coordinates, *Formalized Mathematics,* 20(1), 87–95 (2012) http://dx.doi.org/10.2478/v10037-012-0012-2

[11] A. Grabowski: Automated discovery of properties of rough sets, *Fundamenta Informaticae,* 128(1–2), pp. 65–79 (2013) http://dx.doi.org/10.3233/FI-2013-933

[12] A. Grabowski: Efficient rough set theory merging, *Fundamenta Informaticae,* 135(4), pp. 371–385 (2014) http://dx.doi.org/10.3233/FI-2014-1129

[13] A. Grabowski: Mechanizing complemented lattices within Mizar type system, *Journal of Automated Reasoning,* 55(3), pp. 211–221 (2015) http://dx.doi.org/10.1007/s10817-015-9333-5

[14] A. Grabowski: On the computer-assisted reasoning about rough sets, in *Monitoring, Security and Rescue Techniques in Multiagent Systems*, B. Dunin-Kęplicz, A. Jankowski, M. Szczuka (Eds.), Advances in Soft Computing, 28, 215–226 (2005) http://dx.doi.org/10.1007/3-540-32370-8_15

[15] A. Grabowski: On the computer certification of fuzzy numbers, in *Proceedings of 2013 Federated Conference on Computer Science and Information Systems, FedCSIS 2013,* M. Ganzha, L. Maciaszek, M. Paprzycki (Eds.), pp. 51–54, IEEE (2013)

[16] A. Grabowski, A. Korniłowicz, and A. Naumowicz: Mizar in a nutshell, *Journal of Formalized Reasoning,* 3(2), 153–245 (2010) http://dx.doi.org/10.6092/issn.1972-5787/1980

[17] A. Grabowski and Ch. Schwarzweller: Towards automatically categorizing mathematical knowledge, In *Proceedings of 2012 Federated Conference on Computer Science and Information Systems, FedCSIS 2015,* M. Ganzha, L. Maciaszek, M. Paprzycki (Eds.), IEEE, 63–68 (2012)

[18] A. Grabowski, A. Korniłowicz, and Ch. Schwarzweller: Equality in computer proof-assistants, in *Proceedings of 2015 Federated Conference on Computer Science and Information Systems, FedCSIS 2015,* M. Ganzha, L. Maciaszek, M. Paprzycki (Eds.), IEEE, 45–54 (2015) http://dx.doi.org/10.15439/2015F229
[19] A. Grabowski and Ch. Schwarzweller: Revisions as an essential tool to maintain mathematical repositories, *14th Symposium on Towards Mechanized Mathematical Assistants*, Calculemus'07/MKM'07, Lecture Notes in Computer Science, 235–249 (2007) http://dx.doi.org/10.1007/978-3-540-73086-6_20
[20] H.N. Gupta: Contributions to the axiomatic foundations of geometry, PhD thesis, University of California, Berkeley (1965)
[21] D. Hilbert: *The Foundations of Geometry,* Chicago: Open Court, 2nd ed. (1980)
[22] S. Kanas, A. Lecko, and M. Startek: Metric spaces, *Formalized Mathematics,* 1(3), pp. 607–610 (1990)
[23] A. Korniłowicz: Jordan curve theorem, *Formalized Mathematics,* 13(4), pp. 481–491 (2005)
[24] A. Korniłowicz: Definitional expansions in Mizar, *Journal of Automated Reasoning,* 55(3), pp. 257–268 (2015) http://dx.doi.org/10.1007/s10817-015-9331-7
[25] A. Korniłowicz: On rewriting rules in Mizar, *Journal of Automated Reasoning,* 50(2), 203–210 (2013) http://dx.doi.org/10.1007/s10817-012-9261-6
[26] Library Committee of the Association of Mizar Users, Preliminaries to structures, *Mizar Mathematical Library*, MML Id: `STRUCT_0` (1995)
[27] M. Lombard and R. Vesley: A common axiom set for classical and intuitionistic plane geometry, *Annals of Pure and Applied Logic,* 95, pp. 229–255 (1998)
[28] T. Makarios: A mechanical verification of the independence of Tarski's Euclidean Axiom, MSc thesis (2012)
[29] R. Matuszewski and P. Rudnicki: Mizar: the first 30 years, *Mechanized Mathematics and Its Applications,* 4(1), pp. 3–24 (2005)
[30] W. McCune: Prover9 and Mace4. Available at http://www.cs.unm.edu/~mccune/prover9/, 2005–2010.
[31] L. Meikle and J. Fleuriot: Formalizing Hilbert's Grundlagen in Isabelle/Isar, in Proceedings of TPHOLs'03, *Lecture Notes in Computer Science,* 2758, pp. 319–334 (2003) http://dx.doi.org/10.1007/10930755_21
[32] K. Nakasho and Y. Shidama: Documentation generator focusing on symbols for the HTML-ized Mizar library, Proceedings of the Conference on Intelligent Computer Mathematics, CICM 2015, *Lecture Notes in Computer Science,* 9150, pp. 343–347 (2015) http://dx.doi.org/10.1007/978-3-319-20615-8_25
[33] J. Narboux: Mechanical theorem proving in Tarski's geometry, in *Automated Deduction in Geometry*, F. Botana and T. Recio (eds.), Lecture Notes in Computer Science, 4869, pp. 139–156 (2007) http://dx.doi.org/10.1007/978-3-540-77356-6_9
[34] A. Naumowicz and A. Korniłowicz: A brief overview of Mizar, in *Theorem Proving in Higher Order Logics 2009,* S. Berghofer, T. Nipkow, Ch. Urban, M. Wenzel (Eds.), Lecture Notes in Computer Science, 5674, 67–72 (2009) http://dx.doi.org/10.1007/978-3-642-03359-9_5
[35] K. Pąk: Methods of lemma extraction in natural deduction proofs, *Journal of Automated Reasoning,* 50(2), 217–228 (2013) http://dx.doi.org/10.1007/s10817-012-9267-0
[36] A. Quaife: Automated development of Tarski's geometry, *Journal of Automated Resoning,* 5(1), pp. 97–118 (1989)
[37] W. Richter: Hilbert Axiom Geometry in HOL Light, http://github.com/jrh13/hol-light/tree/master/RichterHilbertAxiomGeometry/
[38] W. Richter, A. Grabowski, and J. Alama: Tarski geometry axioms, *Formalized Mathematics,* 22(2), pp. 167–176 (2014) http://dx.doi.org/10.2478/forma-2014-0017
[39] P. Rudnicki and A. Trybulec: On the integrity of a repository of formalized mathematics, in Proceedings of Mathematical Knowledge Management 2003, A. Asperti, B. Buchberger, O. Caprotti (eds.) *Lecture Notes in Computer Science,* 2594, pp. 162–174 (2003) http://dx.doi.org/10.1007/3-540-36469-2_13
[40] W. Schwabhäuser, W. Szmielew, and A. Tarski: *Metamathematische Methoden in der Geometrie,* Springer-Verlag (1983)
[41] B. Shminke: Routh's, Menelaus' and Generalized Ceva's Theorems, *Formalized Mathematics*, 20(2), pp. 157–159 (2012) http://dx.doi.org/10.2478/v10037-012-0018-9
[42] L.W. Szczerba: Tarski and geometry, *Journal of Symbolic Logic,* 51, pp. 907–912 (1986)
[43] A. Tarski: What is elementary geometry?, in *Studies in Logic and the Foundations of Mathematics,* North-Holland, pp. 16–29 (1959)
[44] A. Tarski and S. Givant: Tarski's system of geometry, *The Bulletin of Symbolic Logic,* 5(2), pp. 175–214 (1999)
[45] W.A. Trybulec: Axioms of incidence, *Formalized Mathematics*, 1(1), pp. 205–213 (1990)
[46] J. Urban and G. Sutcliffe: Automated reasoning and presentation support for formalizing mathematics in Mizar, *Lecture Notes in Computer Science,* 6167, Springer, 132–146 (2010) http://dx.doi.org/10.1007/978-3-642-14128-7_12
[47] F. Wiedijk: A synthesis of the procedural and declarative styles of interactive theorem proving, *Logical Methods in Computer Science,* 8(1):30 (2012) http://dx.doi.org/10.2168/LMCS-8(1:30)2012
[48] F. Wiedijk: Formal proof – getting started, *Notices of the American Mathematical Society,* 55(11), 1408–1414 (2008)
[49] T. Zurek, Modelling of a'fortiori reasoning, *Expert Systems with Applications,* 39(12), 10772–10779 (2012) http://dx.doi.org/10.1016/j.eswa.2012.02.188

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 383–392
DOI: 10.15439/2016F83
ACSIS, Vol. 8. ISSN 2300-5963

Modeling Co-Verbal Gesture Perception in Type Theory with Records

Andy Lücking
Goethe University Frankfurt
Robert-Mayer-Straße 10
D-60325 Frankfurt am Main, Germany
Email: luecking@em.uni-frankfurt.de

***Abstract*—In natural language *face to face* communication interlocutors exploit manifold non-verbal information resources, most notably hand and arm movements, i.e. gestures. In this paper, a type-theoretical approach using Type Theory with Records is introduced which accounts for iconic gestures within an information state update semantics. Iconic gestures are semantically exploited in two steps: firstly, their kinetic representations are mapped onto vector sequence representations from vector space semantics, modeling a perceptual gesture classification; secondly, these vectorial representations are linked to linguistic predicates, giving rise to a computational account to semantic-kinematic interfaces. Each of the steps involves reasoning processes, which are made explicit. The resulting framework shows how various resources have to be integrated in the update mechanism in order to deal with apparently simple multimodal utterances.**

I. Introduction

SEMANTIC theories, artificial intelligence and robotic systems developed for spoken face-to-face interaction eventually have to deal with non-verbal communication means like facial expressions, prosodic features, hand and arm gestures, or proxemic relations. This is because verbal and non-verbal means constitute an integrated communication system [1], [2]. Their tight coupling shows up strikingly in cases where non-verbal means are semantically significant, i.e. when they provide information beyond or even instead of the verbal one – respective data is given in Section II. We are concerned with *iconic* gestures in the sense of representational hand and arm movements in this paper, gestures which, roughly speaking, depict aspects of the scene talked about [3] (for an automatic gesture classification involving iconic ones see e.g. [4]). Such gestures are performed rather spontaneously and presumably do not obey formal constraints [2]. For this reason, they cannot be interpreted according to pre-defined lexical entries, contrary to emblematic gestures as, say, the *thumbs-up* symbol for indicating positive evaluation or agreement, or Karate postures [5] in physical instructions. Rather, the interpretation of such gestures is a challenge that is related to spatial perception [6]. Accordingly, a perceptually oriented iconic gesture classification is proposed which rests on an integration of various semantic resources, as envisaged by, e.g., [7]. The formal framework that provides a unified representational "home" for these resources is *Type Theory with Records* (TTR) [8]. In the following, TTR is applied to capture the semantic impact of co-verbal gestures by combining the following ingredients:

- a detailed kinematic gesture representation [9] – Section III-A;
- the gesture representation is mapped onto vector sequences from vector space semantics [10] – Section III-B;
- vector sequences are linked to the intensions of linguistic expressions along the lines of a formal semantics for perceptual classification [11], [12] – Section III-C;
- perceptual classification and linguistic semantics are finally related within a dynamic information state update semantics [11], [13] – Section III-C.

Finally, in Section IV the account is applied to the benchmark phenomena identified in Section II.

II. Some Data

The integration of speech and co-verbal gesture has been investigated, *inter alia*, by means of the (German) *Speech and Gesture Alignment Corpus* (SaGA) [9], from which the following examples are drawn. Examples are quoted according to their dialog number and start time (e.g. "V13, 3:36" means that the datum can be found in dialog V13 at minute 3:36 within the corresponding video file). Note that only the so-called *stroke* phase of gestures is considered here: it is assumed to be the "meaningful" part of a gestural movement, distinguished from a pre-stroke preparatory movement as well as from a post-stroke retraction movement, which may be required in order to bring hand and arms from an inactive rest position into an active position and back, respectively [14], [2]. The decorated screenshots are all taken from [12].

Fig. 1. *Staircases* (V10, 3:19)

978-83-60810-90-3/$25.00

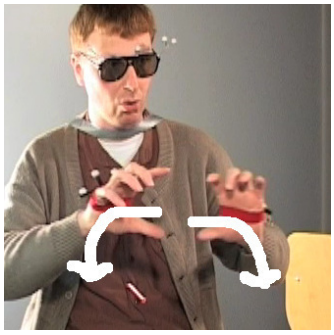


Fig. 2. *It has a concrete base* (V5, 0:39)

Fig. 3. *The house is like this.* (V11, 2:32)

In the first example – given in (1) –, which can be found in V10, 3:19, the speaker speaks about circular staircases. However, in her verbal description she just uses the hypernym *staircases*. The more specific circular information is provided by the affiliated gesture, which is shown in Fig. 1. The part of speech which roughly co-occurs with the gesture is indicated by brackets. Since the first syllable from the noun *Treppen* is not only part of the portion of speech which co-occurs with gesture but also has primary stress (indicated by capital letters), it is the first candidate for providing an integration point for gesture information [15], [12].

(1) *Ich g[laube das sollen TREP]pen sein*
I think that should staircases be

'I think that should be staircases' + Fig. 1

Thus, gestures can be used to specify linguistic expressions to their hyponym meanings.

In a similar manner, the gesture shown in Fig. 2 indicates the shape of a concrete base, which is introduced into dialog in the following way:

(2) *die Skulptur die hat 'n [BeTONsockel]*
the sculpture it has a concrete base

'the sculpture has a concrete base' + Fig.2

From the gesture, but not from speech, we get that the concrete base of the sculpture has the shape of a flat cylinder – the gesture acts as a nominal modifier. Note, however, that the gesture is incomplete since it only depicts about half of a cylinder [16]. Thus, its interpretation interacts with a *good continuation* extension known from gestalt theory.

In the datum given in (3), the speaker speaks about a U-shaped building. However, no shape predicate is given verbally, instead the full shape-representing burden is delegated to the gesture. The gesture in turn is produced within the scope of a verbal demonstrative, which induces a shift in focus to the speaker's gesture. In contrast to examples (1) and (2), the utterance in (3) is not even interpretable without the gesture.

(3) *dann ist das Haus halt so []*
then is the house just this []

'then the house is like this' + Fig. 3

These examples illustrate that the informational enrichment by gestures includes

- invoking hyponymic meanings of affiliated expressions,
- indicating linguistically unexpressed properties,
- providing complete demonstrations.

In the following sections, a type-theoretical model is given that aims at accounting for these gestural enrichments. This account extends previous accounts (most notably [12]) in that it implements a connection to dynamic semantic theories and provides a means for dealing with *good continuations* in formal analyses of co-verbal gestures for the first time. The focus on spontaneous iconic gestures goes beyond functionally restricted *click* or *draw* gestures as predefined in multimodal grammars or dialog systems [17], [18], [19] – see [20] for a comparison of various multimodal approaches.

III. A Type-Theoretical Account to Gestures

Type Theory with Records (TTR) [8] has been developed as a formal framework for natural language semantics which integrates insights from situation semantics [21], Discourse Representation Theory [22] and Montagovian λ-calculus [23]. The basic notion of TTR is a *judgment* of the form $a : T$, meaning that object a is of type T. In order to account for more complex kinds of judgments, TTR develops the notions of *record* and *record type*. The former are matrices of labels and objects, the latter are matrices of labels and types. Record types can be used to regiment records: a record r is of record type RT, $r : RT$, just in case each label of the record type also occurs in the record (the record may contain more fields, though) and the label assignments from the record obey the type constraints imposed by the record type, as schematically exemplified in (4).

(4) $\begin{bmatrix} l_1 & = & o_1 \\ l_2 & = & o_2 \\ l_3 & = & o_3 \end{bmatrix} : \begin{bmatrix} l_1 & : & T_1 \\ l_3 & : & T_2(l_1) \end{bmatrix}$,
just in case $o_1 : T_1$ and $o_3 : T_2(o_1)$.

Record types may depend on other records or record types. For example, type T_2 in (4) depends on object o_1. Dependent types can be used in semantics, for instance, to capture existence

presuppositions imposed by proper names, as illustrated in (5) by example of the name "Max":

(5) $\lambda r : \left[\text{x} \quad : \mathit{Ind}\right] \; . \left[\text{c}_{\text{pn}} \quad : \quad \text{named}(r.\text{x}, \text{“Max”})\right]$

The type in (5) is a functional type in which the value of the range depends on the value of the domain (which in turn is constrained to be of type *Ind(ividual)*) – see [8] for this and various others formal TTR notions.

A. A String Theory of Gesture Events

TTR comes with a string theory of events based on work by [24]. Basically, a(n) (complex) event is segmented into event "snapshots" that are combined by the string concatenator '^'. For example, the event e of opening a door involves the sequence of an agent x gripping the door handle a, pressing the handle and pushing the door b:

(6) $([e : \text{grip}(x, a)]^\frown[e : \text{press}(x, a)]^\frown[e : \text{push}(x, b)])$

Now, gestures can be considered to be events [12]. To begin with, a simple gesture can be represented as a record straightforwardly, as in (7):[1]

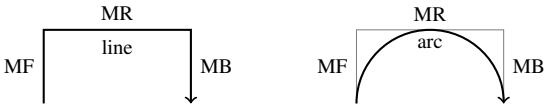
(7)
$$\begin{bmatrix}
\text{hand} & = & \text{right} & \\
\text{hs} & = & \text{claw} & \\
\text{carrier} & = & \begin{bmatrix} \text{boh} & = & \text{none} \\ \text{plm} & = & \text{none} \\ \text{wrst} & = & \text{MR>MB>ML} \\ \text{move} & = & \text{line>line>line} \end{bmatrix} \\
\text{sync} & = & \begin{bmatrix} \text{sloc} & = & \text{CBR-F} \\ \text{eloc} & = & \text{CBR-N} \\ \text{stime} & = & \text{2:32} \\ \text{etime} & = & \text{2:33} \end{bmatrix} \\
\text{rel} & = & \text{none} &
\end{bmatrix}$$

The record in (7) represents the gesture shown in Fig. 3 ('claw' is used to label the American Sign Language (ASL) hand-shape *bent 5*). The values from (7) come from gesture annotation and are imported into TTR as objects of type *annotation predicate* (AP) – see [9] for an overview of a kinetic gesture annotation of this kind. In order to prevent value mismatches like hand-shapes occurring as directions, respective sub-typing of annotation predicates may be employed. For instance, using a type hierarchy from unification-based grammars, the type *AP* can be extended in terms of several sub-types, corresponding to the different kinds of values [12]. By this means, the gesture record entries can be regimented quite specifically, for instance, all *carrier* fields can be required to consist only of movement predicates. However, this rather technical detail is ignored in the following for the sake of brevity and the general type *AP* is used throughout.

The mnemonic labels introduce values for the handedness ('hand'; *left*, *right* or *both*), hand-shape ('hs' according to the *American Sign Language* alphabet), the movement path (where the movement is carried out by one or more 'carriers' [26]) and the relation to the other hand ('rel') as well as the temporal and locational properties ('sync'), where locations follow the gesture space model from Fig. 4.

Gestural movement can be brought about by one or more of three possible movement carriers: back-of-hand (boh), palm (plm) or wrist (wrst). A movement is captured in terms of a *direction* seen from the speaker (e.g. *move forward* (MF)) and a concatenation type which distinguishes straight ('line') from roundish ('arc') trajectories. For example, the same sequence of direction labels, MF>MR>MB, can give rise to an open rectangle or a semicircle, depending on the type of concatenation, as illustrated in (8):

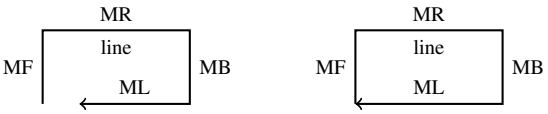
(8)



Note that the *sync*-feature's values allow to discriminate closed from incomplete shapes, which will be important for capturing gestalt properties (see Sec. IV below). For instance, the movement in (9) is underspecified with regard to the respective lengths of the movement parts. It can therefore represent both of the shapes illustrated in (10).

(9)
$$\begin{bmatrix} \text{wrst} & = & \text{MF>MR>MB>ML} \\ \text{move} & = & \text{line>line>line>line} \end{bmatrix}$$

(10)



The incomplete and the closed shape from (10) are distinguished in terms of their *sync* properties: closedness is defined as start (*sloc*) and end location (*eloc*) being the same, as expressed in the closeness condition **C-clos**.

C-clos:A gesture trajectory is closed iff start and end location are the same. The closure constraint has to distinguish one-handed from two-handed gestures:

- One-handed gesture: $\begin{bmatrix} \text{sloc} & : & AP \\ \text{eloc=sloc} & : & AP \end{bmatrix}$
- Two-handed gesture:
$$\begin{bmatrix}
\text{hands} & = & \text{both} & \\
\text{lh} & = & \left[\text{sync} : \begin{bmatrix} \text{sloc} : AP \\ \text{eloc} : AP \end{bmatrix}\right] \\
\text{rh} & = & \left[\text{sync} : \begin{bmatrix} \text{sloc=lh.sync.sloc} : AP \\ \text{eloc=lh.sync.eloc} : AP \end{bmatrix}\right]
\end{bmatrix}$$

The basic representation format introduced above describes gesture events in terms of their kinetic sub-events and can be hooked to the "string theory of gesture events" straightforwardly. The outright string representation for the closed path gesture from (9), for example, is given in (11), where the gesture gets the event variable e:

[1]Matrix-based representations of gestures have been used in robotics at least since [25] and can be considered a standard representation format for gestures already.

$$(11)\quad e : \begin{bmatrix}\text{wrst} = \text{MF} \\ \text{sync} = \begin{bmatrix}\text{sloc} = \text{p1} \\ \text{eloc} = \text{p2}\end{bmatrix}\end{bmatrix} \widehat{\textit{line}} \begin{bmatrix}\text{wrst} = \text{MR} \\ \text{sync} = \begin{bmatrix}\text{sloc} = \text{p3} = \text{p2} \\ \text{eloc} = \text{p4}\end{bmatrix}\end{bmatrix} \widehat{\textit{line}}$$
$$\begin{bmatrix}\text{wrst} = \text{MB} \\ \text{sync} = \begin{bmatrix}\text{sloc} = \text{p5} = \text{p4} \\ \text{eloc} = \text{p6}\end{bmatrix}\end{bmatrix} \widehat{\textit{line}} \begin{bmatrix}\text{wrst} = \text{ML} \\ \text{sync} = \begin{bmatrix}\text{sloc} = \text{p7} = \text{p7} \\ \text{eloc} = \text{p8} = \text{p1}\end{bmatrix}\end{bmatrix}$$

By and large, the string notation using '$\frown$' and the gesture annotation using '>' are equivalent. That is, any record of the form shown in (7) can be translated into the string format illustrated in (11) without loss of information. In order to account for straight and bend movements, however, a small modification to string concatenation has to be made, however: the (temporal) string concatenation '$\frown$' is bifurcated into two spatial variants, '$\widehat{\textit{line}}$' and '$\widehat{\textit{arc}}$'. The string representation of gestures facilitates a rather detailed descriptive resolution. For instance, in order to decide whether movements fragments compose into one gesture event or belong to different gestures, convention **C-loc** can be employed:[2]

C-loc: If the start location of a movement part is identical to the end location of the previous movement part, both movement parts are concatenated within one gesture.

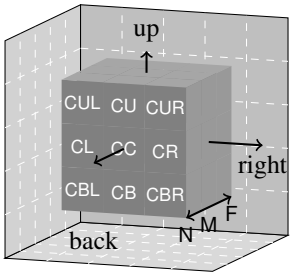
If **C-loc** is fulfilled, the more compact representation in (12) is preferred over the more detailed string representation (11), however:

$$(12)\quad \begin{bmatrix}\text{wrst} = \text{MF>MR>MB>ML} \\ \text{move} = \text{line>line>line>line} \\ \text{sync} = \begin{bmatrix}\text{sloc} = \text{p1} \\ \text{eloc} = \text{sloc} = \text{p1}\end{bmatrix}\end{bmatrix}$$

B. Trajectories within Vector Space

Given a kinetic representation format for gesture events, we need to spell out a semantic interpretation thereof in order to get access to the informativity of co-verbal gestures. To this end, a record type *Vec*(tor) is introduced, which provides an abstract model for configurations or trajectories. Via *Vec*, gesture representations are linked to a vector space semantics as developed by [28], [10]. The linking element in this mapping is the notion of *gesture space* [2], which refers to an inherently oriented space delimited by the reach of the speaker's arms. A respective gesture space model is illustrated in Fig. 4. The central anterior cube of this $3 \times 3 \times 3$-grid of cubes, which is labeled 'CC', is located right in front of the speaker's stomach. Each part of the space can be addressed by a name, which consists of a positioning plus a distance label. For instance, 'CUR-F' (*central upper right far*) is the topmost cube on the right, following the perspective employed in Fig. 4. These names can be used in order to provide values for the locational gesture representation fields conventionally labeled 'sloc' and 'eloc' – see example (7) above. The cube model also provides a regulating screw for the spatial granularity of the gesture space: the more cubes are employed, the higher the spatial resolution. The 27 different regions from the model in Fig. 4 are quite detailed already and are sufficient for present purposes.

[2]For a more sophisticated identity condition for gesture events see the criterion of [12], drawing on the event metaphysics of [27].



CBL: center below left
CL: center left
CUL: center upper left
CB center below
CC: center center
... ...
N: near
M: middle
F: far

Fig. 4. *Gesture Space Model* seen from speaker's perspective

TABLE I
DIRECTIONAL CONSTRAINTS DERIVED FROM THE SAGITTAL PLANE OF THE GESTURE SPACE (EXTRACT). THE VECTOR TRANSLATION v OF THE BASIC CONFIGURATIONS IS ALSO GIVEN.

Configuration	=	Vector π_v	→	Constraints π_d
Handshape $\in$ {C, 5, B, O, Y}	=	$\{\mathbf{u}\}$	→	volume
{MF, MR, MB, ML}	=	$\mathbf{u}$	→	translational
$\varnothing$	=	–	→	–
MF>MR + line	=	$\mathbf{u} \perp \mathbf{v}$	→	orthogonal
MF>ML + line	=	$\mathbf{u} \perp \mathbf{v}$	→	orthogonal
MF>ML + arc	=	$\mathbf{u} \circ \mathbf{v}$	→	quadrant
MF>MR + arc	=	$\mathbf{u} \circ \mathbf{v}$	→	quadrant
...	=	...	→	...
MF + ...+ MB	=	$\mathbf{u}, \mathbf{u}^{-1}$	→	inverse
ML + ...+ MR	=	$\mathbf{u}, \mathbf{u}^{-1}$	→	inverse
sloc = eloc	=	$\mathbf{u}(0) = \mathbf{v}(1)$	→	closed
sloc $\neq$ eloc	=	$\mathbf{u}(0) \neq \mathbf{v}(1)$	→	open
lh.sloc = rh.sloc +	=	$\mathbf{u}(0) = \mathbf{v}(0)$		
lh.eloc = rh.eloc [two-handed]	=	$\mathbf{w}(1) = \mathbf{x}(1)$	→	closed
quadrant + quadrant + invers				semicircle
semicircle + semicircle				circle
orthogonal + orthogonal + invers				rectangular
rectangular + rectangular				rectangle
...				...
translational + crossing planes				diag(onal)

The gesture model spans along the three body planes, transverse (up-down), saggital (left-right) and frontal (front-back). For each plane, rules for inferring vectorial constraints from kinetic gesture representations can be formulated. Additionally, some hand shapes function as "line-thickness modifiers" of the gesture trajectory, giving rise to a three-dimensional body rather than to a two-dimensional sketch (cf. the work of [29]). Some rules that will be used below are collected in Table I by example of the sagittal plane. The rules are the back-bone of the gesture vectorization function π – see (14) below.

The primitive mathematical notion of a vector is used to model *paths*, where a path is a function $\mathbf{p} : [0, 1] \mapsto \mathbf{V}$, $\mathbf{V}$ being a three-dimensional vector space (for reasons of

simplicity we constrain ourselves to a purely spatial model; accounting also for temporal aspects would require $\mathbf{V}$ to be four-dimensional). Simple paths (i.e. lines) can be described by one vector, more complex paths arise out of vector sequences (like MF>MR>MB, respectively its vectorization). Each single vector $\mathbf{u}$ from a vector sequence has its own length in $[0, 1]$, ranging from the origin $\mathbf{u}(0)$ to the end $\mathbf{u}(1)$ ($\mathbf{u}(0.5)$ denotes the half of $\|\mathbf{u}\|$). In a vector sequence like $\mathbf{u} \perp \mathbf{v}$ it holds that $\mathbf{u}(1) = \mathbf{v}(0)$, that is, the described path is consecutive. Additionally, there are various kinds of vectors, discriminating, amongst others, vectorial representations of locations (needed for example for prepositional modifications like "3 cm above x") and of shapes (figuring as spatial denotations of predicates like "round") [30]. Finally, the core vectorial representations are extended with some shape-related features like being translational or circular. These features are due to the work of [31] (who uses partly differently named features, though), where they are used as lexical constraints on path shapes of (mainly) motion verbs. Accordingly, *Vec* consists of three basic fields: *Vtype* determines the kind of vector sequence in question (i.e., axis, path, ...), *Vpath* stores the path's shape and *Vshape* introduces shape-related constraints. For two-handed gestures, each hand gives rise to a record of type *Vec*. In that case, an additional field (conventionally labeled "comb") also of type *Vec* is introduced, which captures hand-crossing, combined information (a simple example is given in Sec. IV by means of the *concrete base*-gesture from Fig. 2).

$$(13)\quad \mathit{Vec} =_{\text{def}} \begin{bmatrix} \text{vt} & : & \text{Vtype} \\ \text{pt} & : & \text{Vpath} \\ \text{sh} & : & \text{set(Vshape)} \end{bmatrix}$$

There is a functional type π which maps records of annotations (i.e., records with entries of type *AP*) – labeled Rec^{AP} for short – onto vectors (*Vec*). According to the division of labor between vectorization and representational feature decomposition (cf. Table I and the above-given explanation), π consists of two sub-types, π_v and π_d:

$$(14)\quad \pi : [[\pi_v : \mathit{Rec}^{AP} \rightarrow \mathit{Vpath}] \rightarrow \pi_d : \mathit{Vpath} \rightarrow \text{set(Vshape)}] \rightarrow \mathit{Vec}$$

Keeping vectorization and feature decomposition apart will also play a role in accounting for the *good continuation* in Sec. IV below.

The vector function π exploits the constrains from Table I and works schematically as follows:

$$(15)\quad \text{if } r\text{:} \begin{bmatrix} \text{hand} & : & \text{AP} \\ \text{hs} & : & \text{AP} \\ \text{carrier} & : & \begin{bmatrix} \text{boh} & : & \text{AP} \\ \text{plm} & : & \text{AP} \\ \text{wrst} & : & \text{AP} \\ \text{move} & : & \text{AP} \end{bmatrix} \\ \text{sync} & : & \begin{bmatrix} \text{sloc} : \text{AP} \\ \text{eloc} : \text{AP} \end{bmatrix} \end{bmatrix},$$

$$\text{then } \pi_v(r) = \begin{cases} \text{pt1} = \begin{bmatrix} v(r.\text{wrst},\ r.\text{move} \\ v(\text{sloc},\ \text{eloc})) \end{bmatrix} : \text{Vpath} & \text{if } r.\text{wrst} \neq \varnothing \\ \text{pt2} = \begin{bmatrix} v(r.\text{plm},\ r.\text{move} \\ v(\text{sloc},\ \text{eloc})) \end{bmatrix} : \text{Vpath} & \text{if } r.\text{plm} \neq \varnothing \\ \text{pt3} = \begin{bmatrix} v(r.\text{boh},\ r.\text{move} \\ v(\text{sloc},\ \text{eloc})) \end{bmatrix} : \text{Vpath} & \text{if } r.\text{boh} \neq \varnothing \end{cases}$$

$$\text{then } \pi_d(\pi_v(r)) = \begin{cases} \begin{bmatrix} \text{sh} = d(\text{pt1}) \end{bmatrix} : \text{set(Vshape)} & \text{if pt1} \neq \varnothing \\ \begin{bmatrix} \text{sh} = d(\text{pt2}) \end{bmatrix} : \text{set(Vshape)} & \text{if pt2} \neq \varnothing \\ \begin{bmatrix} \text{sh} = d(\text{pt3}) \end{bmatrix} : \text{set(Vshape)} & \text{if pt3} \neq \varnothing \end{cases}$$

Since a single gesture may involve more than one movement path – an example is given in Sec. IV below – π_v just numbers the 'pt' values from the possible movement carriers. The procedure is incremental in that immediate pairs as well as skip pairs of directional APs are considered and compared to the table entries above: that is, neighboring APs are evaluated in terms of being quadrant or orthogonal, longer sequences of APs are inspected for fulfilling inversion. The features collected in the constraints of Table I then give rise to shape descriptions. For instance, by dint of the cooperation of Table I and π, the sample gesture from (7) is translated into the following perpendicular vector sequence:

$$(16)\quad \text{a.}\quad \pi_v\left(\begin{bmatrix} \text{wrst} = \text{MR>MB>ML} \\ \text{move} = \text{line>line>line} \\ \text{sync} = \begin{bmatrix} \text{sloc} = \text{p1} \\ \text{eloc} = \text{p2} \neq \text{p1} \end{bmatrix} \end{bmatrix}\right) = \begin{bmatrix} \text{pt1} : \begin{bmatrix} \mathbf{u} \perp \mathbf{v} \perp \mathbf{w} \\ \mathbf{u}(0) \neq \mathbf{w}(\mathbf{1}) \end{bmatrix} \end{bmatrix}$$

$$\text{b.}\quad \pi_d\left(\begin{bmatrix} \text{pt1} : \begin{bmatrix} \mathbf{u} \perp \mathbf{v} \perp \mathbf{w} \\ \mathbf{u}(0) \neq \mathbf{w}(\mathbf{1}) \end{bmatrix} \end{bmatrix}\right) = \begin{bmatrix} \text{sh} : \{\text{rectangular, open}\} \end{bmatrix}$$

Note that the *Vtype* field is underspecified in (16) – there is nothing within the kinetic gesture representation that determines the kind of the vector. Vector types can be instantiated in interaction with the linguistic lexicon when combined with speech.

C. Linking Gesture Perception to Reasoning

In representing meaning, a dynamic information state update semantics as developed in [11] is used. There, the context (e.g., presuppositions) is explicitly separated from the content of linguistic expressions in terms of *backgrounded* (bg) and *foregrounded* (f) information. For instance, the proper name example "Max" from (5) above is recaptured as follows:

$$(17)\quad [\![\text{Max}]\!] = \begin{bmatrix} \text{bg} & = & \begin{bmatrix} \text{x} : \text{Ind} \end{bmatrix} \\ \text{f} & = & \lambda r : \text{bg}\left(\begin{bmatrix} \text{c}_{\text{pn}} : \text{named}(r.\text{x},\ \text{“Max”}) \end{bmatrix}\right) \end{bmatrix}$$

In order to implement the dynamic update of information, the functional type exemplified in (17) has to "accumulate" backgrounded information. This is expressed in terms of a *fixed point type* [8] – cf. also [11]. The fixed point type $\mathcal{F}$ corresponds to the unification of the domain and the range of a functional type. For instance, the fixed point type for the foregrounded meaning of "Max" from (17) is shown in (18):

(18) $\mathcal{F}([\![\text{Max}]\!].\text{f}) = \mathcal{F}(\lambda r : \left[\text{x} : \text{Ind}\right]\left(\left[\text{c}_{\text{pn}} : \text{named}(r.\text{x}, \text{“Max”})\right]\right))$
$= \begin{bmatrix} \text{x} & : & \text{Ind} \\ \text{c}_{\text{pn}} & : & \text{named}(\text{x}, \text{“Max”}) \end{bmatrix}$

In general, context update proceeds as formulated in **C-upc**, where agents' contributions extend their current information state (i.e. their "take of the situation" – see [11, p. 8]):

C-upc:**(preliminary version)** If the current information state s_t is compatible to the background information of expression e, then $[\![\text{e}]\!]$ can be added to s_t to form information state s_{t+1}:
If $s_t \sqsubseteq [\![\text{e}]\!].bg$ then $s_{t+1} = s_t \wedge \mathcal{F}([\![\text{e}]\!].\text{f})$

Operation '$\wedge$' in (18) is the type-theoretic analog to unification – see [8] for details. Gestures constitute a *display situation* (DP) [32] and as such are part of the publicly available information state. The value of DP is a list: every newly produced gesture is put into initial position while all possibly already present gestures are stacked and remain available for eventual repair or clarification.

C-upg:If $G = [v : Vec]$ is the vector interpretation of a gesture g produced at state s_t, the display situation list of s_t gets updated with G:
If $\pi(g) = G$ at s_t, then s_{t+1}.dp.rest = s_t.dp and s_{t+1}.dp.first = G

How is the link between display situations and dynamic meaning implemented? Following the exemplification [33] account of [12], space-related predicates are equipped with a *conceptual vector meaning* (CVM) which extends their intensions by means of vectorial constraints (cf. also the classification approach of [11] and the lexical decomposition approach of [34]). In particular, CVM information uses representations that are of type *Vec*. A spatial predicate like *U-shaped*, for instance, has the following extended dynamic meaning:

(19) $[\![\text{U-shaped}]\!] =$

$$\begin{bmatrix} \text{bg} & = & \left[\text{x} : \text{Ind}\right] \\ \text{f} & = & \lambda r : \text{bg}\left(\left[\begin{array}{l} \text{c}_{\text{u}} : \text{U-shaped}(r.\text{x}) \\ \text{cvm} = \left[\text{pt} : \left[\begin{array}{l} \text{vt} : \text{axis-path}(r.\text{x}, \text{pt}) \\ \text{pt} : \left[\begin{array}{l} \mathbf{u} \perp \mathbf{v} \perp \mathbf{w} \\ \mathbf{u}(0) \neq \mathbf{w}(1) \end{array}\right] \\ \text{sh} : \left\{\text{rectangular, open}\right\} \end{array}\right] : Vec\right] \\ \text{c}_{\text{shape}} : \text{shape}(r.\text{x}, \text{cvm}) \end{array}\right)\right) \end{bmatrix}$$

That is, part of the descriptive meaning of 'U-shaped' is that its argument has a certain perceived shape, namely an axis that is of a particular rectangular configuration. Spatial representations like those in (19) are related to *imagistic description trees* employed in [35] in order to facilitate online recognition of iconic gestures within an artificial intelligence application. There, gesture tracking data is mapped onto (parts of) prototypical object shapes. A similar set of features from a three-axial system has also been employed in [36]. In contrast to such AI recognition approaches, the focus here is on linking multimodal utterances to linguistics and dialog theory.

The occurrence of an iconic gesture adds respective information to the DP of the current information state. Although it is clear that there is an informational interaction between CVM and DP to the effect that the merge operation ('$\wedge$') rules out combinations of predicates and gesture trajectories that are conflicting with regard to their respective *Vec* type information, this is not yet covered by the general context update **C-upc**. Due to the differences of labeling, the revised and final formulation explicitly has to take care of relating "dp" to "cvm":

C-upc:**(final version)** The display situation punctually affects the update mechanism in that DP information has to be merged with linguistic information:
If $s_t \wedge_{s_t.\text{dp.first} \wedge e.\text{cvm}} \sqsubseteq [\![\text{e}]\!].bg$ then $s_{t+1} = s_t \wedge \mathcal{F}([\![\text{e}]\!].\text{f})$

C-upc now correctly predicts that it is not well-formed to co-produce, say, a rectangular gesture and an adjective denoting roundness. The update mechanism is clocked by events: occurrences of speech (words) as well as of gestures trigger **C-upc**. Co-verbal gestures are integrated to the co-occuring linguistic expressions. This provides a co-occurrence based association between gesture and speech events that is not prone to functional ambiguities of the system in [19]. Such ambiguities arise if there are two possible attachment points in speech (say, due to signals involving deictic *here*) but only one attaching (pointing) gesture. Although **C-upc** goes beyond functional systems, it fails to do justice to the temporal and coherence relations that govern speech-gesture integration, in particular with regard to *hold* gestures [37], that are "frozen" gestural configurations that persist beyond the short life of affiliated verbal utterances. In order to provide a more sophisticated account to multimodal integration, event-based information state update mechanisms have to be extended by more detailed grammatical constraints incorporating temporal, intonational and semantic information [38], [12].

IV. Hyponymy, Good Continuation, and Demonstration

The TTR framework for iconic gestures sketched in the previous sections is in the following sections applied to the phenomena from Sec. II, which motivated the semantic integration of gestures in the first place.

The first example in (20), which is repeated from (1), introduces a spiral trajectory into the current information state.

(20) *Ich g[laube das sollen TREP]pen sein*
I think that should staircases be

'I think that should be staircases' + Fig. 1

The gesture event is represented as follows:

(21)
$$\begin{bmatrix} \text{hand} & = & \text{right} \\ \text{hs} & = & \text{G} \\ \text{carrier} & = & \left[\begin{array}{l} \text{wrst} = \text{MU} \\ \text{boh} = \text{MR>MF>ML>MB>MR} \\ \text{move} = \text{arc>arc>arc>arc} \end{array}\right] \\ \text{sync} & = & \left[\begin{array}{l} \text{sloc} = \text{CR-N} \\ \text{eloc} = \text{CUR-N} \end{array}\right] \end{bmatrix}$$

The spiral gesture has two motion aspects: a rotation carried out in the back of the hand and a translation performed by lifting the hand (i.e. wrist) – cf. the decomposition of 'spiralness' given in [31, p. 411] and of the spiral gesture in [12, p. 238]. The vector interpretation thus returns the following record of type *Vec*, where any segment '$\mathbf{u} \circ \mathbf{v}$' indicates that the vector sequence in question describes the fourth of a circle (cf. Table I):

(22)
$$\begin{bmatrix} \text{pt1} & : & \begin{bmatrix} \mathbf{u} \circ \mathbf{v} \circ \mathbf{w} \circ \mathbf{z} \circ \mathbf{y} \\ \mathbf{u}(0) \neq \mathbf{y}(1) \end{bmatrix} \\ \text{pt2} & : & \mathbf{a} \\ \text{sh} & : & \{\text{translational, circle, open}\} \end{bmatrix}$$

Note that the translational component introduced by pt2 distinguishes spirals from circles. The current context thus provides a gesticulated witness for the spiral type within the display situation:

(23)
$$s_t = \begin{bmatrix} \text{dp} = \begin{bmatrix} \text{pt1} & : & \begin{bmatrix} \mathbf{u} \circ \mathbf{v} \circ \mathbf{w} \circ \mathbf{z} \circ \mathbf{y} \\ \mathbf{u}(0) \neq \mathbf{y}(1) \end{bmatrix} \\ \text{pt2} & : & \mathbf{a} \\ \text{sh} & : & \{\text{translational, circle, open}\} \end{bmatrix} \end{bmatrix}$$

The noun "Treppen" (*staircases*) is unspecific with respect to kind and shape. It has a number of sub-kinds, however, that are distinguished according to their layout. This hyponymic relationship is captured within the inheritance type hierarchy in Fig. 5, where descendants provide the informational difference to their parent.

Given s_t from (23) and given the hyponymic relationships in Fig. 5, there are two possible updates to reach state s_{t+1}: one update uses the literally uttered hypernym *staircases*, the other the more specific *spiral staircases*, for which the context provides information due to the iconic gesture (updating with *straight staircases* is blocked, however, due to incompatible CVM constraints). Since more specific updates are generally preferred, the gesture allows us to infer the kind of staircases talked about. After applying **C-upc**, information state s_{t+1} looks as follows:

(24)
$$s_{t+1} = \begin{bmatrix} \text{x} & : & \text{Ind} \\ \text{dp} & : & \begin{bmatrix} \text{pt1} : \begin{bmatrix} \mathbf{u} \circ \mathbf{v} \circ \mathbf{w} \circ \mathbf{z} \circ \mathbf{y} \\ \mathbf{u}(0) \neq \mathbf{y}(1) \end{bmatrix} \\ \text{pt2} : \mathbf{a} \\ \text{sh} : \{\text{translational, circle, open}\} \\ \text{vt1} : \text{axis-path(x, pt1)} \\ \text{vt2} : \text{axis-path(x, pt2)} \end{bmatrix} \\ \text{cvm=dp} & : & \textit{Vec} \\ c_{\text{u}} & : & \text{spiral-staircases(x)} \\ c_{\text{shape}} & : & \text{shape(x, cvm)} \end{bmatrix}$$

In example (2), re-given in (25), we observe a use of hand-shape that cuts out a volume rather than draws a trajectory on a plane (due to the "voluminous" hand shape 'C').

(25) *die Skulptur die hat 'n [Betonsockel]*
the sculpture it has a concrete base

'the sculpture has a concrete base' + Fig.2

The gesture that is part of (25) is produced with both hands, so kinetic features are distributed over the left and the right hand. Accordingly, the *Vec* type of the gesture has double entries, partitioned according to the carriers of the respective hand. Gesture representation and interpretation is shown in (26) and (27), respectively.

(26)
$$\begin{bmatrix} \text{hands = both} \\ \text{rh} = \begin{bmatrix} \text{hand} & = & \text{right} \\ \text{hs} & = & \text{C} \\ \text{carrier} & = & \begin{bmatrix} \text{wrst = MR>MF} \\ \text{move = arc} \end{bmatrix} \\ \text{sync} & = & \begin{bmatrix} \text{sloc=lh.sync.sloc = CC-N} \\ \text{eloc = CR-M} \end{bmatrix} \end{bmatrix} \\ \text{lh} = \begin{bmatrix} \text{hand} & = & \text{left} \\ \text{hs} & = & \text{C} \\ \text{carrier} & = & \begin{bmatrix} \text{wrst = ML>MF} \\ \text{move = arc} \end{bmatrix} \\ \text{sync} & = & \begin{bmatrix} \text{sloc = CC-N} \\ \text{eloc = CL-M} \end{bmatrix} \end{bmatrix} \\ \text{rel = axisymmetric} \end{bmatrix}$$

(27)
$$\begin{bmatrix} \text{pt1lh} & = & \begin{bmatrix} \{\mathbf{u} \circ \mathbf{v}\} \\ \mathbf{u}(0) \neq \mathbf{v}(1) \end{bmatrix} \\ \text{pt1rh} & = & \begin{bmatrix} \{\mathbf{w} \circ \mathbf{x}\} \\ \mathbf{w}(0) \neq \mathbf{x}(1) \end{bmatrix} \\ \text{comb} & = & \begin{bmatrix} \text{pt} = \begin{bmatrix} \mathbf{u}(0) = \mathbf{w}(0) \\ \mathbf{v}(1) \neq \mathbf{x}(1) \\ \mathbf{a} \circ \mathbf{b} \circ \mathbf{c} \\ \mathbf{a}(0) \neq \mathbf{c}(1) \end{bmatrix} \\ \text{sh} = \{\text{semicircle, volume, open}\} \end{bmatrix} \end{bmatrix}$$

The vector representation in the 'comb'-field introduces combined path information for both hands. In the example above, the two quartercircles from the axisymmetrical movement of the hands combine to a semicircle. Modeling two-handed gestures thus adds another level of complexity which can only be mentioned here (with the exception of the closure contraint in **C-clos** – see also Table I).

The lexical entry for *concrete-base* is underspecified with respect to shape and does not have any shape-related hyponyms:

(28)
$$\begin{bmatrix} \text{bg} & = & \begin{bmatrix} \text{x} : \text{Ind} \end{bmatrix} \\ \text{f} & = & \lambda r : \text{bg}\left(\begin{bmatrix} c_{\text{cb}} : \text{concrete-base}(r.\text{x}) \end{bmatrix}\right) \end{bmatrix}$$

The descriptive meaning in (28) imposes no top-down constraints on CVM or DP. Additionally, the gesture path violates the closure constraint **C-clos**. For these reasons, it is likely

$$[\![\text{staircases}]\!] = \begin{bmatrix} \text{bg} & = & \left[\text{x} : \text{Ind}\right] \\ \text{f} & = & \lambda r : \text{bg}\left(\left[\text{c} : \text{staircases}(r.\text{x})\right]\right) \end{bmatrix}$$

$$[\![\text{spiral-staircases}]\!] = \left[\text{f} = \lambda r : \text{bg}\left(\left[\begin{array}{l} \text{cvm} = \left[\begin{array}{l} \text{pt1} : \left[\begin{array}{l} \mathbf{u} \circ \mathbf{v} \circ \mathbf{w} \circ \mathbf{z} \circ \mathbf{y} \\ \mathbf{u}(0) \neq \mathbf{y}(1) \end{array}\right] \\ \text{pt2} : \mathbf{a} \\ \text{sh} : \left\{\text{translational, circle, open}\right\} \\ \text{vt1} : \text{shape-path}(r.\text{x, pt1}) \\ \text{vt2} : \text{axis-path}(r.\text{x, pt2}) \end{array}\right] : \mathit{Vec} \\ \text{c}_{\text{shape}} : \text{shape}(r.\text{x, cvm}) \end{array}\right)\right]$$

$$[\![\text{straight-staircases}]\!] = \left[\text{f} = \lambda r : \text{bg}\left(\left[\begin{array}{l} \text{cvm} = \left[\begin{array}{l} \text{pt} : \mathbf{u} \\ \text{sh} : \left\{\text{translational, diag}\right\} \\ \text{vt} : \text{axis-path}(r.\text{x, pt}) \end{array}\right] : \mathit{Vec} \\ \text{c}_{\text{shape}} : \text{shape}(r.\text{x, cvm}) \end{array}\right)\right]$$

Fig. 5. Type hierarchy showing the hypernym "staircases" and two of its hyponyms.

that the gesture path is elliptical. There are several (in fact, infinitely many) ways to close the path so that the vector sequence has at least two vectors whose coordinates coincide. Out if these options only one is a good continuation, namely that one which brings about the shortest closure while maintaining the concatenation type (*arc*, in this case). In other words, 'good continuation', *GoCont*, is a function mapping record types *Vec* to record types *Vec*, i.e. $\mathit{GoCont} : \mathit{Vec} \rightarrow \mathit{Vec}$. Since *GoCont* changes the vector sequence representation (but not the movement annotation, of course) of a gesture, it gives rise to a re-labeling by means of π_d (cf. (14) above). The above-given features of *GoCont* can be formulated as constraints over an input display situation 'dp_{in}' and an output display situation 'dp_{out}', both being of type *Vec*.

(29) $$\mathit{GoCont} =_{\text{def}} \lambda r : \left[\begin{array}{lcl} \text{ap1} & = & \text{open} : \mathit{AP} \\ \text{cc} & = & \{\circ \mid \perp\} : \text{Vpath} \\ \text{dp}_{\text{in}} & : & \left[\begin{array}{l} \text{sh} : \text{set(AP)} \\ \text{pt} : \text{Vpath} \\ \text{vt} : \text{Vtype} \end{array}\right] \\ \text{c}_{\text{memb}} & : & \text{member(ap1, dp}_{\text{in}}\text{.sh)} \\ \text{c}_{\text{conc}} & : & \text{member(cc, dp}_{\text{in}}\text{.pt)} \\ \text{cvm} & : & \varnothing \end{array}\right] \left(T = \left[\begin{array}{lcl} \text{spt} & : & \text{Vpath} \\ \text{c}_{\text{cond}} & : & \text{member}(r.\text{cc, spt}) \\ \text{dp}_{\text{out}} & : & \left[\begin{array}{l} \text{pt} = \left[\begin{array}{l} r.\text{dp}_{\text{in}}\text{.pt } r.\text{cc spt} \\ r.\text{dp}_{\text{in}}\text{.pt}(0) = \text{spt}(1) \end{array}\right] \\ \text{vt} = r.\text{dp}_{\text{in}}\text{.vt} : \text{Vtype} \end{array}\right] \end{array}\right]\right) . \pi_d(T),$$

where "$\{\circ \mid \perp\}$" means that the concatenation type is either arc-like ($\widehat{arc}$) or straight ($\widehat{line}$).

Applying (the two-handed extension of) *GoCont* to the path from (28) gives rise to a voluminous circle, that is, a cylinder:

(30) $$\mathit{GoCont}\left(\left[\begin{array}{lcl} \text{dp}_{\text{in}} & = & \left[\begin{array}{lcl} \text{pt1lh} & = & \left[\begin{array}{l} \left\{\mathbf{u} \circ \mathbf{v}\right\} \\ \mathbf{u}(0) \neq \mathbf{v}(1) \end{array}\right] \\ \text{pt1rh} & = & \left[\begin{array}{l} \left\{\mathbf{w} \circ \mathbf{x}\right\} \\ \mathbf{w}(0) \neq \mathbf{x}(1) \end{array}\right] \\ \text{comb} & = & \left[\begin{array}{l} \text{pt} = \left[\begin{array}{l} \mathbf{u}(0) = \mathbf{w}(0) \\ \mathbf{v}(1) \neq \mathbf{x}(1) \\ \mathbf{a} \circ \mathbf{b} \circ \mathbf{c} \\ \mathbf{a}(0) \neq \mathbf{c}(1) \end{array}\right] \\ \text{sh} = \left\{\text{semicircle, volume, open}\right\} \end{array}\right] \\ \text{cvm} & = & \varnothing \end{array}\right]\right)$$

$$\rightarrow \left[\begin{array}{lcl} \text{dp}_{\text{out}} & = & \left[\begin{array}{lcl} \text{pt1lh} & = & \left[\begin{array}{l} \left\{\mathbf{u} \circ \mathbf{v} \circ \mathbf{y}\right\} \\ \mathbf{u}(0) \neq \mathbf{y}(1) \end{array}\right] \\ \text{pt1rh} & = & \left[\begin{array}{l} \left\{\mathbf{w} \circ \mathbf{x} \circ \mathbf{z}\right\} \\ \mathbf{w}(0) \neq \mathbf{z}(1) \end{array}\right] \\ \text{comb} & = & \left[\begin{array}{l} \text{pt} = \left[\begin{array}{l} \mathbf{u}(0) = \mathbf{w}(0) \\ \mathbf{y}(1) = \mathbf{z}(1) \\ \mathbf{a} \circ \mathbf{b} \circ \mathbf{c} \circ \mathbf{d} \circ \mathbf{e} \\ \mathbf{a}(0) = \mathbf{e}(1) \end{array}\right] \\ \text{sh} = \left\{\text{circle, volume, closed}\right\} \end{array}\right] \end{array}\right]$$

The good continuation is accomplished by the combined path of both hands.

Since a cylinder is a regular shape that has a lexicalized verbalization, its CVM makes the connection between the trajectory from (30) and the intension of *cylinder* explicit:

(31) $$[\![\text{cylinder}]\!] = \left[\begin{array}{lcl} \text{bg} & = & \left[\text{x} : \text{Ind}\right] \\ \text{f} & = & \lambda r : \text{bg}\left(\left[\begin{array}{l} \text{c}_{\text{cy}} : \text{cylinder}(r.\text{x}) \\ \text{cvm} = \left[\begin{array}{lcl} \text{vt} & = & \text{shape-path}(r.\text{x, cvm}) \\ \text{pt} & = & \left[\begin{array}{l} \left\{\mathbf{a} \circ \mathbf{b} \circ \mathbf{c} \circ \mathbf{d} \circ \mathbf{e}\right\} \\ \mathbf{a}(0) = \mathbf{e}(1) \end{array}\right] \\ \text{sh} & = & \left\{\text{circle, volume, closed}\right\} \end{array}\right] : \mathit{Vec} \\ \text{c}_{\text{shape}} : \text{shape}(r.\text{x, cvm}) \end{array}\right)\right] \end{array}\right]$$

Given the good continuation and the DP triggering of the lexical resource from (31), the most informative information state update s_{t+1} is the following (using only combined paths):

(32)
$$s_{t+1} = \begin{bmatrix} \text{x} & : & \text{Ind} \\ \text{dp} & = & \textit{GoCont}\left(\begin{bmatrix} \text{pt} = \{\mathbf{a} \circ \mathbf{b} \circ \mathbf{c}\} \\ \text{sh} = \{\text{semicircle, volume, open}\} \end{bmatrix}\right) \\ & & \rightarrow \begin{bmatrix} \text{vt} & = & \text{shape-path(x, cvm)} \\ \text{pt} & = & \begin{bmatrix} \{\mathbf{a} \circ \mathbf{b} \circ \mathbf{c} \circ \mathbf{d} \circ \mathbf{e}\} \\ \mathbf{a}(0) = \mathbf{e}(1) \end{bmatrix} \\ \text{sh} & = & \{\text{circle, volume, closed}\} \end{bmatrix} \\ \text{cvm=dp} & : & \textit{Vec} \\ \text{c}_{\text{cb}} & : & \text{concrete-base(x)} \\ \text{c}_{\text{cy}} & : & \text{cylinder(x)} \\ \text{c}_{\text{shape}} & : & \text{shape(x, cvm)} \end{bmatrix}$$

The final example also concerns the depiction of shape that is not realized in speech. The difference, however, being that in the datum given in (33), replicating example (3), the shape of the house talked about is explicitly delegated to the co-verbal gesture due to the demonstrative "so":

(33) *dann ist das Haus halt so []*
then is the house just this []

'then the house is like this' + Fig. 3

For this reason, the DP triggers the (initially empty) shape field of the target noun *house* directly, rather then detouring *via* a collateral expression as in (32), although the resulting information state does not reflect this subtle difference any more.

The lexical entry for house is a standard one-place predicate whose shape-field is unfilled:

(34) $[\![\text{house}]\!] =$
$$\begin{bmatrix} \text{bg} & = & [\text{x} : \text{Ind}] \\ \text{f} & = & \lambda r : \text{bg}\left(\begin{bmatrix} \text{c}_{\text{hs}} : \text{house } (r.\text{x}) \\ \text{cvm} : \textit{Vec} \\ \text{c}_{\text{shape}} : \text{shape}(r.\text{x, cvm}) \end{bmatrix}\right) \end{bmatrix}$$

The information state after processing the noun has the following public information:

(35)
$$s_{t+1} = \begin{bmatrix} \text{x} & : & \text{Ind} \\ \text{c}_{\text{hs}} & : & \text{house (x)} \\ \text{cvm} & : & \textit{Vec} \\ \text{c}_{\text{shape}} & : & \text{shape(x, cvm)} \end{bmatrix}$$

The gesture, which has been described and related to the predicate *U-shaped* in (16) and (19) above triggers a further update **C-upc**, leading to identifying the gesture's trajectory and the exemplified predicate with the shape description of the house. The resulting state s_{t+2} is shown in (36):

(36)
$$s_{t+2} = \begin{bmatrix} \text{x} & : & \text{Ind} \\ \text{c}_{\text{hs}} & : & \text{house (x)} \\ \text{cvm=dp} & : & \textit{Vec} \\ \text{c}_{\text{shape}} & : & \text{shape(x, cvm)} \\ \text{dp} & = & \begin{bmatrix} \text{pt} : \begin{bmatrix} \mathbf{u} \perp \mathbf{v} \perp \mathbf{w} \\ \mathbf{u}(0) \neq \mathbf{w}(1) \end{bmatrix} \\ \text{sh} : \{\text{rectangular, open}\} \end{bmatrix} : \textit{Vec} \\ \text{c}_{\text{u}} & : & \text{U-shaped(x)} \end{bmatrix}$$

Note, however, that the *type* of the vector (the field labeled 'vt') with regard to the shape of the house is not specified. Since the lexical entry for *house* leaves shape information underspecified and the trajectory is neutral about its type, no type information is available from these resource. The most likely type 'axis-path' has to be inferred once again, but this inference is beyond the descriptive coverage of this paper.

V. Conclusion

A formal model has been sketched for relating the perception of iconic gestures to language within an artificial intelligence-oriented information state update framework. The interface between low-level perceptual features and semantic predicates is spelled out in terms of TTR, a large-scale formal theory for language, perception and interaction [39], [40]. The model accounts for semantic key phenomena of multimodal discourse by example of speech-gesture integration like meaning specification, speech-gesture mismatches and semantic enrichments. Following the general framework of [12], a characteristics is that iconic gestures are not interpreted extensionally directly, but rather related to percepts and intensional features of natural language predicates. As the dialog proceeds, gestures are stacked onto the "gesture storage", which allows to approximate the temporal interplay of speech and gesture. Contrary to spoken language, gestures can be "frozen" and kept persistent over a period of talking. Relating both communication means via dynamic information state update mechanisms paves the way for integrating a more detailed time-based notion like that of "communication channels" [37]. Further extensions include the integration of grammatical and dialog-interactive representations of information states as well as a broadened empirical range of non-verbal behavior. In particular two-handed gestures and the combined paths they give rise to need special treatment. Further descriptive extensions might involve methodological extensions as well: the gesture perception part is a typical application area for machine learning approaches.

References

[1] A. Kendon, *Gesture: Visible Action as Utterance*. Cambridge, MA: Cambridge University Press, 2004.

[2] D. McNeill, *Hand and Mind – What Gestures Reveal about Thought*. Chicago: Chicago University Press, 1992.

[3] P. Ekman and W. V. Friesen, "The repertoire of nonverbal behavior: Categories, origins, usage, and coding," *Semiotica*, vol. 1, no. 1, pp. 49–98, 1969.

[4] M. Refice, M. Savino, M. Caccia, and M. Adduci, "Automatic classification of gestures: a context-dependent approach," in *Proceedings of the 2011 Federated Conference on Computer Science and Information System*, 2011, pp. 743–750.

[5] T. Hachaj, M. R. Ogiela, and M. Piekarczyk, "Dependence of Kinect sensors number and position on gestures recognition with gesture description language semantic classifier," in *Proceedings of the 2013 Federated Conference on Computer Science and Information Systems*, M. P. M. Ganzha, L. Maciaszek, Ed. IEEE, 2013, pp. 571–575.

[6] M. W. Alibali, "Gesture in spatial cognition: Expressing, communicating, and thinking about spatial information," *Spatial Cognition and Computation*, vol. 5, pp. 307–331, 2005.

[7] R. Cooper and A. Ranta, "Natural languages as collections of resources," in *Language in Flux: Relating Dialogue Coordination to Language Variation, Change and Evolution*, ser. Communication, Mind and Language, R. Cooper and R. Kempson, Eds. London: College Publications, 2008.

[8] R. Cooper, "Type theory and semantics in flux," in *Philosophy of Linguistics*, ser. Handbook of Philosophy of Science, R. Kempson, T. Fernando, and N. Asher, Eds. Oxford and Amsterdam: Elsevier, 2012, vol. 14, pp. 271–323.

[9] A. Lücking, K. Bergmann, F. Hahn, S. Kopp, and H. Rieser, "The Bielefeld speech and gesture alignment corpus (SaGA)," in *Multimodal Corpora: Advances in Capturing, Coding and Analyzing Multimodality*, ser. LREC 2010. Malta: 7th International Conference for Language Resources and Evaluation, 2010. doi: 10.13140/2.1.4216.1922 pp. 92–98.

[10] J. Zwarts and Y. Winter, "Vector space semantics: A model-theoretic analysis of locative prepositions," *Journal of Logic, Language, and Information*, vol. 9, no. 2, pp. 169–211, 2000.

[11] S. Larsson, "Formal semantics for perceptual classification," *Journal of Logic and Computation*, 2013. doi: 10.1093/logcom/ext059

[12] A. Lücking, *Ikonische Gesten. Grundzüge einer linguistischen Theorie*. Berlin and Boston: De Gruyter, 2013, zugl. Diss. Univ. Bielefeld (2011).

[13] S. Dobnik, R. Cooper, and S. Larsson, "Modelling language, action and perception in Type Theory with Records," in *Proceedings of the 7th International Workshop on Constraint Solving and Language Processing*, ser. CSLP'12, 2012, pp. 51–62.

[14] A. Kendon, "Gesticulation and speech: Two aspects of the process of utterance," in *The Relationship of Verbal and Nonverbal Communication*, ser. Contributions to the Sociology of Language, M. R. Key, Ed. The Hague: Mouton, 1980, vol. 25, pp. 207–227.

[15] D. Loehr, "Apects of rhythm in gesture in speech," *Gesture*, vol. 7, no. 2, pp. 179–214, 2007.

[16] H. Rieser, "Aligned iconic gesture in different strata of mm route-description," in *LonDial 2008: The 12th Workshop on the Semantics and Pragmatics of Dialogue (SEMDIAL)*, King's College London, 2008, pp. 167–174.

[17] M. Johnston, P. R. Cohen, D. McGee, S. L. Oviatt, J. A. Pittman, and I. Smith, "Unification-based multimodal integration," in *Proceedings of the Eighth Conference on European Chapter of the Association for Computational Linguistics*, European Chapter Meeting of the ACL. Madrid, Spain: Association for Computational Linguistics, 1997, pp. 281–288.

[18] M. Johnston, "Deixis and conjunction in multimodal systems," in *Proceedings of the 18th Conference on Computational Linguistics – Volume I*, International Conference On Computational Linguistics. Saarbrücken, Germany: Association for Computational Linguistics, 2000, pp. 362–368.

[19] B. Bringert, R. Cooper, P. Ljunglöf, and A. Ranta, "Multimodal dialogue system grammars," in *Proceedings of the 9th Workshop on the Semantics and Pragmatics of Dialogue*, ser. Dialor'05, 2005.

[20] A. Lücking, H. Rieser, and M. Staudacher, "Multi-modal integration for gesture and speech," in *Proceedings of the 10th Workshop on the Semantics and Pragmatics of Dialogue*, ser. Brandial'06, D. Schlangen and R. Fernández, Eds. Potsdam: Universitätsverlag Potsdam, 2006, pp. 106–113.

[21] J. Barwise and J. Perry, *Situations and Attitudes*, ser. The David Hume Series of Philosophy and Cognitive Science Reissues. Stanford: CSLI Publications, 1983.

[22] H. Kamp and U. Reyle, *From Discourse to Logic*. Dordrecht: Kluwer Academic Publishers, 1993.

[23] R. Montague, *Formal Philosophy: Selected Papers*. New Haven: Yale University Press, 1974.

[24] T. Fernando, "Observing events and situations in time," *Linguistics and Philosophy*, vol. 30, pp. 527–550, 2007. doi: 10.1007/s10988-008-9026-1

[25] S. Kopp, P. Tepper, and J. Cassell, "Towards integrated microplanning of language and iconic gesture for multimodal output," in *Proceedings of the International Conference on Multimodal Interfaces (ICMI'04)*. ACM Press, 2004, pp. 97–104.

[26] G. Johansson, "Visual perception of biological motion and a model for its analysis," *Perception & Psychophysics*, vol. 14, no. 2, pp. 201–211, 1973.

[27] L. B. Lombard, *Events: A Metaphysical Study*. London: Routledge & Kegan Paul, 1986.

[28] J. Zwarts, "Vectors as relative positions: A compositional semantics of modified PPs," *Journal of Semantics*, vol. 14, no. 1, pp. 57–86, 1997.

[29] H. Rieser, "On factoring out a gesture typology from the Bielefeld speech-gesture-alignment corpus," in *Proceedings of GW 2009: Gesture in Embodied Communication and Human-Computer Interaction*, S. Kopp and I. Wachsmuth, Eds. Berlin and Heidelberg: Springer, 2010, pp. 47–60.

[30] J. Zwarts, "Vectors across spatial domains: From place to size, orientation, shape, and parts," in *Representing Direction in Language and Space*, ser. Explorations in Language and Space. Oxford, NY: Oxford University Press, 2003, vol. 1, ch. 3, pp. 39–68.

[31] M. Weisgerber, "Decomposing path shapes: About an interplay of manner of motion and 'the path'," in *Proceedings of the Annual meeting of the Gesellschaft für Semantik*, ser. Sinn und Bedeutung 10, C. Ebert and C. Endriss, Eds. Berlin: Zentrum für allgemeine Sprachwissenschaft, 2006, pp. 405–419.

[32] A. Lücking, "The display situation," Towards a formal description of gesture and the speech-gesture interface: Panel the 6th conference of the International Society for Gesture Studies (ISGS) at the University of California, San Diego.

[33] N. Goodman, *Languages of Art. An Approach to a Theory of Symbols*, 2nd ed. Idianapolis: Hackett Publishing Company, Inc., 1976.

[34] M. Weisgerber, "Where lexical semantics meets spatial description: A framework for "Klettern" and "Steigen"," in *Proceedings of Sinn und Bedeutung 2005*, ser. SuB9, E. Maier, C. Bary, and J. Huitink, Eds., 2005, pp. 507–521.

[35] T. Sowa, *Understanding Coverbal Iconic Gestures in Shape Descriptions*. Berlin: Akademische Verlagsgesellschaft, 2006, zugl. Diss. Univ. Bielefeld.

[36] K. Barczewska and A. Drozd, "Comparison of methods for hand gesture recognition based on dynamic time warping algorithm," in *Proceedings of the 2013 Federated Conference on Computer Science and Information Systems*, M. P. M. Ganzha, L. Maciaszek, Ed. IEEE, 2013, pp. 207–210.

[37] H. Rieser, "When hands talk to mouth. gesture and speech as autonomous communicating processes," in *Proceedings of the 19th Workshop on the Semantics and Pragmatics of Dialogue*, ser. SEMDIAL 2015 goDIAL, C. Howes and S. Larsson, Eds., Gothenburg, Sweden, 2015, pp. 122–130.

[38] K. Alahverdzhieva and A. Lascarides, "Analysing language and co-verbal gesture in constraint-based grammars," in *Proceedings of the 17th International Conference on Head-Driven Phase Structure Grammar (HPSG)*, S. Müller, Ed., Paris, 2010, pp. 5–25.

[39] J. Ginzburg, *The Interactive Stance: Meaning for Conversation*. Oxford, UK: Oxford University Press, 2012.

[40] R. Cooper and J. Ginzburg, "TTR for natural language semantics," in *The Handbook of Contemporary Semantic Theory*, 2nd ed., S. Lappin and C. Fox, Eds. Oxford, UK: Wiley-Blackwell, 2015, ch. 12, pp. 375–407.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 393–402
DOI: 10.15439/2016F272
ACSIS, Vol. 8. ISSN 2300-5963

Modeling conflicts between legal rules

Tomasz Zurek
Institute of Computer Science,
Maria Curie-Sklodowska University in Lublin
Ul. Akademicka 9, 20-033 Lublin, Poland
Email: zurek@kft.umcs.lublin.pl

***Abstract*—The main aim of this work is to formalize the mechanism of resolving conflicts between statutory legal rules with a view to implementing them into a legal advisory system. The model is build on the basis of the *ASPIC*$^+$ argument modeling framework. The paper presents a discussion and a formal model of the mechanism of conflict recognition as well as models of three different mechanisms of conflict solving and a discussion of the relations between them.**

I. Introduction

TYPICAL rule-based expert systems reasoning mechanisms require knowledge bases which are created upon a few spoken and unspoken features and assumptions, such as the closed world assumption, the lack of inconsistencies and circularities in the knowledge base, etc. All of these features allow for the utilization of simple and fast modus ponens-based forward and backward chaining mechanisms. Unfortunately, there are numerous expertise areas for which many of these assumptions cannot be implemented.

One of these areas is law. There certainly have been some experiments whose authors attempted to create legal expert systems, but only a very narrow portion of statutory administrative law appears suitable for implementation in a legal expert system [1], [2]; most of these implementations have also been widely criticized [3]. The main problem of the utilization of classical expert systems in legal expertise lies in the specificity of legal knowledge and a significant presence of commonsense knowledge in legal reasoning. Additionally, legal knowledge (and commonsense knowledge connected with it) very often cannot fulfill the above-mentioned features and assumptions of a properly constructed knowledge base: legal knowledge can be inconsistent, conflicting, imprecise, and there may always appear new circumstances which may change case evaluation. Law is not a perfect and complete system. Legislators cannot foresee all possible situations which should be regulated by law; legal norms may be conflicting, they may lead to unfair conclusions, they require interpretation because it is unclear if the conditions are satisfied, etc. All these reasons hinder the simulation of legal reasoning and creation legal expert systems. A number of models of legal reasoning allowing for the representation of some aspects of legal expertise have been developed, but none of them can be regarded as complete. Practical experience reveals that legal reasoning often makes use of very "peculiar" means of inference, going far beyond the standard modus ponens principle employed in expert systems. Most of these methods are very challenging to formalize; they also are defeasible and do not guarantee correct conclusions. On the basis of the above one may notice that modeling inference processes as performed by lawyers can be perceived as a way to create an advisory legal system. The range of issues connected with modeling legal reasoning includes the problem of resolving conflicts between statutory legal rules. Legal theory and practice have worked out some methods of resolving such conflicts, yet many of them are based on commonsense knowledge, which is extremaly difficult to formalize.

The main aim of this work is to formalize a mechanism of resolving conflicts between statutory legal rules with a view to implementing it into a legal advisory system. Another aim is to incorporate the model into *ASPIC*$^+$, one of the most comprehensive argumentation modeling frameworks, thanks to which the model can be applied to a wide range of real-life legal cases.

In order to create such a model as accurately as possible, it is necessary to make some assumptions. The first one is connected with focusing on statutory legal rules. Many authors (e.g. [4]) point out important distinctions between legal principles and legal rules, where principles are statements which may be applied to various degrees (especially in the case of a conflict), but rules may be either applied or not. If there is a conflict between two or more rules, only one of them may be applied; if there is a conflict between two or more principles, the degrees of how much each may be applied are weighted. In this work I am going to focus on resolving conflicts between legal rules only. The second assumption stems from the fact that to provide a legal opinion, a lawyer usually requires not only legal knowledge (legal rules, principles which are taken from statutes), but also commonsense knowledge and knowledge about mechanisms of legal interpretation and reasoning. In order to simulate such a process as accurately as possible, it is indispensable to make a clear-cut distinction between the provision itself, its interpretation, inference mechanisms, as well as commonsense knowledge required to implement legal provisions. This distinction would allow for preserving both the universal character of the provision and its applicability to various legal problems. The third assumption is connected with the second one: the model should precisely represent the wording of the legal act and as such should contain any provision-specific imperfections, including imprecision, undefined assumptions, loopholes, ambiguity, etc. Inference and interpretation engines should be endowed with mecha-

978-83-60810-90-3/$25.00

nisms allowing for overcoming legal knowledge imperfections. Such procedures should mirror mechanisms which are used by human lawyers in such situations.

Legal theory and practice worked out a group of mechanisms which allow for resolving conflicts between rules. Most of these methods are independent of the circumstances of a given case, yet one of them is based on commonsense relations between conflicting rules and another is based on the axiological estimation of a case. It is important to notice that conflicts may appear not only between literal interpretations of statutory norms. Legal theory says that such conflicts may appear between norms reconstructed by the utilization of any inference rules (per analogiam, a'fortiori, etc.) or interpretation mechanisms.

There are four methods of conflict solving:

- *lex superior derogat legi inferiori*,
- *lex posterior derogat legi priori*,
- *lex specialis derogat legi generali*,
- an axiological method: *argument from social importance*.

The methods do not have equal power: the weaker one is used only if the stronger one does not allow for solving the conflict. The first three methods will be discussed in section III but the fourth (and the weakest) one is presented and modeled in [5], hence the model of the argument from social importance will not be presented here, though it can be easily adopted into the model presented in this paper.

II. The notion of conflict between legal rules

In legal theory there are numerous doubts about the problem of deciding who is able to state the existence of a conflict between legal rules. This is also the crucial aspect of this study.

A. The notion of conflict in the literature

A detailed discussion of the problem of the notion of conflict between legal rules is presented in [5].

Usually, a conflict between the rules of the law is treated as a clash between two distinct arguments (or their subsets) leading to two mutually exclusive conclusions [6], [7], [8]. In [9] rules are deemed conflicting when one of them "attacks" the other one, i.e. when the conclusion of one rule is complementary to the conclusion or condition of the other one. The conflict model proposed by J. Hage [10] is relatively complex. It is based on the assumption that a conflict occurs only if the conclusion from one rule implies A and the conlusion from the other one implies non-A, though it may result from the commonsense-based limitations related to the circumstances analyzed. Comparing the model of conflict from [10] (Chapter 5 concerning the rule coherence) to other models presented in [7], [8], and [11], the author points out, inter alia, that the source of conflicts may not stem only from the complementary nature of the conclusions (P and non-P, referred to as a logical conflict), but also from the incompatibility of the factual states they describe. The compatibility (or incompatibility) of the factual states may be evaluated through additional constraints (the rules may be incompatible with respect to certain constraints) and rules, which may come from commonsense knowledge.

In my opinion the key point of the discussion of the problem of conflict between statutory legal rules lies in the impossibility to detect such a conflict without considering arguments in which such rules appear. Legal rules very seldom exist separately since they are usually used as a part of the whole argument in which other kinds of rules (mainly commonsense ones) also appear. I believe that a real conflict between rules can be discovered by taking into consideration the whole argumentation process.

One of the most recent and important approaches to modeling legal argumentation is *ASPIC*$^+$ presented (among others) in [12]. *ASPIC*$^+$ is a complete framework allowing for the modeling of legal argumentation, one of whose elements is a model of the relation of an attack of one argument on another. The issue of attacks on arguments was described in [13]. The authors of the paper distinguish and define three ways by which one argument can attack another: An undermining attack is an attack on the premises of an argument, an undercutting attack is an attack on the inference step and is a way to provide "exceptions to the rule," and finally, a rebutting attack is performed by constructing a contrary or contradictory conclusion for an attacked argument's (sub)conclusion.

This definition allows for a distinction between direct and indirect attacks: an argument can be indirectly attacked by directly attacking one of its proper subarguments.

The attack relation tells us which arguments are in conflict with each other: if two arguments are in conflict, then they cannot both be accepted. The resolution of such a conflict requires the declaration of additional knowledge. The authors of [12], [13], like many others, assume a binary ordering $\prec$ on the set of all arguments that can be constructed on the basis of the argumentation theory. On the basis of such orders, they define a relation of defeat:

- A successfully rebuts B if A rebuts B on B' and $A \not\prec B'$;
- A successfully undermines B if A undermines B on φ and $A \not\prec \varphi$;
- A defeats B iff A undercuts or successfully rebuts or successfully undermines B;

where A, B, B' are arguments, φ is a one of the premises of argument B.

The issue of conflicts in a knowledge base appears not only in legal decision support systems. For example, the authors of [14] discuss the mechanism of conflict detection in business intelligence systems.

B. *ASPIC*$^+$ argumentation framework

Since the *ASPIC*$^+$ framework is a very powerful and useful tool for legal argumentation modeling, I am going to build the model of legal rules' conflict resolution on the basis of this framework. An *ASPIC*$^+$ based model of conflict detection is presented in [5]; here I am going to sketch a few important aspects of the model.

Due to the length limitation, *ASPIC*$^+$ will not be presented here in detail. I am only going to discuss some basic defini-

tions, which may lead to a better understanding of my idea; a more detailed discussion of the framework can be found in [12], [13], and others.

Generally speaking, *ASPIC*$^+$ is a framework for structured argumentation representation. It is important to emphasize that it is not a system but a framework for specifying systems, hence it does not specify any logical language to represent arguments.

ASPIC$^+$ allows for the definition of the specific logical language $\mathscr{L}$ for the representation of an argument. Arguments are constructed from a knowledge base [12], [13]:

A knowledge base in an argumentation system is a pair $(K, \leq')$ where $K \subseteq \mathscr{L}$ and $\leq'$ is a partial preorder on $K \backslash K_n$. Here $K = K_n \cup K_p \cup K_a \cup K_i$ where these subsets of K are disjoint and there are (unattackable) premises called necessary axioms (K_n), (attackable) ordinary premises (K_p), assumptions (K_a) – which are a weak type of premise always defeated by an attack – and issues (K_i) which are premises that are not acceptable unless backed by further argument.

Below are presented some basics of the framework:

An *argumentation system* is a tuple $AS = (\mathscr{L}, \mathscr{R}, n)$, where [12]:

- $\mathscr{L}$ is a logical language closed under negation;
- $\mathscr{R} = \mathscr{R}_s \cup \mathscr{R}_d$ is a set of strict ($\mathscr{R}_s$) and defeasible ($\mathscr{R}_d$) inference rules of the form $\phi_1, ..., \phi_n \rightarrow \phi$ and $\phi_1, ..., \phi_n \Rightarrow \phi$, respectively (where ϕ_i, ϕ are meta-variables ranging over wff in $\mathscr{L}$) and $\mathscr{R}_s \cap \mathscr{R}_d = \emptyset$;
- n is a naming convention ($n : \mathscr{R} \rightarrow \mathscr{L}$).

An argument in *ASPIC*$^+$ is one of the following constructs:

- ϕ if $\phi \in K_n \cup K_p$;
- $A_1, ..., A_n \rightarrow \psi$ if $A_1, ..., A_n$ are arguments, $Conc(A_1), ..., Conc(A_n) \rightarrow \psi$ is a strict rule in $\mathscr{R}_s$ and $Conc(A_i)$ is a conclusion of an argument A_i;
- $A_1, ..., A_n \Rightarrow \psi$ if $A_1, ..., A_n$ are arguments, $Conc(A_1), ..., Conc(A_n) \Rightarrow \psi$ is a defeasible rule in $\mathscr{R}_s$, and $Conc(A_i)$ is a conclusion of an argument A_i;

A structured argumentation framework (*SAF*) is a triple $\langle \mathscr{A}, \mathscr{C}, \preceq \rangle$ where [12]:

- $\mathscr{A}$ is the smallest set of all finite arguments constructed from a knowledge base in AS;
- $\preceq$ is an ordering on $\mathscr{A}$;
- $(X, Y) \in \mathscr{C}$ iff X attacks (is in conflict with) Y.

ASPIC$^+$ is also meant to generate abstract argumentation frameworks in Dung's understanding [15]. Such frameworks are simply directed graphs in which arguments (nodes) are related to other arguments by attack or defeat relations (arcs).

C. Model of conflict of legal rules

The most important assumption is a clear-cut distinction between legal rules and ordinary commonsense rules. Only the authors of [16] present an approach to formalize legal norms in *ASPIC*$^+$, yet they miss the problem of conflicts between norms; also, their formalization of legal rules is different from the one presented in this paper. Distinguishing between legal and commonsense rules is important for several reasons: firstly, a model of statutory legal norms should mirror the wording of a legal act as precisely as possible and, secondly, the defeaters of legal rules should be, and usually are, precisely regulated by law. Therefore, they cannot be treated in the same way as ordinary commonsense arguments.

On the basis of the above we assume that a set $K_p \in K$ (ordinary premises) consists of two sets: K_l and K_c, where K_l is a set of legal knowledge and K_c is a set of commonsense knowledge. K_l and K_c are separate: $K_l \cap K_c = \emptyset$.

Since a conflict may appear between legal rules, we have to allow to add to a set K_l legal rules in the shape:

$r : Conditions \twoheadrightarrow c$

where *Conditions* is a formula whose satisfaction causes truthfulness of a conclusion c. The binary connective $\twoheadrightarrow$ is used to represent a legal rule, since K_l is in K_p, the legal rule can be defeated by other argument. The defeasibility characteristic of a legal rule differs from ordinary commonsense rules. For example, the authors of [12] discuss a new connective $\rightsquigarrow$ in $\mathscr{L}$ in which $p \rightsquigarrow q$ stands for *"if p then normally/typically/usually q."* Such a rule differs from the legal one in a few important points: firstly, a legal rule is valid not only "usually" or "typically," but it is valid constantly until it is defeated. Secondly, the defeating mechanism of a legal rule is, in opposition to an ordinary commonsense rule (like $p \rightsquigarrow q$), precisely regulated by law and there is no other way to defeat it.

Since we are going to solve the problem of conflict between legal rules, we have to define what is understood as a conflict between legal rules: The authors of *ASPIC*$^+$, on the basis of their previous works, identify three kinds of attack relation on an argument [13]: undercutting, rebutting, and undermining. Since undermining concerns an attack on a premise of an argument, it cannot be treated as an attack on a legal rule.

Undercutting and rebutting concern attacking an argument or its conclusion and if both conflicting arguments use legal rules (from a set K_l), then there may be a conflict between legal rules. If an undercutting attack concerns an inference step based on a legal rule from a set K_l or a rebutting attack concerns a (sub)conclusion of a legal rule and an attacking argument uses at least one legal rule, then there is a conflict between legal rules. More formally, legal rules $r_1, r_2 \in K_l$ are potentially conflicting if:

- A undercuts argument B (on B'), r_1 is used in argument A, r_2 is used in the top rule of B' and $r_1 \neq r_2$ (r_1 and r_2 are different).
  or:
- A rebuts argument B (on B'), r_1 is used in argument A, r_2 is used in argument B' and $r_1 \neq r_2$ (r_1 and r_2 are different).

Without an in-depth analysis of the context and meaning of both arguments, it is very difficult to recognize beyond doubt whether such a conflict exists between legal rules or other subarguments. It is important to notice that conflicting legal rules do not have to be top rules of both arguments (A and B'), except undercutting argument B', because such an attack should strictly address a rule which should be defeated. It is

much easier to discover such a conflict in the special situation when conflicting legal rules r_1, r_2 are the only defeasible rules in both arguments:
Legal rules $r_1, r_2 \in K_l$ are in conflict if:

- A undercuts argument B (on B'), $Defrules(A) = A_l$, argument $r_1 \in Prem(A_l)$, $r_1 \in K_l$, there are no other defeasible rules in A, $Defrules(B') = B_l$, $r_2 \in Prem(B_l)$, $r_2 \in K_l$, $r_1 \neq r_2$ and r_2 is the top rule of B'.
- A rebuts argument B (on B'), $Defrules(A) = A_l$, $r_1 \in Prem(A_l)$, $r_1 \in K_l$, there are no other defeasible rules in A, $Defrules(B') = B_l$, $r_2 \in Prem(B_l)$, $r_2 \in K_l$, $r_1 \neq r_2$ and there are no other defeasible rules in B'.

Where: $Prem(A)$ is a function which returns the premises of an argument (by $r \in Prem(A)$ we denote that a rule r is one of the premises of argument A), $Defrules(A)$ is a function which returns defeasible inference rules used in argument A.

It is important to notice that there will be a conflict between legal rules even if the attack relation is not symmetrical (argument A attacks argument B and argument B does not attack argument A). By $conflictingRules(A,B)$ we denote that two arguments A and B contain conflicting legal rules.

III. Methods of conflict resolving between legal rules

Although in this work I am going to focus on the Polish legal system, the mechanisms which are described here may be adjusted to other statutory law systems in a relatively easy way.

At the beginning we have to look at the problem of conflict resolution from a more general point of view. If we have recognized that there is a conflict between two (or more) legal rules and we know that we cannot overcome the conflict by reconciling the conflicting norms using other legal norms, then we have to use one of the conflict resolution methods. All of these methods work in a similar way: in the case of a collision between legal rules, on the basis of some reasons listed below it is recognized which of the conflicting rules has a higher priority and all conflicting rules with lower priorities are then excluded from the reasoning process.

The theory of law distinguishes 4 main ways of resolving conflicts between legal rules [17]:

1) *Lex superior derogat legi inferiori*, based on the structural nature of law, where every legal act has its own position in the hierarchy. If one of the conflicting norms comes from the act which is at a higher position in the hierarchy, then such a norm prevails over the one from the act which is at a lower position in the hierarchy.
2) *Lex posterior derogat legi priori*, based on the time of establishing a given legal act. A legal act established later prevails over an act established earlier.
3) *Lex specialis derogat legi generali* in which a specific act (provision) derogates from (prevails over) a general regulation.
4) The final and most controversial method, known as an argument from social importance, where a rule which is more important from the axiological point of view prevails over a less important one.

Ad 1. Discussing the problem of hierarchy between legal norms requires some consideration of legal norms form several points of view:

a the first one is based on a strict hierarchy of legal acts,
b the second one is based on the relation between general law and internal law,
c the third one is based on the relation between a law which binds over the whole country and a law which binds over a part of the country,
d the fourth one is based on the relation between national and international law.

ad [a] From the point of view of the strict hierarchy of legal acts, we may state that in Polish law the constitution prevails over a legal act, which prevails over a regulation (where a regulation is a normative act issued on the basis of a specific authorization contained in a legal act aiming to allow for the execution of the act).

ad [b] From the point of view of conflict between general and internal law, the theory of law states that general law prevails over internal law.

ad [c] From the territorial point of view, a law which binds over the whole country prevails over a law which binds over a part of the country.

ad [d] Relations between national and international law are not the topic of this paper because we are going to discuss only the relations between the norms of national law.

Ad 2. The *lex posterior* mechanism of conflict resolution is based on the analysis of the dates when legal acts were established: an act established earlier has a lower priority than an act established later. If there is a conflict between these acts, the second one prevails over the first one. It is important to notice that such a mechanism works only if both conflicting acts are at the same level in the hierarchy and it is not possible to determine whether either is more specific. In other words, the *lex posterior...* mechanism has a lower priority than *lex superior...* or *lex specialis...* and it can only be employed if the utilization of *lex superior...* or *lex specialis...* cannot resolve the conflict.

Ad 3. *Lex specialis derogat legi generali* is a principle under which a specific act (provision) derogates from (prevails over) a general regulation. This mechanism is based on the analysis of the scope of conflicting legal rules and it allows for resolving conflicts which may appear, unlike in *lex posterior* or in *lex superior*, between the rules from the same legal act. This is a very strong mechanism which should be used very carefully because it may even change the conclusion made on the basis of the *lex superior* principle. There is a problem of superiority of *lex superior* over *lex specialis* of which legal literature does not include a clear view. In general, *lex superior* prevails over *lex specialis* (for example in [18] it is stated that *lex superior* is absolutely valid) but sometimes it may not work (following [17]): a legal act (as a more specific one) may be an exception to a constitutional norm or a local law (as a more

specific one) may prevail over a national one. Unfortunately, [17] does not explain clearly when or in what conditions such exceptions may occur.

Ad 4. An argument from social importance is the most controversial one and it should be used only if other ways of conflict resolution do not allow for resolving an existing conflict. This mechanism is based on the distinction between axiological contexts of conflicting norms. One of the norms may be more significant from the point of view of social importance and this norm should prevail over the less significant one. Unlike the abovementioned mechanisms, this mechanism is based on reasons which come from outside the law, making it more difficult to justify and apply. One of the main problems connected with this conflict resolution method is the uncertainty of interpretation and evaluation of social importance. One of the most clear (though rather seldom) situations may appear when one of the conflicting norms is strictly based on an expressly stated legal principle, which clearly states this norm's social importance. In other situations it is very difficult to undoubtedly decide which of the analyzed norms is more significant from the point of view of social importance. This is the reason why such a method is used very seldom and usually only in higher courts. It is also important that an argument from social importance can be strengthened by supporting it by previously decided cases. A detailed discussion and the model of an argument from social importance can be found in [5].

IV. Model of conflict resolution

Since I am going to model conflicts between legal rules, it is necessary to make some assumptions with regard to modeling such rules. In order to keep the model simple, I am going to represent them using the propositional logic, but it is also possible to use more expressive logics (the example presented in the section VI will be extended with deontic modalities) as well as additional interpretation mechanisms (for example teleological, like in [19]) and inference mechanisms (like a'fortiori [20] or instrumental [21]). The utilization of such mechanisms is important because in the literature [17] it is emphasized that such conflicts may appear between the norms reconstructed by the utilization of any inference rules (per analogiam, a'fortiori, etc.) or interpretation mechanisms.

At the beginning we have to assume some elements of a language $\mathscr{L}$ which will allow for the representation of legal rules. We assume a set of operators $OP \subset \mathscr{L}$, where $OP = \{\neg, \sim, \vee, \wedge, \twoheadrightarrow, \supset\}$ which will be used to model legal rules.

- $\neg$ is classical (strong) negation;
- $\sim$ is negation as failure;
- $\vee$ is a disjunction;
- $\wedge$ is a conjunction;
- $\twoheadrightarrow$ is a binary connective which stands for a defeasible legal rule;
- $\supset$ is a classical (material) implication used in commonsense rules.

A language $\mathscr{L}$ can also include other operators (defeasible implication, etc.) which should allow for a more adequate representation of various kinds of commonsense arguments. Let $F = \{f_1, f_2, ...\}$ be a set of propositional atoms called facts. We assume that a legal rule is a formula in the form:

$$r_n : Conditions \twoheadrightarrow Conclusion;$$

where:

- n is a rule's name
- *Conditions* is a (possibly empty) antecedent formula;
- *Conclusion* is a rule's non-empty conclusion in the form: $Conclusion = (lx \wedge ly \wedge \ldots)$, where: lx, ly are atomic conclusions which can be positive (f) or negative facts (negated by classical negation only $\neg f$).

An antecedent formula is a formula: $c1$ ***func*** $c2$ ***func*** $\ldots$ cn, where ***func*** are the operators from the set $\{\vee, \wedge\}$ and $\{c1, c2, \ldots, cn\}$ are atomic conditions, each being either a positive fact(f) or one negated by classical negation ($\neg f$), negation as failure ($\sim f$), or both ($\sim \neg f$).

Legal reasoning uses various interpretation and reasoning mechanisms; however, the legal literature points out that the basic one is linguistic interpretation (a more detailed description of linguistic interpretation can be found in [19]). As an unstrict way satisfying a rule's antecedent we understand the satisfaction of a rule's condition by utilizing any non-standard (non-linguistic) interpretation mechanism, e.g. teleological (an example of a model of teleological interpetation can be found in [19]), systematic, etc. A detailed discussion of the issue of modeling legal interpretation can be found in [22], [23], and [24].

By $K \bullet Conditions$ we will denote that a knowledge base K satisfies the conditions of a given legal rule.

On the basis of the above, a new defeasible inference rule should be added to a set $\mathscr{R}_d$:

$$r_n : Conditions \twoheadrightarrow Conclusion \wedge K \bullet Conditions \Rightarrow conclusion$$

Most authors working on modeling the resolution of conflicts between legal rules assume the existence of an order between rules or prioritising them ([9], [7], [1], [10], [25], etc.). Such an order allows to decide which of the conflicting rules is strongest and prevails over the weaker ones. Unfortunately, in real-life situations it is difficult to assume in advance that one rule is always stronger than another. It usually depends on many circumstances, which are sometimes difficult to express. To overcome this disadvantage, G. Sartor and H. Prakken ([9], [25]) propose defeasible priorities and rules which allow for inference about priorities.

A. Model of lex superior...

An interesting model of *lex superior...* is presented in [26], where the authors treat the level of authority which establishes a norm as a root of preference between norms. Unfortunately, such a conception does not fit the Polish legal system in which the hierarchy of legal acts is not strictly based on authority hierarchy, but is established by legal theory.

The *lex superior...* principle is based on a hierarchy of norms. It is obvious that formalization of a legal norm should

represent the meaning of this norm in a most accurate way, with all its features and imperfections. If we try to reconstruct the content of a legal rule from a legal text, it is important not to forget that not only the strict wording of the norm is important. There is also some relevant information connected with the norm from which a given legal rule comes from, like the position in the legal system, the date of issue, etc. Since the model of the *lex superior...* principle requires information about the position of the analyzed norms in the legal system hierarchy, such information has to be added to our model.

Let us assume that there are 2 conflicting legal rules:

$$r_1 : Conditions1 \twoheadrightarrow Conclusion1$$
$$r_2 : Conditions2 \twoheadrightarrow Conclusion2$$

It is clear that we have to add some information about the source of the rule, on the basis of which we can conclude the hierarchy between rules. We have three main levels of the legal norms hierarchy: the constitution, a legal act, and a regulation; however, apart from the general, national law there are other acts: internal legal norms, local norms, norms whose scope is narrowed to some parts of the country, etc. As $HCH \in K_n$ we denote a set of the levels of a hierarchy ($HCH = \{hch_1, hch_2, \ldots\}$). A strict partial order $OH = (HCH, >_{hch}) \in K_n$ represents a hierarchy of norms. A set $ACT = \{act_1, act_2, \ldots\} \in K_l$ represents a set of all statutory legal acts. Let an act $act \in ACT$ be a set of legal rules. By $r_l \in act_x$ we denote that a legal rule r_l is taken from an act act_x. A function $H : ACT \to HCH$ assigns to a given legal act a hierarchy level.

Every act belongs to only one hierarchy level. By $H(act_n) = hch_m$ we denote that act_n belongs to hierarchy level hch_m.

If we recognize r_n and r_m as conflicting rules, then we can solve the conflict using the *lex superior...* principle $lexSuperior \in \mathscr{R}_d$:

$$lexSuperior : (r_n \in act_k) \wedge (r_m \in act_l) \wedge (H(act_k) = hch_x) \wedge (H(act_l) = hch_y) \wedge (hch_x >_{hch} hch_y) \wedge (r_n \in Prem(A)) \wedge (r_m \in Prem(B)) \Rightarrow A \succeq B$$

where A, B are arguments built on conflicting legal rules. If both of the arguments attack each other (rebut or undercut) and it is possible to conclude (on the basis of *lexSuperior*) an order $A \succeq B$, then argument A defeats argument B.

B. Model of lex posterior...

The *lex posterior...* principle is in some points similar to *lex superior...*: both of them are also based on the properties of legal acts from which the conflicting rules are taken. While *lex superior...* is based on the position of a statute in the legal system, *lex posterior...* is based on the date of issue of a statute. If we have two conflicting rules issued on different dates, then, on the basis of the *lex posterior...* principle, a rule issued later prevails over a rule issued earlier. Similarly to *lex superior...*, we have to rely on a specific feature of acts from which conflicting rules are taken.

Let us assume that we have 2 conflicting legal rules:

$$r_1 : Conditions1 \twoheadrightarrow Conclusions1$$
$$r_2 : Conditions2 \twoheadrightarrow Conclusions2$$

By $DATE \in K_n$ we denote a set of all dates of issue of all legal acts. A function $D : ACT \to DATE$ assigns date of issue to a given act. By:

$$D(act_m) = date_m$$

we denote that a norm included in act_m was issued on date $date_m$. A strict order between the dates of issue of norms $OD = (DATE, >_{time}) \in K_l$ reflects the later-sooner relation. By $date_m >_{time} date_n$ we denote that $date_m$ was earlier than $date_n$.

If we recognize r_n and r_m as conflicting rules, then we may solve the conflict using the *lex posterior...* principle ($lexPosterior \in \mathscr{R}_d$):

$$lexPosterior : (r_n \in act_k) \wedge (r_m \in act_l) \wedge (D(act_k) = date_k) \wedge (D(act_l) = date_l) \wedge (date_k >_{time} date_l) \wedge (r_n \in Prem(A)) \wedge (r_m \in Prem(B)) \Rightarrow B \succeq A$$

where A, B are arguments built on conflicting legal rules. If both of the arguments attack each other (rebut or undercut) and it is possible to conclude (on the basis of *lexPosterior*) an order $B \succeq A$, then argument B defeats argument A.

C. Model of lex specialis...

Since *lex superior...* and *lex posterior...* are based on the knowledge which comes from legal sources, their models do not require any external sources of knowledge. Unlike them both, the *lex specialis...* principle is based on commonsense knowledge, which makes modeling this method much more challenging, because the representation and collection of commonsense knowledge is still one of the most complicated and difficult problems in the field of artificial intelligence.

Lex specialis... in the literature:

The *lex specialis...* principle has been mentioned in many papers concerning the problems of defeasible reasoning, argumentation, or normative conflicts, but in most of these papers the authors do not make attempts to formalize its nature. They usually assume an order declared in advance, which represents a relation of generality. Only in [27] a model of such a mechanism is presented:

A *normative position np* in an activity state q is more specific than np' (denoted as $np \succ_S np'$), if $np \in N_q$ and $np' \in N_q^{in}$, where: *normative position* is a deontic state of activity, N_q is a set of normative positions of an activity state q, and N_q^{in} is a set of normative positions propagated from a state of a super activity to an activity state q.

Unfortunately, it is unclear what the authors of the paper understand as a relation of activity – super activity. It may be understood as a superclass – subclass relation or an aggregation (a super activity consists of activities). Following the example presented in the paper, the relation should be understood as a kind of aggregation. It is, in my opinion, an oversimplification of the problem because not every relation

of being more or less specific can be described in such a way. I believe that such a way of modeling is appropriate for only relatively small and well-structured cases like, for example, small multi-agent systems (for which the model described in [27] was designed). Real-life legal cases are usually too fuzzy and ambiguous to let us assume without any doubt that activity q_1 is a super activity to q_2.

The model presented in [27] is an interesting approach to discuss *lex specialis...*, but due to the abovementioned controversies, I am going to present my own version of the formalization of the principle, disregarding the already assumed relations of inheritance between activities.

The model

From the model presented in [27] we adopt the idea of utilization of a partial order representing the relation being more or less specific, but its origin will be different than in [27].

Let $SPEC = (K_l, >_{spec})$ will be a partial order representing a generality relation between legal rules.

Let us assume two conflicting legal rules:

$$r_1 : Conditions1 \twoheadrightarrow Conclusion1$$

$$r_2 : Conditions2 \twoheadrightarrow Conclusion2$$

We have to decide which of them is more specific. What does it mean? As stated earlier, this mechanism is based on the analysis of the scope of conflicting legal rules. A rule which is more specific (for example, r_1) regulates a group of cases which is a subgroup of cases regulated by a more general rule (for example, r_2). Basing on the above, we may state that the scope of a rule depends on the conditional (left-hand) part of the rule, because the decision which case can be classified within the range of the scope of the rule is based on this part of a given rule.

The issue of modeling of *lex specialis...* can be divided into two separate tasks: the first one is to recognize which (if any) of the rules is more specific; the second one is to model the process of defeating a more general rule. Firstly, I am going to focus on the first task and to look at the problem of *lex specialis...* from a purely theoretical point of view. If we are going to model the principle, then we have to investigate if a set of cases which satisfies the conditions of one of the conflicting rules subsumes a set of cases which satisfies the conditions of another one. More formally, the condition of subsumption of a rule's antecendent can be modeled in such a way:

Where $Conditions1$ and $Conditions2$ are antecedents of the conflicting rules, and if for any possible to occur case P expressed by wff of $\mathscr{L}$:

$$\forall_P((P \bullet Conditions1) \rightarrow (P \bullet Conditions2))$$

and,

$$\exists_P((P \bullet Conditions2) \not\rightarrow (P \bullet Conditions1))$$

then we recognize that in a view of more restrictive character of a rule r_1 we may conclude that a rule r_1 is more specific than a rule r_2. The key challenge of such a model is an unrealistic assumption in which we have to list all cases which satisfy a rule's conditions. Firstly, it is impossible to predict all possible real-life cases (except some trivial ones); secondly, how can we recognize if a given case can possibly occur?

Since we cannot foresee all possible real-life cases, our model does not allow for recognizing all general-specific relations between rules. The only thing we can do is analyze the antecedants of conflicting rules to discover whether the condition of subsumption is fulfilled. There are some kinds of specific situations which allow us to make inferences, for example:

- Restricting Rule
  $restrictingRule$:
  $(r_1 : Conditions1 \twoheadrightarrow Conclusion1) \wedge$
  $(r_2 : (Conditions1) \wedge (Conditions1a) \twoheadrightarrow Conclusions) \wedge$
  $(Conditions1 \neq Conditions1a) \Rightarrow$
  $r_2 >_{spec} r_1$
  If every case which satisfies the conditions of a legal rule r_2 also satisfies the conditions of r_1, then rule r_2 is more specific than r_1.
- Subsuming Rule:
  $subsumingRule$:
  $(r_1 : Conditions1 \twoheadrightarrow Conclusion1) \wedge$
  $(r_2 : (Conditions1) \vee (Conditions1a) \twoheadrightarrow Conclusion2) \wedge$
  $(Conditions1 \neq Conditions1a) \Rightarrow$
  $r_1 >_{spec} r_2$
  A legal rule r_1 is more specific than a rule r_2, because every case which satisfies conditions of r_1 also satisfies conditions of r_2.

Both restricting and subsuming inference rules are a part of $\mathscr{R}_d$ ($restrictingRule \in \mathscr{R}_d$, $subsumingRule \in \mathscr{R}_d$).

Looking at the problem of *lex specialis...* in a more general way one can notice that the above mechanism does not allow for the recognition of all general-specific relations between legal rules. However, it is also worth to emphasize that in real legal practice it is also not easy to recognize them without any doubts.

Having the order $>_{spec}$ representing the specificity-generality relation between legal rules, we can model *lex specialis...*: If we recognize r_n and r_m as conflicting rules, then we may solve the above conflict on the basis of the *lex specialis...* principle ($lexSpecialis \in \mathscr{R}_d$):
$lexSpecialis : (r_n >_{spec} r_m) \wedge (r_n \in Prem(A)) \wedge (r_m \in Prem(B)) \Rightarrow A \succeq B$

If A, B are arguments built on conflicting legal rules, both arguments attack one another (rebut or undercut), and it is possible to conclude (on the basis of $lexSpecialis$) an order $A \succeq B$, then argument A defeats argument B.

The *lex specialis...* principle is slightly different from the previous ones. The most important difference lies in the commonsense background of the method: both *lex superior...* and *lex posterior...* are based on purely legal knowledge taken from statutes, while *lex specialis...* is (similarly to argument from social importance [5]) based on commonsense knowledge.

V. Ordering of inference rules

Legal practice and theory have developed a collection of methods of legal rules' conflict resolution, but it is important to notice that their results do not have to be compatible in the sense that one of the methods can give preference to one rule and second method can give preference to another one. Generally speaking, if there is a conflict between two conflict resolution methods, the theory of law assumes the following order of methods: *lex superior* prevails over *lex specialis...* which prevails over *lex posterior...* which prevails over axiological methods. Unfortunately, it is not so obvious in real-life legal cases: sometimes, in specific cases, such an order does not work and, for example, *lex specialis...* defeats *lex superior...*. Prakken and Sartor's logic ([9], [25]) allows for reasoning about priorities between arguments and rules whose elements may be helpful in such conflict resolution.

However, for the sake of this study we assume that in the case of incompatibility of conflict resolution methods (two methods infer different results), the above order will be applied.

Although our conflict resolution inference rules can be seen as a kind of higher level inference rules, in the argumentation process they are treated in the same way as ordinary inference rules. Moreover, $ASPIC^+$ does not distinguish any particular kinds of arguments except strict and defeasible ones. Hence arguments created on the basis on our inference rules can be attacked and defeated by other arguments whose strength can be regulated by the above order. How can it be applied in our framework? First of all, we have to notice that not all conflict resolution methods work in all cases. If conflicting norms are at the same level in the legal act hierarchy and were released at the same time, the other methods (for example, *lex specialis...*) can be applied. Also, if two methods can be applied and their results are not compatible, the order:
$lexSuperior \succeq lexSpecialis \succeq lexPosterior \succeq$ axiological methods
allows for defeating the conflicting arguments.

VI. Example

Let us illustrate our ideas by a simple example. There are two legal defeasible rules:

$$r_1 : vehicle \twoheadrightarrow \neg allow(enterThePark)$$

$$r_2 : vehicle \wedge emergency \twoheadrightarrow allow(enterThePark)$$

as well as 3 necessary axioms:
$r_3 : ambulance \supset vehicle$
$f1 : ambulance$
$f2 : emergency$

$r_1 \in act_k$, $r_2 \in act_l$ $act_k, act_l \in K_l$
$r_3, f1, f2 \in K_n$

On the basis of the above knowledge, we may build two argument chains A and B:

$A_1 : f1$
$A_2 : A_1, r_5 \rightarrow vehicle$
$A_3 : A_2, r_1 \Rightarrow \neg allow(enterThePark)$
$B_1 : f1$
$B_2 : f2$
$B_3 : B_1, B_2, r_2 \Rightarrow allow(enterThePark)$

In the above example, argument A_3 attacks (rebuts) B_3 and B_3 attacks (rebuts) A_3. Since both arguments have only one defeasible rule (respectively r_1 and r_2), r_1 and r_2 are legal rules and $r_1 \neq r_2$, then we may conclude that r_1 and r_2 are in conflict.
Let us assume that both r_1 and r_2 are taken from acts which are at the same level in the hierarchy and have the same date of release:

$K_n \vdash (H(act_k) \not\succ_{HCH} H(act_l)) \wedge (H(act_l) \not\succ_{HCH} H(act_k))$.
$K_n \vdash (D(act_k) \not\succ_{time} D(act_l)) \wedge (D(act_l) \not\succ_{time} D(act_k))$.

The above knowledge does not allow us to solve the conflict on the basis of *lex superior...* or *lex posterior...*, hence the possibility of using *lex specialis...* will be checked.

We do not have any additional knowledge about the case, but if we compare two conflicting rules:

$r_1 : vehicle \twoheadrightarrow \neg allow(enterThePark)$
$r_2 : vehicle \wedge emergency \twoheadrightarrow allow(enterThePark)$

we can notice that on the basis of the *restrictingRule* inference rule, argument B_4 can be constructed:

$B_4 : r_1, r_2, restrictingRule \Rightarrow r_2 >_{spec} r_1$

and, on the basis of *lexSpecialis*, argument B_6 can be constructed:

$B_5 : B_4, r_2 \in Prem(B), r_1 \in Prem(A) \Rightarrow B \succeq A$

then since we know that B attacks A (rebuttal on A_3), $B \succeq A$ and the only defeasible steps in argument chains are rules based on r_1 and r_2, argument B defeats argument A.

VII. Conclusions and future work

Legal decision support systems as well as argumentation mining systems require an adequate and comprehensive formal model of various aspects of the argumentation process. AI and law researchers agree that legal reasoning cannot be seen as a simple, mechanical, deduction-based inference, like it was treated in classical expert systems. The key point lies in the issue of argumentation: Most legal decisions are, in fact, results of a trade-off between various arguments built on the basis of legal knowledge, commonsense knowledge, legal and non-legal inference rules. This is why formal modeling of argumentation is one of the crucial elements of legal advisory systems as well as argumentation mining systems (which are

probably the future of legal informatics and a tool which can help to search, analyze, and construct new arguments).

The problem of modeling of conflicting arguments has been widely discussed in the AI and law literature. A number of authors have presented their own models of the mechanisms of conflict resolution. However, most of the existing models do not distinguish between arguments based on legal statutes and ordinary commonsense reasoning, which, in my opinion, is an oversimplification. In legal reasoning, the issue of legal rules' conflict solving is very precisely regulated and cannot be treated the same way as it is in ordinary commonsense arguments.

The main aim of this work is to formalize the mechanism of resolving conflicts between statutory legal rules with a view to implementing it into the legal advisory system. The additional aim is to incorporate the model into *ASPIC*$^+$, one of the most comprehensive argumentation modeling frameworks, thanks to which the model can be used to represent a wide range of complex real-life legal cases including various kinds of arguments. The model was created on the basis of Polish law, however, it can be easily adapted into most statutory legal systems.

In summing up the above, the most important contributions of this paper are as follows:

- a discussion of the nature of conflict between legal rules,
- a comprehensive formal model of such a conflict,
- the distinction between legal and commonsense rules,
- a formal model of three main methods of conflict solving,
- the incorporation of the model into the APSIC+ argumentation framework.

There are two important issues calling for further discussion which I am going to elaborate in my future work. The first one is the problem of modeling the strength of an argument and the balance between two conflicting commonsense arguments. Real-life arguments are very often evaluated in the light of their strength, rightfulness, adequacy, etc., which are challenging to estimate and compare. However, this is the basis on which commonsense arguments defeat one another and it is difficult to imagine a system modeling real-life argumentation without a possibility of reasoning about its strength, rightfulness, adequacy, etc. The model of the strength of an argument will be tested by the MIZAR proof checker [28]. The second (parallel) direction of future work is implementation of the above model into a small decision support system similar to the ones presented in [29] or [30].

REFERENCES

[1] R. Kowalski and F. Toni, "Abstract argumentation," *Artificial Intelligence and Law*, vol. 4, pp. 275–296, 1996. doi: 10.1007/BF00118494. [Online]. Available: http://dx.doi.org/10.1007/BF00118494

[2] R. Heller, F. Teeseling, and M. Gulpers, "A knowledge infrastructure for the Dutch Immigration Office," in *The Semantic Web: Research and Applications*, ser. Lecture Notes in Computer Science, L. Aroyo, G. Antoniou, E. Hyvonen, A. Teije, H. Stuckenschmidt, L. Cabral, and T. Tudorache, Eds. Springer Berlin Heidelberg, 2010, vol. 6089, pp. 386–390. ISBN 978-3-642-13488-3. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-13489-0_29

[3] T. Gordon, "Some problems with PROLOG as a knowledge representation language for legal expert systems," *Yearbook of Law, Computers and Technology*, pp. 52–67, 1987. doi: 10.1080/13600869.1987.9966253

[4] R. Alexy, "On balancing and subsumption. a structural comparison," *Ratio Juris*, vol. 16, no. 4, pp. 433–449, Dec. 2003. doi: 10.1046/j.0952-1917.2003.00244.x. [Online]. Available: http://dx.doi.org/10.1046/j.0952-1917.2003.00244.x

[5] T. Zurek, "Model of argument from social importance," in *Legal Knowledge and Information Systems - JURIX 2014: The Twenty-Seventh Annual Conference, Jagiellonian University, Krakow, Poland, 10-12 December 2014*, 2014. doi: 10.3233/978-1-61499-468-8-23 pp. 23–28.

[6] G. Sartor, "A simple computational model for nonmonotonic and adversarial legal reasoning," in *ICAIL '93: Proceedings of the 4th international conference on Artificial intelligence and law*, 1993. doi: 10.1145/158976.159001 pp. 192–201. [Online]. Available: http://doi.acm.org/10.1145/158976.159001

[7] F. Gordon, T, "Constructing arguments with a computational model of an argumentation scheme for legal rules: Interpreting legal rules as reasoning policies," in *Proceedings of the 11th International Conference on Artificial Intelligence and Law*. New York, NY, USA: ACM, 2007. doi: 10.1145/1276318.1276340 pp. 117–121. [Online]. Available: http://doi.acm.org/10.1145/1276318.1276340

[8] D. Nute, "Defeasible logic," in *INAP'01: Proceedings of the Applications of prolog 14th international conference on Web knowledge management and decision support*. Berlin, Heidelberg: Springer-Verlag, 2003. doi: 10.1007/3-540-36524-913. ISBN 3-540-00680-X pp. 151–169. [Online]. Available: http://dx.doi.org/10.1007/3-540-36524-913

[9] H. Prakken and G. Sartor, *A system for defeasible argumentation, with defeasible priorities*, ser. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 1996, vol. 1085, pp. 510–524. [Online]. Available: http://dx.doi.org/10.1007/3-540-48317-915

[10] J. Hage, *Studies in Legal Logic*. Springer, 2005.

[11] N. Van Der Torre L, W and N. Yao-Huata, *Defeasible Deontic Logic*. Kluwer, 1997, ch. The Many Faces Of Defeasibility In Defeasible Deontic Logic, pp. 79–122. [Online]. Available: http://dx.doi.org/10.1007/978-94-015-8851-5_5

[12] S. Modgil and H. Prakken, "The ASPIC+ framework for structured argumentation: a tutorial," *Argument and Computation*, vol. 5, no. 1, pp. 31–62, 2014. doi: 10.1080/19462166.2013.869766

[13] B. van Gijzel and H. Prakken, "Relating Carneades with abstract argumentation via the ASPIC + framework for structured argumentation," *Argument & Computation*, vol. 3, no. 1, pp. 21–47, Mar. 2012. doi: 10.1080/19462166.2012.661766

[14] M. Hernes and K. Matouk, "Knowledge conflicts in business intelligence systems," in *Proceedings of the 2013 Federated Conference on Computer Science and Information Systems*, M. P. M. Ganzha, L. Maciaszek, Ed. IEEE, 2013, pp. pages 1241–1246.

[15] P. M. Dung, "On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games," *Artificial Intelligence*, vol. 77, no. 2, pp. 321 – 357, 1995. doi: http://dx.doi.org/10.1016/0004-3702(94)00041-X. [Online]. Available: http://www.sciencedirect.com/science/article/pii/000437029400041X

[16] H. Prakken and G. Sartor, "Formalising arguments about norms," in *Legal Knowledge and Information Systems - JURIX 2013: The Twenty-Sixth Annual Conference, December 11-13, 2013, University of Bologna, Italy*, 2013. doi: 10.3233/978-1-61499-359-9-121 pp. 121–130.

[17] L. Leszczynski, *Zagadnienia teorii stosowania prawa: Issues of theory of application of law*. Krakow: Zakamycze, 2001.

[18] J. Stelmach, *Kodeks argumentacyjny dla prawnikow*. Zakamycze, 2003.

[19] T. Zurek and M. Araszkiewicz, "Modeling teleological interpretation," in *Proceedings of ICAIL 2013*. ACM, 2013. doi: 10.1145/2514601.2514619 pp. 160–168. [Online]. Available: http://doi.acm.org/10.1145/2514601.2514619

[20] T. Zurek, "Modelling of a'fortiori reasoning," *Expert Systems with Applications*, vol. 39, no. 12, pp. 10 772–10 779, 2012. doi: http://dx.doi.org/10.1016/j.eswa.2012.02.188. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0957417412004617

[21] T. Zurek, "Instrumental inference in legal expert system," in *Proceedings of JURIX 2011*. IOS Press, 2011. doi: 10.3233/978-1-60750-981-3-155 pp. 155–159.

[22] M. Araszkiewicz, "Scientia juris : a missing link in the modelling of statutory reasoning," in *Legal knowledge and information systems : JURIX 2014 : the twenty-seventh annual conference*, ser. Frontiers in

Artificial Intelligence and Applications, R. Hoekstra, Ed. Amsterdam: IOS Press, 2014, pp. 1–10. ISBN 978-1-61499-467-1

[23] M. Araszkiewicz and T. Zurek, "Comprehensive framework embracing the complexity of statutory interpretation," in *Legal Knowledge and Information Systems - JURIX 2015: The Twenty-Eighth Annual Conference, Braga, Portual, December 10-11, 2015*, 2015. doi: 10.3233/978-1-61499-609-5-145 pp. 145–148.

[24] D. Walton, G. Sartor, and F. Macagno, "An argumentation framework for contested cases of statutory interpretation," *Artificial Intelligence and Law*, vol. 24, no. 1, pp. 51–91, 2016. doi: 10.1007/s10506-016-9179-0. [Online]. Available: http://dx.doi.org/10.1007/s10506-016-9179-0

[25] H. Prakken and G. Sartor, "Argument-based extended logic programming with defeasible priorities," *Journal of Applied Non-classical Logics*, vol. 7, no. 1, pp. 25–75, 1997. doi: 0.1080/11663081.1997.10510900

[26] W. W. Vasconcelos, M. J. Kollingbaum, and T. J. Norman, "Normative conflict resolution in multi-agent systems," *Autonomous Agents and Multi-Agent Systems*, vol. 19, no. 2, pp. 124–152, 2008. doi: 10.1007/s10458-008-9070-9

[27] A. García-Camino, P. Noriega, and J. A. Rodríguez-Aguilar, "An algorithm for conflict resolution in regulated compound activities," in *Engineering Societies in the Agents World VII, 7th International Workshop, ESAW 2006. Revised Selected and Invited Papers*, G. M. P. O'Hare, A. Ricci, M. J. O'Grady, and O. Dikenelli, Eds., vol. 4457, Springer-Verlag. Dublin, Ireland: Springer-Verlag, September 6-8 2007. doi: 10.1007/978-3-540-75524-1 11. ISBN ISBN 978-3-540-75522-7 pp. 193–208. [Online]. Available: http://dx.doi.org/10.1007/978-3-540-75524-111

[28] A. Grabowski, A. Kornilowicz, and C. Schwarzweller, "Equality in computer proof-assistants," in *Proceedings of the 2015 Federated Conference on Computer Science and Information Systems*, ser. Annals of Computer Science and Information Systems, M. Ganzha, L. Maciaszek, and M. Paprzycki, Eds., vol. 5. IEEE, 2015. doi: 10.15439/2015F229 pp. 45–54. [Online]. Available: http://dx.doi.org/10.15439/2015F229

[29] T. Zurek and E. Kruk, "Supporting of legal reasoning for cases which are not strictly regulated by law," in *ICAIL '09: Proceedings of the 12th International Conference on Artificial Intelligence and Law*, 2009. doi: 10.1145/1568234.1568263 pp. 220–221. [Online]. Available: http://doi.acm.org/10.1145/1568234.1568263

[30] T. Zurek, "Conflicts in legal knowledge base," *Foundations of Computing and Decision Sciences*, vol. 37, no. 2, pp. 129–145, 2012. doi: 10.2478/v10209-011-0006-9

6th International Workshop on Advances in Semantic Information Retrieval

RECENT advances in semantic technologies form a solid basis for a variety of methods and instruments that support multimedia information retrieval, knowledge representation, discovery and analysis. They influence the way and form of representing documents in the memory of computers, approaches to analyze documents, techniques to mine and retrieve knowledge. The abundance of video, voice and speech data also raises new challenging problems to multimedia information retrieval systems.

We believe that our workshop will facilitate discussions of new research results in this area, and will serve as a meeting place for researchers from all over the world. Our aim is to create an atmosphere of friendship and cooperation for everyone, interested in computational linguistics and semantic information retrieval. The ASIR'15 workshop will continue to maintain high standards of quality and organization, set in the previous years. We welcome all the researchers, interested in semantic information retrieval, to join our event.

TOPICS

The workshop addresses semantic information retrieval theory and important matters, related to practical Web tools. The topics and areas include but not limited to:

- Domain-specific semantic applications.
- Evaluation methodologies for semantic search and retrieval.
- Models for document representation.
- Natural language semantic processing.
- Ontology for semantic information retrieval.
- Ontology alignment, mapping and merging.
- Query interfaces.
- Searching and ranking.
- Semantic multimedia retrieval.
- Visualization of retrieved results.

EVENT CHAIRS

- **Klyuev, Vitaly,** University of Aizu, Japan
- **Mozgovoy, Maxim,** University of Aizu, Japan

PROGRAM COMMITTEE

- **Carrara, Massimiliano,** Universita di Padova, Italy
- **Dobrynin, Vladimir,** Saint Petersburg State University, Russia
- **Goczyła, Krzysztof,** Gdansk University of Technology, Poland
- **Haralambous, Yannis,** Institut Telecom - Telecom Bretagne, France
- **Homenda, Wladyslaw,** Warsaw University of Technology, Poland
- **Jin, Qun,** Waseda University, Japan
- **Lai, Cristian,** CRS4, Italy
- **Leonelli, Sabina,** University of Exeter, United Kingdom
- **Nalepa, Grzegorz J.,** AGH University of Science and Technology, Poland
- **Pyshkin, Evgeny,** University of Aizu, Japan
- **Shtykh, Roman,** CyberAgent Inc., Japan
- **Ślęzak, Dominik,** University of Warsaw & Infobright Inc., Poland
- **Soldatova, Larisa,** Brunel University, United Kingdom
- **Suárez-Figueroa, Mari Carmen,** Ontology Engineering Group, Scool of Computer Science at Universidad Politécnica de Madrid, Spain
- **Tadeusiewicz, Ryszard,** AGH University of Science and Technology, Poland
- **Vacura, Miroslav,** University of Economics, Czech Republic
- **Zadrożny, Sławomir,** Systems Research Institute of Polish Academy of Sciences, Poland
- **Ławrynowicz, Agnieszka,** Poznan University of Technology, Poland

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 405–409
DOI: 10.15439/2016F172
ACSIS, Vol. 8. ISSN 2300-5963

Game with a Purpose for Mappings Verification

Tomasz Boiński
Department of Computer Architecture
Faculty of Electronics, Telecommunication and Informatics
Gdańsk University of Technology
11/12 Narutowicza Street
80-233 Gdańsk, Poland
Email: tobo@eti.pg.gda.pl

***Abstract*—Mappings verification is a laborious task. The paper presents a Game with a Purpose based system for verification of automatically generated mappings. General description of idea standing behind the games with the purpose is given. Description of TGame system, a 2D platform mobile game with verification process included in the gameplay, is provided. Additional mechanisms for anti-cheating, increasing player's motivation and gathering feedback are also presented. Example of the system usage for verification of mappings between WordNet synsets and Wikipedia articles is presented. The evaluation of proposed solution and future work is also described.**

I. Introduction

NOWADAYS people tend to spend a lot of time playing computer games. Increased availability of powerful mobile devices further increases time spend on this form of entertainment. In 2012 Samsung Electronics Polska performed a study among people in Poland on the time spend on video games [1]. Almost half of the population aged 27-35 spends 1 to 2 hours weekly playing games and 14% spends over 20 hours a week. High percentage of the players use mobile devices like smartphones (20%) and tablets (5%). It can be seen that in many cases playing games occupies the amount of time equal to at least a part-time job. On the other hand many nowadays problems still cannot be solved by a computer algorithm and assigning human resources to perform such tasks is often economically inefficient.

The question arises what if we could use all that potential resources (time and knowledge of the users and hardware capabilities of their devices) to solve also non-algorithmic problems? One can imagine that if we would treat a group of users as a distributed system, then it is sufficient to divide a problem into small portions, distribute them to the players and finally aggregate achieved results. This however introduces some difficulties, from technical ones like how to divide a problem into sub-problems, how to distribute them and how to gather results, to more social oriented like how to trust the results and more importantly how to convince people to spend their time on solving our problem. Some research shows however that even educational games can be well perceived if constructed properly [2] so that an algorithm verification system within games seems plausible.

Numerically solvable problems adopted volunteer computing model [3] where the users donate the power of their machines when it is not needed (the calculations are done between the periods of active hardware usage). Using this model it is not possible to solve all type of problems, as some of them cannot be successfully turned into a computer algorithm [4]. We can use heuristics but than we still have to verify the results manually. In crowdsourcing [5] approach the user is encouraged to perform a task for some type of gratitude. The task can be both algorithmic and non-algorithmic.

It is also worth mentioning that the problems that are difficult for computers are usually quite easy for humans. Example of such problem is image recognition, image tagging, natural language understanding and processing or verification of results obtained by traditional heuristics. The results often very complicated algorithms are quite easy to grasp by an average human. Linking those two areas could prove to be useful for performing laborious tasks without a need of hiring additional workers especially that many research results needs some evaluation and sometimes it is the sole purpose of the research [6].

In this paper we focus on a so called human-based computation (HBC) [7] model. It is using human brain directly to solve a computational problem. The term was defined by Kosorukoff in 2001 in a paper about human enhanced genetic algorithm [8].

HBC can be viewed as similar to crowdsourcing. The later focuses solely on solving the problem by human, while in the former model part of the problem is solved by a computer. Usually the machine performs sub-problem organization, distribution and retrieval of results, sometimes some calculations are done using heuristics. The human part usually contains the verification of computer generated results or performing the calculations itself [9].

In our research we applied HBC-model for verification of mappings. As an example we used mappings between WordNet and Wikipedia [10], [11], [12], [13] that were obtained during Colabmap[1] project and are a result of running heuristics on a computer. Currently we are working on generalization of the proposed solution hoping to provide a general framework suitable for solving different types of problems.

II. Games with a Purpose

In 2006 Luis von Ahn proposed usage of computer games as something more than pure entertainment and thus creating

[1] http://kask.eti.pg.gda.pl/colabmap

978-83-60810-90-3/$25.00

the idea of so called GWAP (Game With A Purpose) [4]. GWAPs are typical games that provide standard entertainment value that users expect but are designed in a way that allows generation of added value by solving a problem requiring intellectual activity. It is worth noticing that GWAPs does not allow financial gratification for the work. The will to continue playing should be treated as the only way of gratifying users [14].

Ahn defined three types of GWAPs:

- output-agreement game,
- inversion-problem game,
- input-agreement game.

In the first type of GWAPs two randomly selected players are presented with identical input data and each produces results only based on the available information. Both players should achieve identical results without any knowledge about the other player - they are awarded only when both will give identical answers. In this case an identical answer provided by both players is treated as highly probable to be correct as it comes from independent sources. Example of such game is The ESP Game [15], where users task was to tag images with keywords. The players were presented with an image and were given 2,5 minute to enter all keywords that are related to the image. The game proved to be very popular. During first few months authors managed to gather around 10 million tags with statistics showing many users playing for over 40 hours a week [4]. In 2006 Google released their own version of the game called Google Image Labeler[2] (it was shut down in 2011) which was used to extend capabilities of Google Graphics.

The second type, the inversion-problem game, also selects players randomly. The players are however divided into two groups - describers and guessers. The describer is presented with input data and is responsible for creating tips allowing the guessing player to correctly point out the input data. The players are awarded points when the output given by the guesser is equal to the input. One of the examples of such game is Phetch [16]. One of the players is presented with a random image from the Internet. His or her task is to describe the image to other players. Other players task is to find an identical image. Other example is the Peekaboom game [17]. The task of the players is to quest words that are describing an image. The "boom" player is presented with an image and its description in a few words. The "peek" player is presented with empty page on which the "boom" player gradually reveals parts of the image. The "peek" player have to guess, based on the revealed fragments, the exact words describing the image.

The third type, input agreement game, also selects players randomly. Both players have to achieve agreement on the input data. More precisely they have to guess whether the other player received the same or different input data. Each player describes what he or she sees on the screen and the other player have to state whether the input is similar to theirs or different. The example of such game is TagATune [18] where players should describe their feelings about a tune that is played. Based on the description the players have to decide whether they heard the same or different tune.

What is common for all above types of games is that the players unknowingly generate added value that is not possible to calculate using computers. The problem behind such games is a way to lure players - only very large user base can provide viable results. During implementation many techniques can be used to enrich the game and encourage more players, like time limits, awards in form of points and achievements, difficulty levels, leader boards or randomness of input data [14].

The quality of target game can be described by two parameters: average lifetime play (the time that average player spent playing the game) and throughput (average number of problems solved per hour of playtime) [4]. Simko [19] also pointed out that GWAP should be characterized also by the total number of players that took part in the game.

III. Wikipedia - WordNet Connections Validation

During the Colabmap project we created a set of mappings between English Wikipedia articles and WordNet synsets. Sample mappings are presented in Table I. Each mapping consists of a WordNet synset, definition of the synset and the title of Wikipedia article with special characters coded using RFC 3986. Such mappings, when proved to be correct, will allow formalization of Wikipedia structure. The obtained set of mappings contained algorithmically created 54475 connections that required verification. Tempted by the results obtained during the Samsung's survey we decided on implementing a GWAP for validation of those connections.

The originally obtained mappings were extended with three additional "next best" mappings with the idea of presenting the user a question (definition of a synset) with 4 possible answers (Wikipedia article titles). At the beginning the 3 other answers were randomly selected from the set of Wikipedia's pages but such approach quickly proved to be incorrect as the "next best" mappings were not related at all to the question. Instead we used Wikipedia search functionality to select alternative answers (according to Wikipedia) following the Algorithm 1.

Algorithm 1 Algorithm for selecting alternative answers

```
for all synonyms of WordNet synset do
    Read the synonym
    Perform a Wikipedia search using the synonym
    Store 3 top elements from search results
end for
```

The example of extended mappings are presented in Table II. For the tests we randomly selected 100 synsets from our database to limit the time needed to gather the results and verify the viability of the game.

IV. TGame

We decided to implement TGame[3] ("Tagger Game") as a 2D platform game following the output-agreement model.

[2]http://en.wikipedia.org/wiki/Google_Image_Labeler

[3]https://play.google.com/store/apps/details?id=pl.gda.eti.kask.tgame, http://kask.eti.pg.gda.pl/tgame/

TABLE I
SAMPLE WORDNET – WIKIPEDIA MAPPINGS

Synset (WordNet)	Definition (WordNet)	Article (Wikipedia)
Sept. 11, September 11, 9-11, 9/11, Sep 11	the day in 2001 when Arab suicide bombers hijacked United States airliners and used them as bombs	September_11
interval, time interval	a definite length of time marked off by two instants	Time
ice age, glacial epoch, glacial period	any period of time during which glaciers covered a large part of the earth's surface	Ice_age Glacial_period
man hour, person hour	a time unit used in industry for measuring work	Man-hour
entity	that which is perceived or known or inferred to have its own distinct existence (living or nonliving)	Entity
French leave	an abrupt and unannounced departure (without saying farewell)	French_leave
hunt, hunting	the pursuit and killing or capture of wild animals regarded as a sport	Huntingdon
blindman's bluff, blindman's buff	a children's game in which a blindfolded player tries to catch and identify other players	Blind_man%27s_bluff
landler	a moderately slow Austrian country dance in triple time	L%C3%A4ndler
coup d'oeil, glimpse, glance	a quick look	Coup_d%27%C5%93il

TABLE II
SAMPLE OF EXTENDED MAPPINGS, THE ORIGINAL MAPPING IS EMPHASIZED

Synset (WordNet)	Articles (Wikipedia)
Sept. 11, September 11, 9-11, 9/11, Sep 11	*September 11*, 9/11 Commission, 9/11 conspiracy theories, United Airlines Flight_93
interval, time interval	*Time*, Interval (music), Interval, Interval (mathematics)
ice age, glacial epoch, glacial period	*Ice age*, Pleistocene, Wisconsin glaciation, Gravettian
man hour, person hour	*Man-hour*, Hourman, Man of the Hour, 24 Hours of Le Mans
entity	*Entity*, Administrative divisions of Mexico, Administrative division
French leave	*French leave*, French leave (disambiguation), Desertion, French Leave (1930 film)
hunt, hunting	*Huntingdon*, Hunting, Fox hunting, Seal hunting

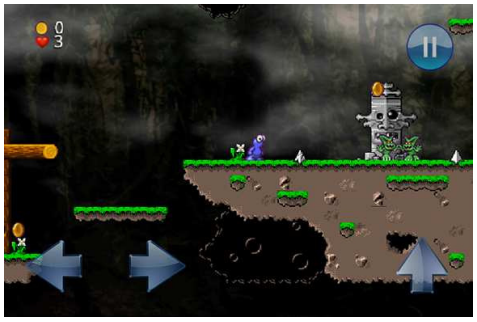


Fig. 1. TGame.

Fig. 2. Checkpoint activation.

We chose Android platform as a test environment due to its popularity and ease of access for users and developers. The game implements standard features like different levels and collectibles (coins, hearts), need of finishing one level before the other one is accessible. The player is encouraged to replay levels by a simple point system that awards the player for killing monster, gathering stars and hearts (Figure 1).

A. Tying questions with the game

One of the biggest challenge is to properly include the mappings (or any type of a general question) into the game. We tried to implement the questions to be as non intrusive as possible but still easy to stumble upon. In TGame the verification of mappings is done when the player wants to activate a checkpoint (a respawn place when player is moved when killed). To activate the checkpoint player needs to answer the question provided by marking the correct mapping (Figure 2). When the answer is identical to the one stored in the database the checkpoint is activated. If the player chose other answer then the checkpoint is not enabled. When the player is certain that he or she selected a correct answer then he or she can report his or her answer using the proper option in the pause menu. The checkpoint is then activated for one use.

From the technical point of view the communication between the client and the server goes as follows. Each client upon first connection downloads pack of configurable number of questions and possible answers so connection to the Internet

is required only at first start of application and later at user chosen intervals. Whenever possible the game sends gathered results with statistical information and downloads new pack of questions (if needed).

B. Answer Verification

The process of reporting wrong answers requires explicit action from the user. It was designed to require some activity but not too much so not to discourage the users. Very easy access to submission would encourage people to skip reading the question and just submitting information about wrong answer. In general the game has to be paused and proper menu have to be selected. Only the last question can be reported.

Furthermore when submitting results also time elapsed between displaying the question and selecting the answer is also submitted. Such extensions allows us to eliminate submissions that i.e. are so short that the user would not be able to read the question. Randomly selected batch of questions required on average 5 seconds to be properly read and understood by players. We decided to discard all answers that took less than 4 seconds (8% of all results).

The answers that the players gave (correct, incorrect and reported) are later compared with the one calculated by Colabmap algorithms. All the selected answers are visible in administrators panel of the server side of the solution and can be exported using csv format to an external tool.

C. Results Analysis

During first two months of tests the game was downloaded by 25 people, mainly students and friends (currently according to Google Play web page there are between 50 and 100 downloads without any additional advertising). The original 25 players gave 626 answers for 100 questions. The game run 10 hours in total on multiple devices. Each player solved 44,42 questions per hour. At this rate, with average playtime of each player at 5 hours, we would need minimum 2500 players to answer each question at least once. Judging by other similar games available on Google Play such number can be achieved with proper advertising of the product given the user base and popularity of mobile games.

During the evaluation of the proposed solution we faced two main type of problems. In some cases the additional answers generated using the Wikipedia search functionality provided very similar pages which introduced difficulty in choosing the correct one. Selection of 100 random questions also introduced problem with high variety of difficulty level among questions. It became obvious that some of them require expert-level knowledge. Examples of questions belonging to those two groups are presented in Table III. The column "Answer 1" contains the correct mapping. Furthermore the type of game implemented (a simple platform game) did not match the questions asked. In future work we will reorganize the questions to Yes/No/Unsure form to lower the difficulty level and implement other type of games to better match the type of mappings that are verified.

TABLE III
SAMPLE QUESTIONS WITH HIGH LEVEL OF DIFFICULTY

The Question	Answer 1	Answer 2	Answer 3	Answer 4
Asiatic nut trees: wing nuts	Pterocarya	Pterocarya fraxinifo-lia	Pterocarya stenoptera	Cyclocarya
a colorless flammable volatile liquid hydrocarbon used as a solvent	Xylene	O-Xylene	P-Xylene	Xylene cyanol
a former large county in northern England	Yorkshire	Yorkshire 6	Yorkshire captaincy affair of 1927	South Yorkshire Fire and Rescue
fine porcelain that contains bone ash	Bone china	Aynsley China	Bisque (pottery)	Porcelain

V. MAPPING UPDATE

We tried three approaches for deciding whether the mapping, based on the answers provided by the players, should be updated or not:

- The mapping was considered correct when 75% of the player answers agreed. This approach however did not give any results as only 50% of original mappings managed to get enough answers, none of the incorrect mappings were marked as correct.
- The mapping was considered correct when at least 50% of player answers agreed. In this case 64% of all mappings were marked as correct which covered 75% of all mappings marked as correct in our database. Unfortunately this method generated some false positives.
- The mapping which gathered the most of the player answers was considered correct. In this case 74% of all mappings were marked as correct witch covered 80% of all mappings originally marked as correct in our database. This method also generates false positives.

Currently in our solution we implemented the third method as it provided the best results. Still this method does not allow us to automatically state whether the given mapping is correct or not. However "problematic" mappings are clearly pointed out by the players (by either majority of wrong answers or reports). Such mappings can than be verified manually by experts. In our further work we plan on extending the procedure with additional parameters like user reputation, level and field of education, history of answers etc. which should improve the level of trust that can be put in user the generated answers.

During the evaluation period the players submitted 17 mappings update requests regarding 12 questions. Sample reports are presented in Table IV. The corrected mappings are emphasized.

TABLE IV
UPDATED MAPPINGS

WordNet Definition	Original mappings	Other available answers
an advanced student or graduate in medicine gaining supervised practical experience ('houseman' is a British term)	Internet Movie Database	Houseman, Julius Houseman, *Internship (medicine)*
large voracious aquatic reptile having a long snout with massive jaws and sharp teeth and a body covered with bony plates	Crocodile tears	*Crocodile*, Crocodylus, Schnappi
(elections) more than half of the votes	Supermajority	*Majority*, Simple majority, Absolute majority

VI. CONCLUSIONS AND FURTHER WORKS

We proposed a platform for verification of the results of heuristic algorithms. Currently verification of mappings is supported. We verified the solution using Wikipedia – WordNet mappings and managed to get some promising results and were able to correct some of the mappings. The problems that still need to be solved include better formulation of the question and the trust that the system can put in answers provided by the users.

We also plan on extending the proposed solution by generalizing it for other type of tasks, inclusion of different clients, not only game based, designed for certain types of questions or with required user knowledge in mind. We are also currently implementing social features like achievements and leader boards that should lure more players and create a wider user base. In case of Wikipedia - WordNet mappings we plan on tagging questions with difficulty levels and include them in a quiz-like game similar to "Fifteen to One"[4] or "1 of 10"[5]). Such type of client could be more suitable for such defined problems. The TGame can be a great application for crowd base image tagging or a client when the questions will be redesigned to a Yes/No format.

Our research shows that popularity of computer games and wide availability of devices that can be used for playing at any time makes GWAPs an approach that has some unexplored potential. Our first implementation, despite its drawback, shows that this potential is relatively easy to unlock. Even for small user base we managed to find some errors in our mappings. Implementation of different client applications more fitting the types of tasks needed to be done (image tagging, sound analysis etc.) and careful question formulation should enable us to fully unlock the possibility of crowdsourcing based task execution. When succeeded such possibility can be of great help to researchers around the world as it reduces resources and time needed to verify the results of designed algorithms and implementations. Furthermore it can be implemented as an alternative to in app purchases or advertisements. This way the users can be provided with content with their work be treated as another means to "pay" for it.

[4]http://en.wikipedia.org/wiki/Fifteen_to_One

[5]http://pl.wikipedia.org/wiki/Jeden_z_dziesi%C4%99ciu

REFERENCES

[1] *Prawie połowa Polaków gra codziennie w gry wideo (in Polish)*, URL http://samsungmedia.pl/pr/223805/prawie-polowa-polakow-gra-codziennie-w-gry-wideo. [Online], Biuro Prasowe Samsung Electronics Polska Sp. z o.o.

[2] U. Świerczyńska-Kaczor and J. Wachowicz, "Student response to educational games – an empirical study," in *Proceedings of the 2013 Federated Conference on Computer Science and Information Systems*, M. P. M. Ganzha, L. Maciaszek, Ed. IEEE, 2013, pp. pages 1293–1299.

[3] D. P. Anderson and G. Fedak, "The computational and storage potential of volunteer computing," in *Cluster Computing and the Grid, 2006. CCGRID 06. Sixth IEEE International Symposium on*, vol. 1. IEEE, 2006, pp. 73–80.

[4] L. Von Ahn, "Games with a purpose," *Computer*, vol. 39, no. 6, pp. 92–94, 2006.

[5] J. Howe, "Crowdsourcing: A definition," *URL http://www.crowdsourcing.com/cs/2006/06/crowdsourcing_a.html.[Online]*, p. 29, 2006.

[6] V. Osinska, A. Jozwik, and G. Osinski, "Mapping evaluation for semantic browsing," in *Proceedings of the 2015 Federated Conference on Computer Science and Information Systems*, ser. Annals of Computer Science and Information Systems, M. Ganzha, L. Maciaszek, and M. Paprzycki, Eds., vol. 5. IEEE, 2015, pp. 329–335. [Online]. Available: http://dx.doi.org/10.15439/2015F50

[7] D. Wightman, "Crowdsourcing human-based computation," in *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries*. ACM, 2010, pp. 551–560.

[8] A. Kosorukoff, "Human based genetic algorithm," in *Systems, Man, and Cybernetics, 2001 IEEE International Conference on*, vol. 5. IEEE, 2001, pp. 3464–3469.

[9] J. Simko and M. Bieliková, "Games with a purpose: User generated valid metadata for personal archives," in *Semantic Media Adaptation and Personalization (SMAP), 2011 Sixth International Workshop on*. IEEE, 2011, pp. 45–50.

[10] R. Korytkowski and J. Szymanski, "Collaborative approach to WordNet and Wikipedia integration," in *The Second International Conference on Advanced Collaborative Networks, Systems and Applications, COLLA*, 2012, pp. 23–28.

[11] J. Szymański, "Mining relations between Wikipedia categories," in *Networked Digital Technologies*. Springer, 2010, pp. 248–255.

[12] ——, "Words context analysis for improvement of information retrieval," in *Computational Collective Intelligence. Technologies and Applications*. Springer, 2012, pp. 318–325.

[13] J. Szymański and W. Duch, "Self organizing maps for visualization of categories," in *Neural Information Processing*. Springer, 2012, pp. 160–167.

[14] L. Von Ahn and L. Dabbish, "Designing games with a purpose," *Communications of the ACM*, vol. 51, no. 8, pp. 58–67, 2008.

[15] ——, "Labeling images with a computer game," in *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 2004, pp. 319–326.

[16] L. Von Ahn, S. Ginosar, M. Kedia, and M. Blum, "Improving image search with phetch," in *Acoustics, speech and signal processing, 2007. icassp 2007. ieee international conference on*, vol. 4. IEEE, 2007, pp. IV–1209.

[17] L. Von Ahn, R. Liu, and M. Blum, "Peekaboom: a game for locating objects in images," in *Proceedings of the SIGCHI conference on Human Factors in computing systems*. ACM, 2006, pp. 55–64.

[18] E. L. Law, L. Von Ahn, R. B. Dannenberg, and M. Crawford, "TagATune: A game for music and sound annotation," in *ISMIR*, vol. 3, 2007, p. 2.

[19] J. Simko, "Semantics discovery via human computation games," *Semantic Web: Ontology and Knowledge Base Enabled Tools, Services, and Applications*, p. 286, 2013.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 411–420
DOI: 10.15439/2016F428
ACSIS, Vol. 8. ISSN 2300-5963

An Ontology-based Contextual Pre-filtering Technique for Recommender Systems

Aleksandra Karpus*, Iacopo Vagliano†, Krzysztof Goczyła*, Maurizio Morisio†
* Faculty of Electronics Telecommunication and Informatics,
Gdańsk University of Technology
ul. Gabriela Narutowicza 11/12, 80-233 Gdańsk-Wrzeszcz, Poland
Email: {aleksandra.karpus, krzysztof.goczyla}@eti.pg.gda.pl
† Dept. Control and Computer Engineering,
Politecnico di Torino,
C.so Duca degli Abruzzi, 24, 10129 Turin, Italy
Email: {iacopo.vagliano, maurizio.morisio}@polito.it

***Abstract*—Context-aware Recommender Systems aim to provide users with the most adequate recommendations for their current situation. However, an exact context obtained from a user could be too specific and may not have enough data for accurate rating prediction. This is known as the data sparsity problem. Moreover, often user preference representation depends on the domain or the specific recommendation approach used. Therefore, a big effort is required to change the method used. In this paper we present a new approach for contextual pre-filtering (i.e. using the current context to select a relevant subset of data). Our approach can be used with existing recommendation algorithms. It is based on two ontologies: Recommender System Context ontology, which represents the context, and Contextual Ontological User Profile ontology, which represents user preferences. We evaluated our approach through an offline study which showed that when used with well-known recommendation algorithms it can significantly improve the accuracy of prediction.**

I. Introduction

RECOMMENDER Systems are software tools and techniques providing suggestions for items to be of use to a user. Various kind of items can be suggested, such as music tracks, movies and news. Context-Aware Recommender Systems (CARS) are a particular category of recommender systems which exploits contextual information to provide more effective recommendations. For example, in a temporal context, vacation recommendations in winter should be very different from those provided in summer. Or a restaurant recommendation for a Saturday evening with your friends should be different from that suggested for a workday lunch with co-workers [1].

We distinguish three forms of context-aware recommendation process: *contextual pre-filtering*, *contextual post-filtering* and *contextual modeling* [2]. *Pre-filtering* approaches use the current context to select a relevant subset of data on which recommendation algorithm is applied. *Post-filtering* approaches exploit contextual information to select only relevant recommendations returned by some algorithm. *Contextual modeling* differs from others techniques as it incorporates the context into recommendation algorithm.

Nowadays context information such as time and location are easy to be obtained with modern devices. However, also other parameters may be considered, such as company (*alone*, *with friends*, *with girlfriend*) which may be relevant when recommending movies or vacations. In addition, the exact context sometimes can be too narrow, as Adomavicius and Tuzhilin [2] exemplified by considering the context of watching a movie with a girlfriend in a movie theater on Saturday. Using this exact context may be problematic for several reasons. First, certain aspects of the overly specific context may not be significant. For example, user's movie watching preferences with a girlfriend in a theater on Saturday may be exactly the same as on Sunday, but different from Wednesday's. Therefore, it may be more appropriate to use a more general context specification, i.e. weekend instead of Saturday. Second, exact context may not have enough data for accurate rating prediction, which is known as the data sparsity problem. Thus it may be useful to refer to a more general context such as watching a movie with a girlfriend in a movie theater on weekend, watching a movie with someone in a movie theater on weekend and so on.

Additionally, often user preferences and items representation depends on the application domain addressed or on the specific recommendation approach used. Thus, a big effort is required to adapt the recommender system to another domain or to change the approach to use.

In this paper, we address the problems previously mentioned and we focus on the following research questions:

- *Is it possible to represent context by combining different dimensions (such as time, location, mood, etc.) and representing different granularities for each dimension (e.g. the precise time moment, the day of the week or the season)?*
- *Is it possible to represent user preferences and items in such a way that can be adapted to different application domains and combined with different recommendation approaches?*

The second author was supported by a fellowship from TIM.

We present a new approach to represent context and user preferences, which is based on two ontologies: Recommender System Context (RSCtx)[1] which represents the context, and another ontology, Contextual Ontological User Profile (COUP), which represents user preferences. RSCtx is an ontology in a classical sense, while COUP is an ontology build according to Structured-Interpretation Model (SIM) [3] and it consists of multiple ontological modules. Moreover, we propose a new ontology-based contextual pre-filtering method which could be used with existing recommendation algorithms.

We evaluated our approach by means of an offline study with a rating prediction task which showed that the usage of proposed ontologies and pre-filtering technique with recommendation algorithms could significantly improve the accuracy of prediction according to the Mean Absolute Error (MAE) measure.

The rest of the paper is organized as follows: Section II presents related work, Section III introduces our ontology to represent the context, while Section IV addresses the overall recommendation approach and the representation of user preferences. We detail the evaluation process and its results in Section V. Conclusions and future work close the paper.

II. Related Work

We distinguish related work in works which addressed representation of context and other ontology based recommender systems proposed. The first is presented in Section II-A while the latter is briefly described in Section II-B.

A. Context Representation

In this section, firstly we address ontology-based context modeling and then we review context representation for recommender systems.

Many context ontologies have been proposed in the context awareness community. There are a number of surveys which review the literature relevant to context modeling in general [4], [5] or focusing on ontology-based models [6], [7]. In addition, Costabello [8] presented and compared a number of ontology-based context models against a set of requirements. This requirements fit also for our purpose therefore in the following we present the requirements and summarize Costabello's comparison, obviously also considering RSCtx.

The relevant context aware and ontology engineering requirements are:

R1. Domain independence. A number of context ontologies have been created to model a given domain-specific scenario. Others adopt a domain-independent approach.

R2. Coverage. The ontology must guarantee a proper level of completeness for what concerns the desired contextual dimensions. In particular, the model must support multiple context dimensions such as device features, user preferences, location and time.

R3. Formality. Some ontology-based context models rely on formal definitions, while others adopt a more intuitive approach.

R4. Variable Context Granularity. Certain ontologies model context dimensions at different level of granularity. For example, location might be expressed in terms of latitude and longitude, or with a label assigned to a place (e.g. office, beach, cinema, etc.).

R5. User Friendliness Evaluation. Context-aware application developers must spend a reasonable amount of effort dealing with the context model, thus the ontology must be sufficiently easy to adopt and well documented. The presence of a user evaluation campaign to assess such feature is assessed by certain context models.

R6. Core ontology approach. The vocabulary must adopt a modular design, thus focusing on modeling core classes and properties that will be extended by third-party domain specialists.

Linked Data[2] is a set of best practice to publish structured data on the Web. The set of data, vocabularies and ontologies which follows these practices made up the Web of Data. Costabello [8] considered also a number of requirements related to the Linked Data principles, which also fit for our purpose:

R7. Open World Assumption. The Web of Data is an open environment, and describing context in this scenario must consider third-party extensions unknown beforehand. Extensibility must be obtained with a low effort, thus add-ons must not impact on the already existing model.

R8. Lightweight Ontology. According to Linked Data best practices [9], the goal is to keep ontologies small and simple.

R9. Reuse of Existing Terms. Linked Data best practices favor the reuse and the combination of classes and properties of existing vocabularies. This is done to prevent the proliferation of terms and reduce the range of choices when modeling data.

R10. Availability on the Web. Web of Data vocabularies are published on the Web, and accessible according to Web of Data best practices. Moreover, they are associated to an HTML page, the "namespace document", whose task is to provide a textual description of the vocabulary rationale, along with classes and properties explanation and examples.

Following these requirements, Costabello [8] compared a number of ontologies which modeled context and proposed PRISSMA[5], a vocabulary designed to model client generated context data. We present in the following the main feature of this vocabulary and other related works showed in Table I. PRISSMA satisfies the most of the aforementioned requirements, although *variable context granularity* only partially. It miss formality and user friendliness evaluation, but none of the other works satisfies these two. On the contrary, all

[1] http://softeng.polito.it/rsctx/

[2] http://linkeddata.org

[5] http://ns.inria.fr/prissma

TABLE I
A COMPARISON OF ONTOLOGY-BASED CONTEXT MODELS [8]. FULL SUPPORT IS IDENTIFIED BY ●, PARTIAL SUPPORT BY ○, NO SUPPORT BY THE EMPTY CELL.

Work	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10
PRISSMA[3] [8]	●	●		○		●	●	●	●	●
DCO[4]	●	○				●	○			●
SOUPA [10]	●	●				●	●		○	
CoOL [11]	●	○		○		●	○			
CONON [12]	●	●		●		●	●			
CoDaMos [13]	●	●					●			○
Korpipää et al. [14]	●	○					○			
Hervás and Bravo [15]	●	●				●				
RSCtx	●	●		●		●	●	●	●	●

the works provide coverage and are domain independent, and all but one support (at least partially) the open world assumption. The only other ontology published on the Web is the Delivery Context Ontology (DCO)[6], a modular and fine-grained vocabulary to model mobile platforms. It does not provide linking with other vocabulary, thus it is not considered a lightweight ontology. The SOUPA ontology [10] is an OWL ontology which is extensible, i.e. support the open world assumption, and reuses external ontologies, but it does not comply with Linked Data principles, for example it is not publicly available on the Web. CoOL [11] is a modular OWL ontology, which is grounded on F-Logic and uses features typically avoided in lightweight ontology. CONON [12] is another modular OWL ontologies, which is not published on the Web and does not reuse existing vocabularies. CoDaMos [13] is an extensible OWL ontology that is available on the Web but no namespace vocabulary is present. It is not lightweight and does not reuse other vocabularies. Korpipää et al. [14] present a context model designed for mobile, context-aware applications. It is general, but does not reuse existing terms and it is not extensible. Hervás and Bravo [15] propose a modular context model composed by independent ontologies which support extensions, although they do not reuse already existing linked data ontologies.

Various works addressed context representation for recommender systems. Abowd et al. [16] distinguish among primary and secondary context: the first can be directly measured, while the second not and needs to be derived from other types of contextual information. Kaminskas and Ricci [17] reviewed literature about contextual music retrieval. They distinguish among environmental, user-related and multimedia context: the first refers to information about the location of the user, the current time, weather, temperature, etc.; the second to information about the activity of the user, the user's demographic information, emotional state; and the third to other types of information the user is exposed to besides music, e.g., text and images. In addition to traditional dimensions (time location etc.) the authors suggested traffic, noise and light level. As multimedia context, they mention text and images. They indicated some cases in which it can be useful consider this kind of context, e.g. for adapting music to text context as done by Cai et al. [18]. Baltraunas et al. [19] proposed an approach to assess which contextual factors are important and to which degree they influence user ratings. They conducted a study in which users where asked to judge whether a contextual factor influences the rating given a certain contextual condition. In their survey they focus on tourism domain and consider budget, time availability, transport in addition to traditional dimensions. RSCtx supports most of the addressed dimensions and distinguish among user-related and environmental context. It does not address multimedia context, but it considers the device features.

[6]http://www.w3.org/TR/2009/WD-dcontology-20090616/

B. Ontology based Recommender System

It has been proved that ontological user profile improves recommendation accuracy and diversity [20]. More specifically, a number of ontology-based and context-aware recommender system have been proposed. AMAYA allows management of contextual preferences and contextual recommendations [21]. AMAYA also uses an ontology-based content categorization scheme to map user preferences to entities to recommend. News@hand [22] is an hybrid personalized and context aware recommender system, which retrieves news via RSS feed and annotates by using system domain ontologies. User context is represented by a weighted set of classes from the domain ontology. Rodriguez et al. [23] proposed a CARS which recommens Web services. They use a multi-dimensional ontology model to describe Web services, a user context and an application domain. The multi-dimensional ontology model consists of a three independent ontologies: a user context ontology, a Web service ontology and an application domain ontology, which are combined into one ontology by some properties between classes from different ontologies. The recommendation process consists in assigning a weight to the items based on a list of interests in the user ontology. All this works focus on a specific domain and an ad-hoc algorithm, while our approach for representing user preferences is cross-domain and can be applied to different recommendation algorithm.

Hawalah and Fasli [24] suggest that each context dimension should be described by its own taxonomy. Time, date, location and device are considered as default context parameters in the movie domain. It is possible to add other domain specific context variables as long as they have a clear hierarchical

representations. Besides context taxonomies, this approaches uses a reference ontology to build contextual personalized ontological profiles. The key feature of this profile is the possibility of assigning user interests in groups, if these interests are directly associated with each other by a direct relation, sharing the same super-class, or sharing the same property.

Other works uses ontologies and taxonomy to improve the quality of recommendations. Middleton et al. [25] uses an ontological user profile to recommend research papers. Both research papers and user profiles are represented through a taxonomy of topics and the recommendation are generated considering topics of interest for the user and papers classified in those topics. Mobasher et al. [26] proposed a measure which combines semantic knowledge about items and user-item rating, while Anand et al. [27] inferred user preferences from rating data using an item ontology. Their approach recommends the items using the ontology and inferred preferences while computing similarities. A more detailed description of ontology based techniques is available in [28] and [29].

III. The Recommender System Context Ontology

Recommender System Context (RSCtx) extends PRISSMA[7], a vocabulary based on Dey's definition of context [30]. PRISSMA relies on the W3C Model-Based User Interface Incubator Group proposal[8], which describes mobile context as an encompassing term, defined as the sum of three different dimensions: user model and preferences, device features, and the environment in which the action is performed. A graph-based representation of PRISSMA is provided Figure 1.

We designed RSCtx following METHONTOLOGY [31], a well know ontology design method. We assumed there is a predefined set of contextual dimensions in a given application, each with a defined set of attributes and we modeled the contextual information relevant to provide recommendations. We did not focus on any particular domain, on the contrary we aimed at reusing the ontology in different applications. As in PRISSMA, the point of view used to describe the context itself is the application point of view, thus we considered, the user itself as part of the context.

We needed a more detailed representation of the environment, in order to consider other contextual dimensions such as the purpose of the user and the weather. Figure 2 shows how `prissma:Environment` has been extended, by adding a number of properties and related concepts. To represent the weather we integrate `hw:WeatherState` from the Weather Ontology[9]. In this ontology the temperature is represented with respect to the room temperature, thus we defined a new class to represent symbolic values of temperature (such as warm, cold, etc.) and an attribute to represent numeric values, as show in Figure 3.

[7]http://ns.inria.fr/prissma

[8]http://www.w3.org/2005/Incubator/model-based-ui/XGR-mbui/

[9]https://www.auto.tuwien.ac.at/downloads/thinkhome/ontology/WeatherOntology.owl

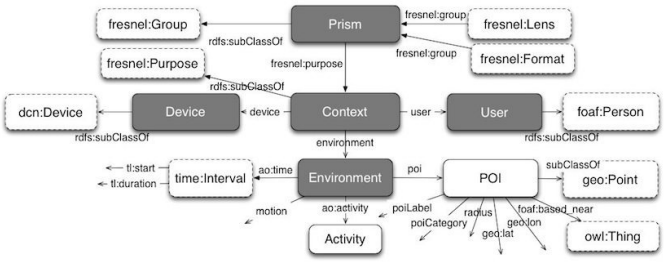


Fig. 1. The PRISSMA vocabulary [8].

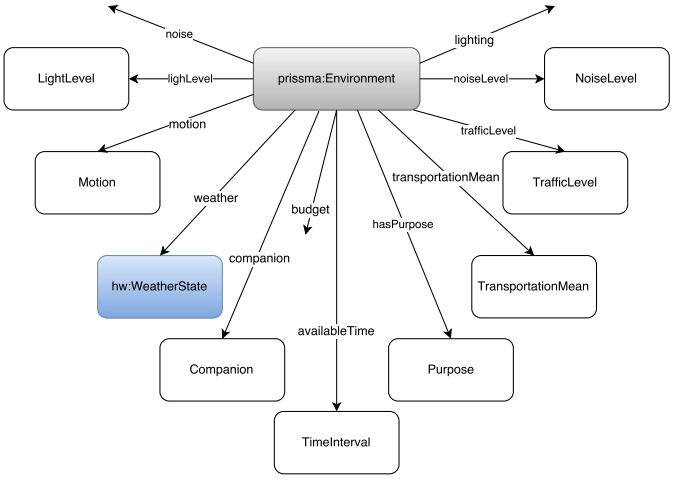


Fig. 2. Relations and concepts which extend `prissma:Environment`

We also extended the time and location representations. We needed a more expressive model of these two dimensions, since asking for recommendations which have the same time stamp and the coordinates of the actual context is too restrictive and the recommender system may not have enough data. On the contrary, by generalizing the context (for example distinguishing among weekend and working day, or considering the city or neighborhood instead of the actual user position) may enable the recommender system to provide recommendations. The concept `prissma:POI` has been extended with various properties to represent the location in the context of a specific site by integrating the Buildings and Rooms vocabulary[10]. Furthermore, other properties related to the hierarchical organization of the location (such as the neighborhood, the city and the province of the current user position) have been added and some concepts from the Juso ontology[11] have been reused. Figure 4 depicts relations and attributes which characterize a location. Yellow rectangles indicate concepts from rooms vocabulary, while blue ones are taken from Juso. The representation of time augments `time:Instant` defined in the Time ontology[12]. Some time intervals have been defined: the hours and the parts of day

[10]http://vocab.deri.ie/rooms

[11]rdfs.co/juso/latest/html

[12]https://www.w3.org/2006/time

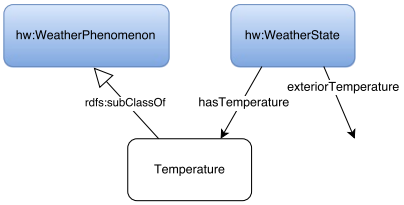


Fig. 3. Temperature representation in our ontology.

(morning, afternoon, etc.). In addition, days of week are classified in weekdays or weekend and seasons are represented. Figure 5 illustrates how time is represented and the relations with PRISSMA and the Time ontology.

Furthermore, we extended the user representation adding some dimensions which may be of interest, as the emotional, mental and physiological state of the user or his fitness. This can be interesting mainly in the medical or fitness domain, but emotional state can affect the user also in taking other kind of decisions, like choosing a movie to watch or music to listen to. Emotional, mental and physiological state concepts are equivalent to emotional, mental and physiological state in the General User Model Ontology (GUMO) [32], an ontology to describe the user which is available on the Web, although it is not compliant with Linked Data principles since it has not a namespace assigned. In addition, the emotional state is an extension of `emoca:Emotion`, which is defined in the Emotion Ontology for Context Awareness (EmOCA)[13]. We added some attributes to the physiological state and also defined an arousal relation which reuse `emoca:Arousal`. Figure 6 depicts the user representation in our ontology.

The emotion in EmOCA are represented according to Russel's model [33]. We extended `emoca:Emotion`, in particular we added pleasure and dominance as subclasses of `emoca:Component` in order to represent emotions according to the Pleasure Arousal Dominance (PAD) model [34] as well, as it is showed in Figure 7. In this way, we can indicate that the emotion is defined by valence and arousal by means of `emoca:isDefinedBy` to refer to Russel's model, while we can indicate that the emotion is defined by pleasure, arousal and dominance to refer to the PAD model. Furthermore, in psychology it is possible to refer to emotion just by indicating its category (such as joy, anger, disgust, etc.). In EmOCA, six categories have been already defined, which can be used also in RSCtx since the emotional state is a subclass of `emoca:Emotion`. We can add more categories in our ontology if it will be needed, although at the moment it has not been done.

IV. Recommendation Approach

A. Contextual User Profile Ontology

To model user profiles we used the Structured-Interpretation Model (SIM) [35], [36], which consists of two types of ontological modules, i.e. *context types* and *context instances*. Context types describe the terminological part of an ontology (TBox) and are arranged in hierarchy of inheritance. Context instances describe assertional part of an ontology (ABox) and are connected with corresponding context types through a relation of instantiation. There is another kind of relation, i.e. aggregation, which links context instances of more specific context types to a context instance of a more general context type. In the class hierarchy in a classical ontology there always exists a top concept, i.e. `Thing`. In SIM ontology there is a top context type and a top context instance connected by instantiation. It is possible to add multiple context instances to one context type and aggregate multiple context instances into one context instance. The idea of SIM is shown in Figure 8.

The idea of adaptation SIM ontology as a user profile was proposed by Karpus and Goczyła [37]. They modeled contextual user profiles using only three context variables, i.e. location, time and mood, which influences a split of terminology into ontological modules. Our approach is different in some crucial aspects. First of all, we allow storage of many user profiles in one SIM ontology. We also support a storage of preferences from multiple domains by adding context types related to different domains. Another difference is the number of context variables permitted. We add context types and context instances related to contextual parameters in a dynamic way. As a consequence, we can use as many variables as needed in our approach. An example of contextual profile for one user is shown in Figure 9.

Only three modules in the example illustrated in Figure 9 are fixed: `topContextType`, `topContextInstance` and `UserType`. All others are configurable or can be added in a dynamic way. In `topContextType` we defined the concept `Rating` and its corresponding roles, e.g. `isRatedWith` and `hasValue`. `UserType` is artificial and is present in the SIM ontology because it enable to add many user profiles to the ontology. In the next level of the hierarchy, there are context types that describe domains of interests related to a recommender system which will use the profile. In the next levels, all context types and instances are added to the contextual user profile during the learning phase or later, when a new context situation occurs.

The general process of learning the user profile is as follows. At the beginning there is just the RSCtx ontology and an empty contextual ontology, i.e. with terminological part only. For a given user, an item is taken with the rating and the situation in which it was consumed from the user's history. The level of granularity of the context is checked with the RSCtx ontology and is changed if needed, e.g. shifting from *time = 2 p.m.* to *time = afternoon*. A context instance is created for this context, if it is not already available. Finally, an item with its rating is added to the identified context instance. Each item is represented as a set of individuals of appropriate concepts defined in a domain context type.

B. Recommendation

We use the ontologies previously presented for pre-filtering in the recommendation process. The aim is to provide a

[13]http://ns.inria.fr/emoca/

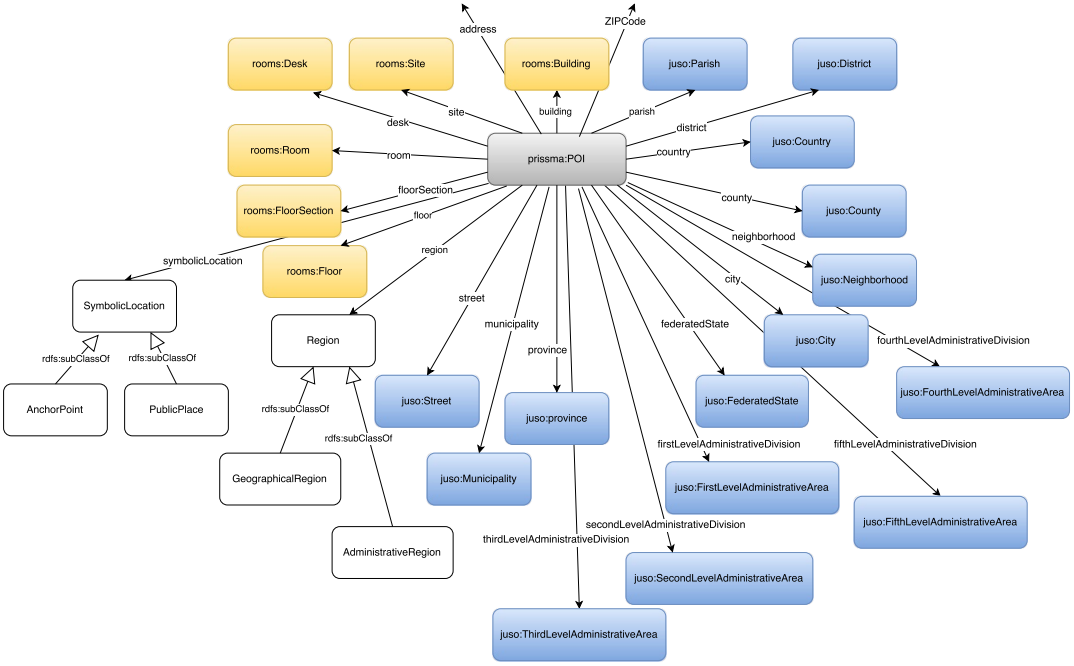


Fig. 4. Concepts and relations of RSCtx representing the location dimension.

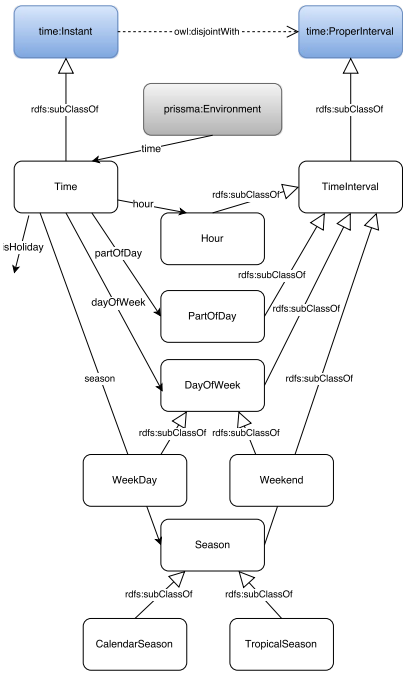


Fig. 5. Time representation in RSCtx ontology as extension of Time and PRISSMA ontologies

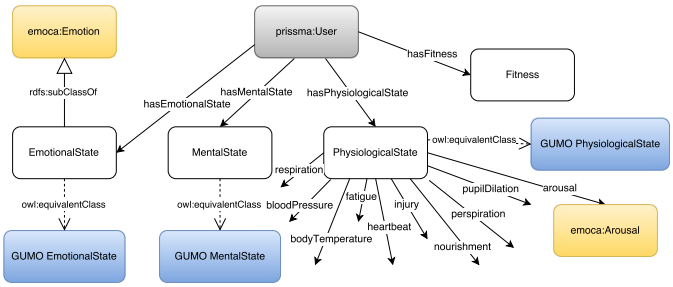


Fig. 6. User representation in RSCtx ontology.

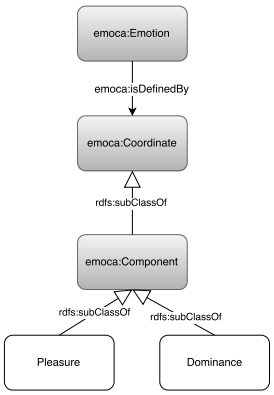


Fig. 7. Emotion representation in RSCtx ontology.

universal context-aware improvement for existing algorithms.

The system consists of three main functional parts: context detection and generalization, user profile and pre-filtering, and recommendation. In the first part, we used the RSCtx ontology to identify the user context from raw data and generalize it in the desired granularity level . The second part is responsible for building user profile, finding a context instance that fits the actual user context, and returning only relevant user

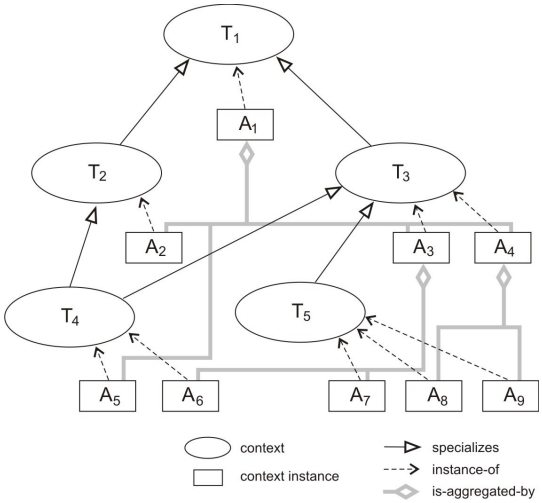


Fig. 8. Structured-Interpretation Model [3]

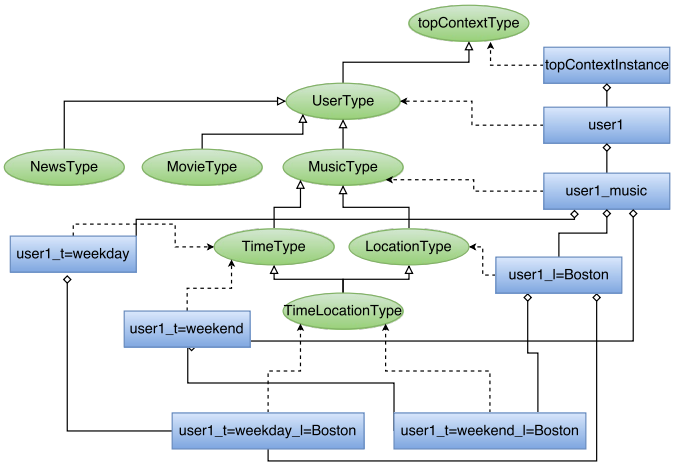


Fig. 9. An example of COUP

preferences. The last part uses well-know algorithms, e.g. Item kNN, User Average, SVD++, for providing recommendations. For this task we exploit implementations from the LibRec[14] library.

The general recommendation process is as follows. Given a user and his current situation, a proper generalization of his context is generated by using the RSCtx ontology. Then, an appropriate context instance from COUP is identified by using the generalized context. If a context instance is not found in the user profile, the generalization step is repeated to search for a module that corresponds to the new context. If it is found, user preferences are prepared to be used with a recommendation algorithm.

V. Evaluation

We conducted an offline study in order to evaluate the RSCtx ontology and the COUP ontology. We selected a number of algorithms and we compared the accuracy of each algorithm when used as is and when combined with proposed ontologies. We aimed to answer the following question: *does our context and user preferences representation improve accuracy of recommendation algorithms?*

We relied on ConcertTweets dataset [38], which combines implicit and explicit user ratings with rich content as well as spatio-temporal contextual dimensions and social network data. It contains ratings that refer to musical shows and concerts of various artists and bands. Since the dataset was generated automatically, there were some duplicated events, for example the same concert was occurring twice, once with country *United Kingdom* and once *UK*. We fixed this kind of situations. Another problem with the dataset is the usage of two rating scales: one numerical scale with ratings in the range [0.0, 5.0] and one descriptive scale with possible values equal to *yes*, *maybe* and *no*, although *no* never occurred. We decided to split the dataset into two separate sets according to the scale type and we mapped the descriptive values *yes*, *maybe* and *no* with the numerical values 2, 1 and 0. Table II presents some statistics about the data by considering the whole dataset and each of the sets generated when splitting by scale type. We prepared two pairs (one for each scale) of training and test sets for hold-out validation. In the test set we put 20% of the newest ratings of each user. All other ratings were placed in the training set. The split was performed based on rating timestamp value.

Because of the dataset domain, we needed to add to contextual user profile a new *context type*, `MusicType`. For this purpose we reused two existing ontologies related to music, i.e. `musicbrainz`[15] and `music vocabulary`[16]. We used their terminological parts only. Any item can be represented as a pair of individuals of type `mb:Artist` and `m:Musical_Event` and with their corresponding properties.

Our approach is a pre-filtering approach and can be used with existing recommendation methods. We evaluated the ontologies with five algorithms: Random Guess, Item kNN, User Average, SVD++ and Time SVD++. We compared the results of first four algorithms without pre-filtering and with pre-filtering, while the fifth was executed without pre-filtering only, because it already contains the *time* as a contextual factor [39]. We used it as a baseline for comparing our contextual pre-filtering technique combined with SVD++ algorithm.

We performed an experiment for rating prediction task and measured accuracy with MAE. Results are presented in Table III and Figures 10 and 11. It should be noticed that without pre-filtering, the User Average algorithm outperforms SVD++. This may be due to the way in which users rate musical events: it may be possible that they do not use the whole rating scale but just a part of it, e.g. a user evaluates only those events that he likes (his ratings are always greater that 3.0). As it can be seen in Figure 10 and Table III, when our ontological pre-filtering approach is applied, results on the *numerical scale*

[14]http://www.librec.net/

[15]https://musicbrainz.org/

[16]http://www.kanzaki.com/ns/music

TABLE II
STATISTICS ON THE DATA CONTAINED IN CONCERTTWEETS DATASET AT THE TIME OF THE EXPERIMENT

	All	**Descriptive ratings**	**Numeric Ratings**
Number of users	61803	56519	16479
Number of musical events	116320	110207	21366
Number of pairs artist and musical events	137382	129989	23383
Number of ratings	250000	219967	30033
Maximum number of ratings per user	1423	1419	92
Minimum number of ratings per user	1	1	1
Average number of ratings per user	4.045	3.892	1.823
Maximum number of ratings per item	218	216	38
Minimum number of ratings per item	1	1	1
Average number of ratings per item	2.149	1.996	1.406
Number of users who ranked at least 5 items	13241	11548	962
Number of users who ranked at least 10 items	5369	4639	190
Number of users who ranked at least 50 items	289	244	4
Number of users who ranked at least 100 items	66	54	0
Sparsity	0.999971	0.999970	0.999922

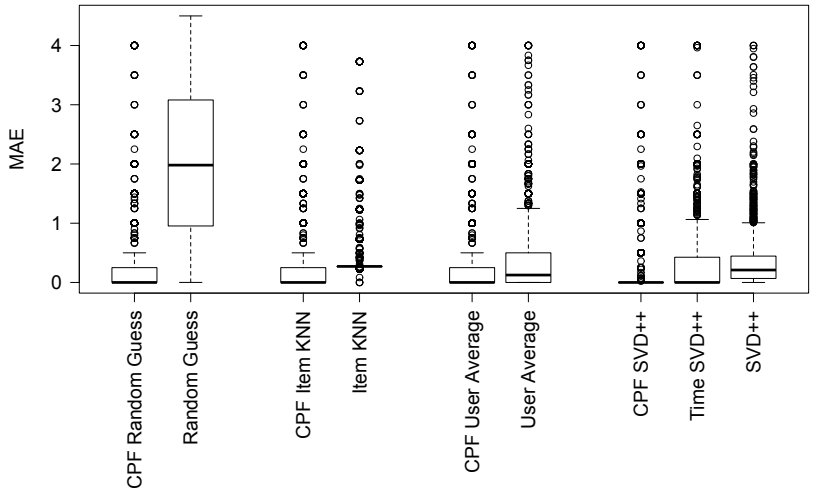


Fig. 10. Boxplots of MAE values of different algorithms computed per user on subset with numeric ratings

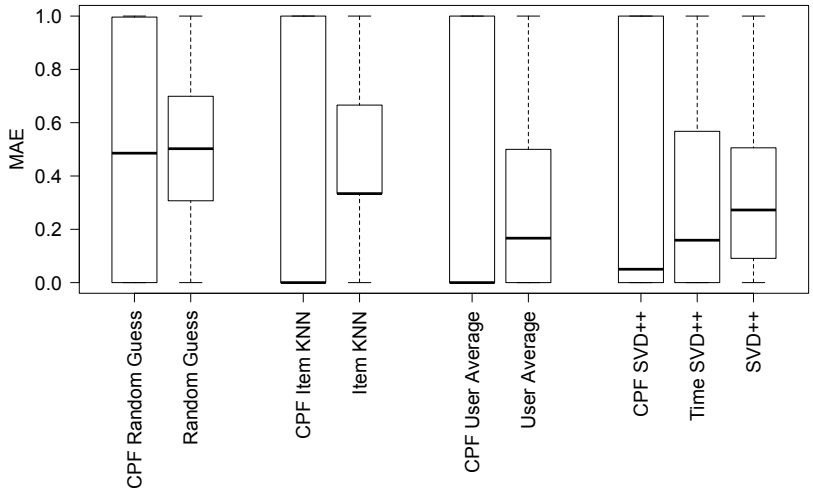


Fig. 11. Boxplots of MAE values of different algorithms computed per user on subset with descriptive ratings

TABLE III
MAE VALUES COMPUTED FOR WHOLE TEST SETS

	Numeric ratings		Descriptive Ratings	
Contextual pre-filtering	**YES**	**NO**	**YES**	**NO**
Random Guess	0.2315	2.0998	0.4694	0.4989
User Average	0.2312	0.3026	0.3624	0.2570
Item kNN	0.2312	0.3976	0.3624	0.4374
SVD++	0.2514	0.3511	0.3621	0.3101
Time SVD++	NA	0.2693	NA	0.2975

subset are better. Our contextual pre-filtering combined with classical SVD++ performs better than Time SVD++. There could be two reasons for this behavior. First, the usage of more contextual parameters than just one, the time, gives more improvement to prediction accuracy. Second, our approach (even if used with time parameter only) with SVD++ is truly better than Time SVD++ algorithm. This should be addressed in further work.

From Figure 11 we see that a median value for our approach is improved for all algorithms but the overall MAE value for *descriptive scale* subset is worst for all of the cases. This suggest that in the case of binary scale (*yes*/*maybe*) contextual pre-filtering may increase the sparsity and noisiness of the data. Thus, the recommendation algorithm may not always predict the rating. However, the difference between results for two subsets could be caused not by wrong pre-filtering method but by psychological differences between *a priori* and *a posteriori* evaluation by a user. It is more reliable when a user evaluates an item after he consumed it than when he declares what he would do or prefer. This could lead us to conclusion that this approach could be successfully applied in recommender systems where numeric scale is used to rate items in *a posteriori* way. Currently, we have not identified any other limitations for applying proposed contextual pre-filtering approach.

To check the statistical significance of the results, we applied Wilcoxon test with *p-value* 0.01. We chose this statistical test because we cannot guarantee the normal distribution of obtained results. The test confirmed the statistical significance of our results.

VI. CONCLUSIONS AND FUTURE WORK

In this paper we presented a new approach for contextual pre-filtering in Recommender Systems. It is based on two ontologies: Recommender System Context (RSCtx), which represents the context, and Contextual Ontological User Profile (COUP), which represents user preferences. RSCtx extends PRISSMA and represents different context dimensions on different granularity levels. COUP was modeled according to SIM approach for modularization. Different users' parts of profile are represented in different ontological module. This allows us to: (I) store multiple users in one ontology, (II) clearly distinguishing user preferences from different domains, but keeping all the user preferences together, and (III) split user interests from one domain into "micro profiles" related to some contextual situation without loosing the possibility to reason on different level of context granularity.

We successfully applied RSCtx for context identification and generalization tasks, showing that it is possible to represent context by combining different dimensions and representing different granularities for each dimension. We used COUP for representing user preferences in different context in the domain of musical events and for obtaining user data relevant to his current context for rating prediction task with baseline algorithms. This offline study showed that the usage of proposed ontologies with recommendation algorithms could significantly improve the accuracy of prediction according to MAE measure. This confirmed part of the second research question, i.e. that it is possible to represent user preferences and items in such a way that can be combined with different recommendation approaches. The next step in our research is proving that we can adapt our user representation for different domains.

As future work, we plan to extend our experiment to ranking task as well as to investigate on the influence of the proposed approach on diversity and novelty of recommendations.

REFERENCES

[1] F. Ricci, L. Rokach, and B. Shapira, "Introduction to recommender systems handbook," in *Recommender Systems Handbook*, F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor, Eds. Springer US, 2011, pp. 1–35. ISBN 978-0-387-85819-7. [Online]. Available: http://dx.doi.org/10.1007/978-0-387-85820-3_1

[2] G. Adomavicius and A. Tuzhilin, *Recommender Systems Handbook*. Boston, MA: Springer US, 2011, ch. Context-Aware Recommender Systems, pp. 217–253. ISBN 978-0-387-85820-3. [Online]. Available: http://dx.doi.org/10.1007/978-0-387-85820-3_7

[3] K. Goczyla, W. Waloszek, and A. Waloszek, "Contextualization of a DL knowledge base," in *Proc. of the 2007 Int. Workshop on Description Logics (DL2007)*, 2007. [Online]. Available: http://ceur-ws.org/Vol-250/paper_55.pdf

[4] C. Bettini, O. Brdiczka, K. Henricksen, J. Indulska, D. Nicklas, A. Ranganathan, and D. Riboni, "A survey of context modelling and reasoning techniques," *Pervasive and Mobile Computing*, vol. 6, no. 2, pp. 161 – 180, 2010. doi: 10.1016/j.pmcj.2009.06.002 Context Modelling, Reasoning and Management. [Online]. Available: http://dx.doi.org/10.1016/j.pmcj.2009.06.002

[5] C. Bolchini, C. A. Curino, E. Quintarelli, F. A. Schreiber, and L. Tanca, "A data-oriented survey of context models," *SIGMOD Rec.*, vol. 36, no. 4, pp. 19–26, Dec. 2007. doi: 10.1145/1361348.1361353. [Online]. Available: http://dx.doi.org/10.1145/1361348.1361353

[6] R. Krummenacher and T. Strang, "Ontology-based context modeling," in *In Workshop on Context-Aware Proactive Systems*, 2007.

[7] J. Ye, L. Coyle, S. Dobson, and P. Nixon, "Ontology-based models in pervasive computing systems," *Knowl. Eng. Rev.*, vol. 22, no. 4, pp. 315–347, Dec. 2007. doi: 10.1017/S0269888907001208. [Online]. Available: http://dx.doi.org/10.1017/S0269888907001208

[8] L. Costabello, *Context-Aware Access Control and Presentation for Linked Data*, 2013, ch. A Declarative Model for Mobile Context, pp. 21–32.

[9] T. Heath and C. Bizer, *Linked Data: Evolving the Web into a Global Data Space*, 1st ed. Morgan & Claypool, 2011. ISBN 9781608454303. [Online]. Available: http://dx.doi.org/10.2200/S00334ED1V01Y201102WBE001

[10] H. Chen, T. Finin, and A. Joshi, *Ontologies for Agents: Theory and Experiences*. Basel: Birkhäuser Basel, 2005, ch. The SOUPA Ontology for Pervasive Computing, pp. 233–258. ISBN 978-3-7643-7361-0. [Online]. Available: http://dx.doi.org/10.1007/3-7643-7361-X_10

[11] T. Strang, C. Linnhoff-Popien, and K. Frank, *Distributed Applications and Interoperable Systems: 4th IFIP WG6.1 Int. Conf., DAIS 2003, Paris, France, November 17-21, 2003. Proc.* Berlin, Heidelberg: Springer Berlin Heidelberg, 2003, ch. CoOL: A

Context Ontology Language to Enable Contextual Interoperability, pp. 236–247. ISBN 978-3-540-40010-3. [Online]. Available: http://dx.doi.org/10.1007/978-3-540-40010-3_21

[12] X. H. Wang, D. Q. Zhang, T. Gu, and H. K. Pung, "Ontology based context modeling and reasoning using owl," in *Proc. of the Second IEEE Annual Conf. on Pervasive Computing and Communications Workshops*, ser. PERCOMW '04. Washington, DC, USA: IEEE Computer Society, 2004. doi: 10.1109/PERCOMW.2004.1276898. ISBN 0-7695-2106-1 pp. 18–. [Online]. Available: http://dx.doi.org/10.1109/PERCOMW.2004.1276898

[13] D. Preuveneers, J. Bergh, D. Wagelaar, A. Georges, P. Rigole, T. Clerckx, Y. Berbers, K. Coninx, V. Jonckers, and K. Bosschere, *Ambient Intelligence: Second European Symposium, EUSAI 2004, Eindhoven, The Netherlands, November 8-11, 2004. Proc.* Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, ch. Towards an Extensible Context Ontology for Ambient Intelligence, pp. 148–159. ISBN 978-3-540-30473-9. [Online]. Available: http://dx.doi.org/10.1007/978-3-540-30473-9_15

[14] P. Korpipää and J. Mäntyjärvi, *Modeling and Using Context: 4th Int. and Interdisciplinary Conf. CONTEXT 2003 Stanford, CA, USA, June 23–25, 2003 Proc.* Berlin, Heidelberg: Springer Berlin Heidelberg, 2003, ch. An Ontology for Mobile Device Sensor-Based Context Awareness, pp. 451–458. ISBN 978-3-540-44958-4. [Online]. Available: http://dx.doi.org/10.1007/3-540-44958-2_37

[15] R. Hervás and J. Bravo, "Towards the ubiquitous visualization: Adaptive user-interfaces based on the semantic web," *Interacting with Computers*, vol. 23, no. 1, pp. 40 – 56, 2011. doi: http://dx.doi.org/10.1016/j.intcom.2010.08.002. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0953543810000676

[16] G. D. Abowd, A. K. Dey, P. J. Brown, N. Davies, M. Smith, and P. Steggles, "Towards a better understanding of context and context-awareness," in *Proc. of the 1st Int. Symposium on Handheld and Ubiquitous Computing*, ser. HUC '99. London, UK, UK: Springer-Verlag, 1999. ISBN 3-540-66550-1 pp. 304–307. [Online]. Available: http://dx.doi.org/10.1007/3-540-48157-5_29

[17] M. Kaminskas and F. Ricci, "Contextual music information retrieval and recommendation: State of the art and challenges," *Computer Science Review*, vol. 6, no. 2–3, pp. 89 – 119, 2012. doi: 10.1016/j.cosrev.2012.04.002. [Online]. Available: http://dx.doi.org/10.1016/j.cosrev.2012.04.002

[18] R. Cai, C. Zhang, C. Wang, L. Zhang, and W.-Y. Ma, "Musicsense: Contextual music recommendation using emotional allocation modeling," in *Proc. of the 15th ACM Int. Conf. on Multimedia*, ser. MM '07. New York, NY, USA: ACM, 2007. doi: 10.1145/1291233.1291369. ISBN 978-1-59593-702-5 pp. 553–556. [Online]. Available: http://dx.doi.org/10.1145/1291233.1291369

[19] L. Baltrunas, B. Ludwig, S. Peer, and F. Ricci, "Context relevance assessment and exploitation in mobile recommender systems," *Personal Ubiquitous Comput.*, vol. 16, no. 5, pp. 507–526, Jun. 2012. doi: 10.1007/s00779-011-0417-x. [Online]. Available: http://dx.doi.org/10.1007/s00779-011-0417-x

[20] Z. Su, J. Yan, H. Ling, and H. Chen, "Research on personalized recommendation algorithm based on ontological user interest model," *J. of Computational Information Systems*, vol. 8, no. 1, pp. 169–181, Jan. 2012.

[21] C. Rack, S. Arbanowski, and S. Steglich, "Context-aware, Ontology-based Recommendations," in *SAINT-W '06: Proc. of the Int. Symposium on Applications on Internet Workshops*. Washington, DC, USA: IEEE Computer Society, 2006. doi: 10.1109/saint-w.2006.13. ISBN 0769525105 pp. 98–104. [Online]. Available: http://dx.doi.org/10.1109/saint-w.2006.13

[22] I. Cantador, A. Bellogín, and P. Castells, "Ontology-based personalised and context-aware recommendations of news items," in *Proc. of the 2008 IEEE/WIC/ACM Int. Conf. on Web Intelligence and Intelligent Agent Technology - Volume 01*, ser. WI-IAT '08. Washington, DC, USA: IEEE Computer Society, 2008. doi: 10.1109/WIIAT.2008.204. ISBN 978-0-7695-3496-1 pp. 562–565. [Online]. Available: http://dx.doi.org/10.1109/WIIAT.2008.204

[23] J. Rodríguez, M. Bravo, and R. Guzmán, "Multidimensional ontology model to support context-aware systems," 2013. [Online]. Available: http://www.aaai.org/ocs/index.php/WS/AAAIW13/paper/view/7187

[24] A. Hawalah and M. Fasli, "Utilizing contextual ontological user profiles for personalized recommendations," *Expert Systems with Applications*, vol. 41, no. 10, pp. 4777 – 4797, 2014. doi: http://dx.doi.org/10.1016/j.eswa.2014.01.039. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0957417414000633

[25] S. E. Middleton, N. R. Shadbolt, and D. C. De Roure, "Ontological user profiling in recommender systems," *ACM Trans. Inf. Syst.*, vol. 22, no. 1, pp. 54–88, Jan. 2004. doi: 10.1145/963770.963773. [Online]. Available: http://doi.acm.org/10.1145/963770.963773

[26] B. Mobasher, X. Jin, and Y. Zhou, *Web Mining: From Web to Semantic Web: First European Web Mining Forum, EWMF 2003, Invited and Selected Revised Papers*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, ch. Semantically Enhanced Collaborative Filtering on the Web, pp. 57–76. ISBN 978-3-540-30123-3. [Online]. Available: http://dx.doi.org/10.1007/978-3-540-30123-3_4

[27] S. S. Anand, P. Kearney, and M. Shapcott, "Generating semantically enriched user profiles for web personalization," *ACM Trans. Internet Technol.*, vol. 7, no. 4, Oct. 2007. doi: 10.1145/1278366.1278371. [Online]. Available: http://doi.acm.org/10.1145/1278366.1278371

[28] P. Lops, M. de Gemmis, and G. Semeraro, *Recommender Systems Handbook*. Boston, MA: Springer US, 2011, ch. Content-based Recommender Systems: State of the Art and Trends, pp. 73–105. ISBN 978-0-387-85820-3. [Online]. Available: http://dx.doi.org/10.1007/978-0-387-85820-3_3

[29] T. Di Noia and V. C. Ostuni, *Reasoning Web. Web Logic Rules: 11th Int. Summer School 2015, Berlin, Germany, July 31-August 4, 2015, Tutorial Lectures.* Cham: Springer International Publishing, 2015, ch. Recommender Systems and Linked Open Data, pp. 88–113. ISBN 978-3-319-21768-0. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-21768-0_4

[30] A. K. Dey, "Understanding and using context," *Personal Ubiquitous Comput.*, vol. 5, no. 1, pp. 4–7, Jan. 2001. doi: 10.1007/s007790170019. [Online]. Available: http://dx.doi.org/10.1007/s007790170019

[31] M. Fernández-López, A. Gómez-Pérez, and N. Juristo, "Methontology: from ontological art towards ontological engineering," in *Proc. Symposium on Ontological Engineering of AAAI*, 1997.

[32] D. Heckmann, T. Schwartz, B. Brandherm, M. Schmitz, and M. Wilamowitz-Moellendorff, *User Modeling 2005: 10th Int. Conf., UM 2005, Edinburgh, Scotland, UK, July 24-29, 2005. Proc.* Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, ch. Gumo – The General User Model Ontology, pp. 428–432. ISBN 978-3-540-31878-1. [Online]. Available: http://dx.doi.org/10.1007/11527886_58

[33] J. A. Russell, "A circumplex model of affect." *J. of Personality and Social Psychology*, vol. 39, no. 6, pp. 1161–1178, Dec. 1980. doi: 10.1037/h0077714. [Online]. Available: http://dx.doi.org/10.1037/h0077714

[34] A. Mehrabian, "Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament," *Current Psychology*, vol. 14, no. 4, pp. 261–292. doi: 10.1007/BF02686918. [Online]. Available: http://dx.doi.org/10.1007/BF02686918

[35] K. Goczyła, A. Waloszek, W. Waloszek, and T. Zawadzka, *Intelligent Tools for Building a Scientific Information Platform*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, ch. Modularized Knowledge Bases Using Contexts, Conglomerates and a Query Language, pp. 179–201. ISBN 978-3-642-24809-2. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-24809-2_11

[36] K. Goczyla, A. Waloszek, and W. Waloszek, "Towards context-semantic knowledge bases," in *Federated Conf. on Computer Science and Information Systems - FedCSIS 2012, Wroclaw, Poland, 9-12 September 2012, Proc.*, M. Ganzha, L. A. Maciaszek, and M. Paprzycki, Eds., 2012. ISBN 978-83-60810-51-4 pp. 475–482. [Online]. Available: https://fedcsis.org/proceedings/2012/pliks/388.pdf

[37] A. Karpus and K. Goczyla, "A multi-domain hybrid recommender systems based on a dynamic contextual ontological user profile," in *Doctoral Consortium (IC3K 2014)*, 2014. doi: 10.5220/0005174300830087. ISBN Not Available pp. 83–87. [Online]. Available: http://www.scitepress.org/DigitalLibrary/PublicationsDetail.aspx?ID=6Kh9MDlu7qs=&t=1

[38] P. Adamopoulos and A. Tuzhilin, "Estimating the Value of Multi-Dimensional Data Sets in Context-based Recommender Systems," in *8th ACM Conf. on Recommender Systems (RecSys 2014)*, 2014.

[39] Y. Koren, "Collaborative filtering with temporal dynamics," in *Proc. of the 15th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, ser. KDD '09. New York, NY, USA: ACM, 2009. doi: 10.1145/1557019.1557072. ISBN 978-1-60558-495-9 pp. 447–456. [Online]. Available: http://doi.acm.org/10.1145/1557019.1557072

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 421–428
DOI: 10.15439/2016F141
ACSIS, Vol. 8. ISSN 2300-5963

Predicting Star Ratings based on Annotated Reviews of Mobile Apps

Dagmar Monett and Hermann Stolte
Computer Science Department
Faculty of Cooperative Studies
Berlin School of Economics and Law, Germany
Email: {dagmar.monett-diaz, hermann.stolte}@hwr-berlin.de

***Abstract*—This paper presents and evaluates different computational models for review rating prediction. The models rely solely on star ratings from an annotated corpus of customer reviews of mobile apps that were collected from the Google Play Store in a related work. Fine-granular opinions and the classification of their sentiment orientation were already available. The models build upon them to make predictions based on their polarity. Predicting star ratings is of importance to the sentiment analysis community because it can better be understood how customers subjectively rate products. Rating them consistently with corresponding written reviews, however, remains a difficult task for automated predictors. This paper sheds new light in that direction.**

***Index Terms*—Mobile apps, review rating prediction, semantic sentiment analysis.**

I. Introduction

MOBILE app star ratings and reviews drive apps' rankings, downloads, updates, and in-app purchases. That is what a study from Apptentive has found after surveying smartphone owners and after analysing "historical data from delivering over 160 million interactions and ratings prompts" [1]. According to the study, both star ratings and reviews strongly influence not only the success of mobile apps but also the consumers' engagement with them.

The analysis and interpretation of mobile app star ratings and reviews are not straightforward tasks, however. Monitoring star ratings and reviews is expensive, difficult to accomplish, laborious, and error-prone [2]; many of the ratings and reviews are biased (e.g. app users are more likely to leave ratings or reviews after a negative experience with the app [1]); reviews are in general short and often use abbreviations, emoticons, and informal language; and even star ratings are sometimes unrelated to the experiences with the app itself (e.g. [3] analyses how people give poor ratings just because they are asked to rate the app, explicitly).

Star ratings and reviews are extremely important for brands, for example, for improving their products based on customers' feedback. Ratings also matter for marketing purposes and companies' reputation: it is not only crucial that a top app is highly rated but also that it has at least four stars and many ratings. According to Walz [4], 88% of top-100 Android apps (51% of top-100 iOS apps) have a rating greater than four stars, and the average top-100 Android app (top-100 iOS app) has 3.1 million (196 thousand) ratings.[1] But how to predict or to influence users' star ratings?

Star ratings and reviews are also crucial for customers and their future behaviour when using and recommending the apps. If new customers trust an app's ratings and reviews, then they are more willing to download the app and to benefit from its functionality, e.g. to buy products easily, or to connect and communicate instantly with others, or to simplify daily activities at the office, to name a few benefits. If their experiences with the app are positive, then they would recommend it further and even give feedback to the company for improvements to the app: a win-win situation. Although customers and companies value feedback differently [5], it is true that not only star ratings but also the reviews' content play an important role for both parts.

However, could we *teach* users how to rate apps *consistently* with the review they are writing for a mobile app? For example, would it be possible to improve recommendation accuracy by suggesting to users the most adequate star rating they should give to a product depending on the semantic orientation of what they have already written in the review? How does it compare to previously reviewed mobile apps? Would an improvement in the accuracy also mean an improvement of users' engagement and satisfaction with the apps?

The remaining sections of this paper continue as follows: Section II introduces both the task of review rating prediction and related work in this area. A corpus of annotated reviews of mobile apps from different domains that is used for analysis is presented in Section III. Computational models that are proposed to predict star ratings based on the annotated reviews of the corpus are topic of Section IV. These models are analysed and evaluated in several experimental settings that are defined in Section V. Finally, results are discussed before the conclusions of the paper are presented together with some ideas for further work.

II. Related Work

The prediction of star ratings (e.g., ratings ranging from 1 to 5 stars) has been the focus of many academic and business applications to date. In particular, *review rating prediction*,

[1]Figures from June 2015.

also known as *sentiment rating prediction*, is a task that deals with the inference of an author's implied numerical rating, i.e. on the prediction of a rating score, from a given written review [6], [7]. Recommendation systems, for instance, often suggest products based on star ratings of similar products previously rated by other users.

Yet analysing a textual review is a much more difficult task than guessing the rating by only considering other available numerical scores. This is why not only classifying sentiment [8], [9] but also predicting rating scores has captured the attention of the sentiment analysis community in the last few years. For example, Pang and Lee apply classification and regression, supervised learning techniques to rate movie reviews [10], and Goldberg and Zhu extend their approach by applying a graph-based semi-supervised learning algorithm that achieves better performance [11]. Tang and co-authors follow a similar approach [12], and present a neural network-based method that considers not only the review texts but also author information. They claim that their method "performs better than several strong baseline methods which only use textual semantics." Li and co-authors go beyond the review texts and their authors, and add information also about the product that is reviewed, by modelling all three features using a three-dimensional tensor [13]. Then, they apply tensor factorisation techniques and optimise their model using gradient descent. Their results outperform other similar approaches. Furthermore, Qu et al. introduce the *bag-of-opinions* representation for which their method learns rating scores from domain-independent corpora using constrained ridge regression [14].

Zhang and co-authors delve deeper into the polarity[2] of a review by stating that "it might not be appropriate to use overall ratings as ground-truth to label the sentiment orientations of review texts, as users tend to act differently when making overall ratings and expressing their true feelings on detailed product aspects or features" [15]. This means that rating predictors should consider the subtle differences between review texts as a whole, and reviews of individual aspects. [16] and [17] come to the same conclusions, and affirm that textually derived ratings are better predictors than numerical star ratings. In their experiments, Zhang and co-authors first let three annotators manually label the polarity orientation of sample reviews from a restaurant dataset and then compare them against automatically generated annotations using unsupervised review-level sentiment classification [15]. Afterwards, the annotators label not reviews as a whole but their aspects or features individually. Again, the results are compared to those obtained with the methods the authors propose, showing the inconsistency between textual reviews and numerical ratings when the latter do not consider phrase-level sentiment polarity.

Gupta and co-authors also apply supervised learning with a multi-aspect rating prediction for textual reviews of restaurants [18]. They consider numerical ratings for aspects like food, service, and overall experience, inter alia, as well as considering the interdependence of aspects for around eight sentences per review on average. Orimaye and co-authors introduce a sentence-level polarity correction [19]. Their technique identifies sentences with inconsistent polarities that are handled as outliers and, as such, are discarded from the reviews. This approach might not be convenient for mobile app reviews, where the length of subjective phrases might be about two words long on average, and the reviews are not long enough either [20]. Discarding information in the case of mobile apps would introduce an extra bias to the problem.

Sänger [20] introduces an aspect-based opinion mining of mobile apps ratings that extends Klinger and Cimiano's work [21], [22]. According to Sänger, Klinger and Cimiano's approach was chosen because it deals with fine-granular aspect-based opinion mining, its implementation is open-sourced (see https://bitbucket.org/rklinger/jfsa), and it is suitable for mining text written in German, as is the case of the dataset he uses (see next section). Sänger concludes that such a technique is also appropriate for analysing mobile app reviews; he both adapts and validates Klinger and Cimiano's work for such reviews.

Sänger's approach serves as the background to, and the basis for, the work presented here. It is worth mentioning, however, that the goal of the work presented in this paper *is not* to deal with aspect identification nor with sentiment classification; but assuming that these tasks are performed *before* the star ratings are predicted. A complement to Sänger's work, in other words. Thus, *unlike* other approaches that identify aspects or classify sentiment at a fine-granular level, like most of the works reviewed above (e.g. [10]–[12], [17], [21], to cite but a few), the idea of our approach is to provide a method for predicting star ratings based *solely* on available annotated, fine-granular opinions.

The next section introduces the dataset that is used for analysis and validation.

III. Corpus of Annotated Customer Reviews

The annotated corpus used here was initially provided by Sänger as constructed in [20], later named *SCARE* as introduced in [23]. It consists of 1,760 randomly selected, annotated reviews for a total of 130 mobile apps from different domains. The annotations consider fine-granular opinions as well as the app aspects and their relationships. Each textual review includes a customer evaluation of the app, and has an associated rating. All textual reviews are in German. Each evaluation consists of at least one phrase. There is a total of 6,446 phrases from which 3,959 are manually annotated subjective phrases. The corpus contains a total of 2,487 aspects.

Sänger claims that his mobile app dataset is the first of its kind. It comprises a total of 802,860 reviews in German of 148 mobile apps from 11 different categories, the annotated corpus introduced above being a subset of it. The reviews were collected from the Google Play Store (see https://play.google.

[2]See next section for more on polarity.

com/store) using an open-source API for the Android Market (see https://code.google.com/archive/p/android-market-api/).

Specifically, the annotated corpus that is used here contains the following information, which follows the structure presented in [24]:

all.data A list of all reviews as retrieved through the Android Market API, including the app's name, the full review text, and the star rating given by the user.

all.txt A list of all review texts as they were used in the annotation process (the review title and its content are concatenated).

all.csv A list of all annotated subjective phrases and aspects, each subjective phrase with an internal ID, its corresponding ID, and its polarity.

all.rel A list of all annotated relations between the subjective phrases and their aspects.

Figure 1 shows all major steps of the prediction process that makes use of the annotated corpus. It starts by *parsing* the lists introduced above and by creating workspace variables with which to work.

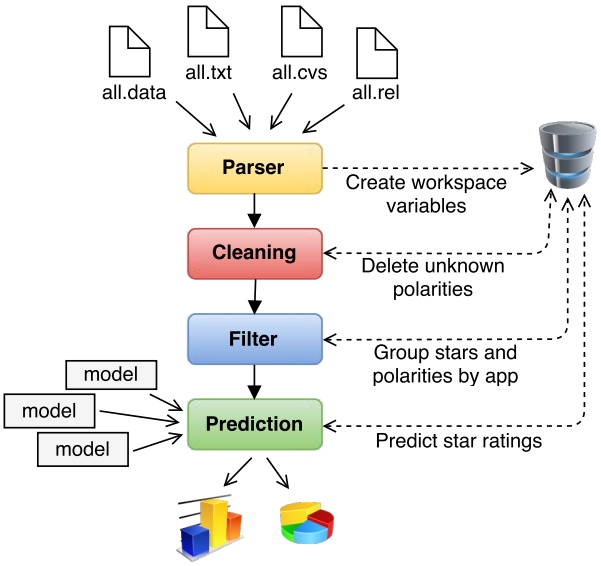


Fig. 1 Prediction process.

The polarity of a phrase depends on the expressed opinion, and thus on the semantic orientation or *sentiment* of the phrase, i.e., whether the expressed opinion of the opinion holder[3] is positive, negative, or neutral [7]. Since a review might have more than one phrase, calculating the polarity of the review would depend on the polarities of its phrases. In particular, mobile app reviews are much shorter than other product reviews, use language constructs that are similar to those used in micro-blogging (e.g., Twitter), have unstructured sentences in general, and often use more concise words [20].

According to Sänger [20], the subjective phrases were annotated and their polarity determined following a rigorous process that comprised the development of annotation guidelines, the explicit training of four annotators on these guidelines, the annotation of random phrases in iterative rounds, as well as a later controlling and improvement of the performed annotations. The final version of the annotations during the training process achieved a substantial inter-annotator agreement with a kappa value $\kappa = 0.72$, computed using the Fleiss' kappa measure (see Chapter 3 in [20] for more). Then, the actual annotations to be considered for the corpus were carried out.

It is worth mentioning that the polarity of type *unknown* was handled as a default value in the tool that was used for annotating the corpus (see http://brat.nlplab.org/). According to Sänger [25], this relates to reviews where the annotators forgot to specify the polarities. Because the phrase polarity was not of further interest in his work, there was no need to correct that issue. Thus, unknown sentiments are not considered for the experiments that will be introduced in succeeding sections: they are deleted from the corpus in a *cleaning* procedure (see Figure 1).

After cleaning the unknown polarities out, the new annotated corpus consists of 1,751 reviews, 130 apps, 6,398 phrases, and 3,927 subjective phrases. Table I shows the distribution of all phrases from the corpus according to their polarity, before and after the cleaning process has taken place. Almost two thirds of the subjective phrases express a positive opinion, and about one-third have a negative polarity.

TABLE I
POLARITY DISTRIBUTION OF ANNOTATED SUBJECTIVE PHRASES.

	Before cleaning		**After cleaning**	
Polarity	Annotated phrases	%	Annotated phrases	%
positive	2,463	62.2	2,458	62.6
negative	1,433	36.2	1,416	36.1
neutral	53	0.01	53	1.3
unknown	10	0.002	–	–

The star ratings associated with the entries from the corpus, i.e., to the annotated mobile apps reviews, after the cleaning process are summarised in Table II.

TABLE II
DISTRIBUTION OF STAR RATINGS.

Star rating	No. of annotated reviews	%
1	295	16.8
2	111	6.3
3	136	7.8
4	299	17.1
5	910	52.0

If reviews with 4-5 stars are considered positive reviews and those with 1-2 stars are considered negative reviews (the *thumbs-up-thumbs-down* approach suggested by Liu in [7]), then over two-thirds of the reviews from the annotated corpus have a positive polarity (69.1%) and only about one out of four reviews is negative (23.1%). Compared to the

[3]The person that holds the opinion [9]. Also, opinion source.

subjective phrases polarities from Table I, these are slightly smaller values (62.6% positive polarity). This means that the expressed opinions from the corpus are in general more positive when they are given as an overall numerical rating than when taking into account their individual subjective phrases (probably aspect-related) polarity. It can be observed in Figure 2 that the line depicting the average of star ratings is above the expected line averaging the subjective phrases polarity. The fine-granular analysis suggested by Klinger and Cimiano [21], [22] and extended by Sänger [20] confirms the findings from other approaches [15]–[17] with respect to the subtle differences between ratings of reviews as a whole and as differentiated subjective phrases.

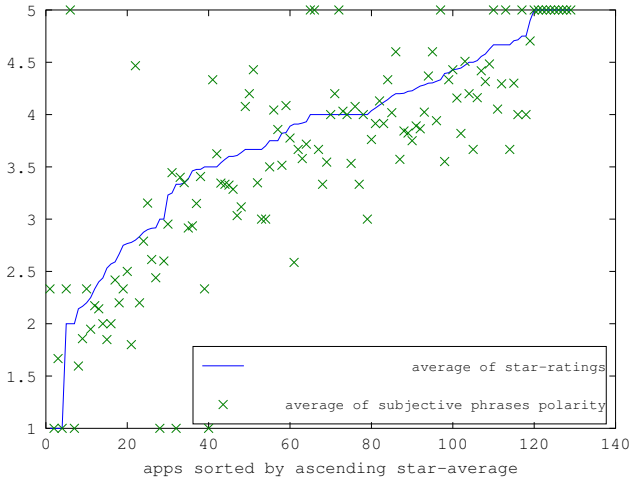


Fig. 2 Average of labelled star ratings versus average of subjective phrases polarity.

Figure 3 shows the number of star ratings and subjective phrases for each app after a *filtering* procedure (see Figure 1) that groups them together according to the reviews associated with that app. There are about twice as many subjective phrases than star ratings per app. They have a strong linear dependency: there is a positive correlation, with the Pearson's correlation coefficient $\rho = 0.8$. This is a good indicator for considering linear regression models that can predict the star ratings without human intervention.

Yet another possibility to plot the data from the corpus is shown in Figure 4. This time, the number of reviews per app is taken into account. Such a visualisation was helpful when analysing apps according to their importance or to the number of reviews that are provided. We do not consider further implications in our experiments but were better aware of the distribution of the ratings when analysing the data.

Not only is a visual analysis of the data concerning the number of reviews and their ratings interesting, but also in which relation stay positive and negative opinions to each other. As can be seen in Figure 5, negative reviews have higher impact than positive reviews. There is a negative correlation between both of them, with Pearson's correlation coefficient $\rho = -0.78$ (apps with no positive subjective phrases were

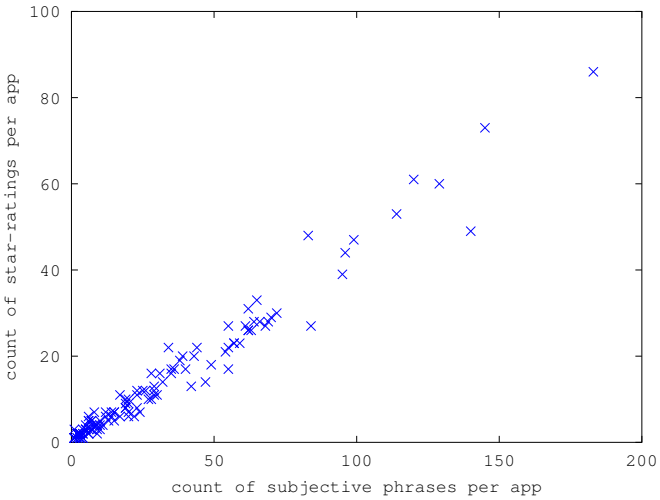


Fig. 3 Number of star ratings and subjective phrases for each app in the annotated corpus.

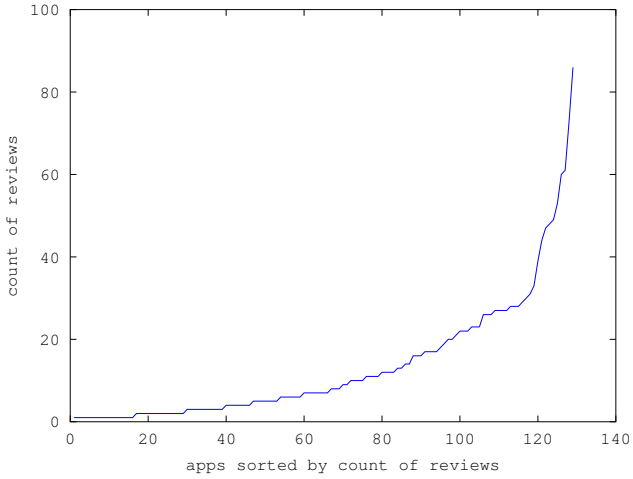


Fig. 4 Count of reviews per app sorted in ascending order.

filtered out to avoid indetermination when considering the negative vs. positive sentiment ratio).

All these observations determined which computational models should be considered in order to predict star ratings. Some of these models will be presented in the next section. In this paper and for the reasons commented above (like the strong linear dependency between subjective phrases and star ratings per app), we place great emphasis on multivariate regression models.

IV. Prediction of Star Ratings

Let $h_\Theta : \mathbb{R}^{n+1} \to \mathbb{R}$ be the hypothesis of a multivariate regression model,

$$h_\Theta(\mathbf{x}) = \theta_0 x_0 + \theta_1 x_1 + \cdots + \theta_n x_n = \Theta^\intercal \mathbf{x}, \qquad (1)$$

with $\Theta \in \mathbb{R}^{n+1}$ being a vector of parameters, $\mathbf{x} \in \mathbb{R}^{n+1}$ being a vector of features or independent variables, $n \in \mathbb{N}$, and $i = 0, \ldots, n$.

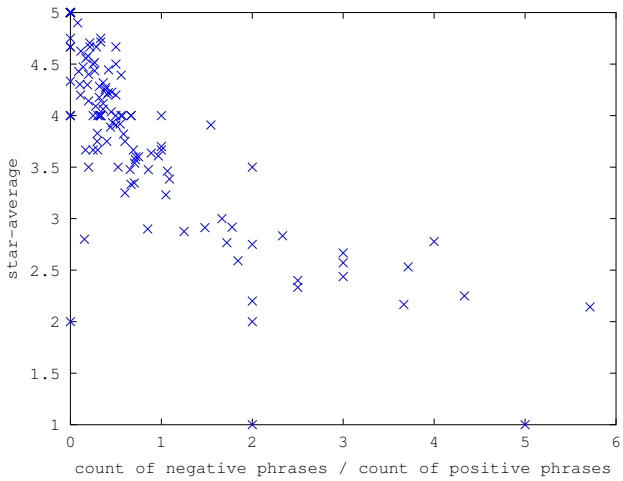


Fig. 5 Star average according to the negative vs. positive sentiment ratio.

The cost function which is to be minimized in order to find the optimal values of the parameters θ_i is the following:

$$c(\Theta) = \frac{1}{2m} \sum_{j=1}^{m} (h_\Theta(\mathbf{x})^{(j)} - y^{(j)})^2, \tag{2}$$

with $m = 1701$ being the number of reviews in the corpus and y being the dependent variable or star rating for each annotated review j.

For predicting star ratings of mobile apps, a model with four variables (or features) could be considered, where

- x_0 is equal to 1 for convenience of notation,
- x_1 is the number of subjective phrases with positive polarity,
- x_2 is the number of subjective phrases with negative polarity, and
- x_3 is the number of subjective phrases with neutral polarity.

Let $h_{1_\Theta} : \mathbb{R}^4 \to \mathbb{R}$ be the corresponding hypothesis:

$$h_{1_\Theta}(\mathbf{x}) = \theta_0 x_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3. \tag{3}$$

This is the *baseline model*.

In case the neutral polarities are not considered, as will be discussed in the next section, the above model can be simplified as follows:

$$h_{2_\Theta}(\mathbf{x}) = \theta_0 x_0 + \theta_1 x_1 + \theta_2 x_2, \tag{4}$$

with $h_{2_\Theta} : \mathbb{R}^3 \to \mathbb{R}$.

Since some apps might have many more reviews than others, the values of the features could be normalised using the following scaling:

$$x_i' = \frac{x_i - \mu_i}{\sigma_i}, \tag{5}$$

with μ_i the mean value and σ_i the standard deviation of the feature i in the vector of features $\mathbf{x}$.

An average-based, simpler model could also be considered by taking into account only the average value of the polarities of a review (e.g., average polarity between all positive, negative, and neutral sentiments of the review) in one feature. Let $h_{3_\Theta} : \mathbb{R}^2 \to \mathbb{R}$ be the hypothesis for that case:

$$h_{3_\Theta}(\mathbf{x}) = \theta_0 x_0 + \theta_1 x_1. \tag{6}$$

The average polarity (numerical) value of a review can be calculated by mapping the polarities to the following values: 5 for a positive polarity, 3 for a neutral polarity, and 1 for a negative polarity.

The average polarity value can also be calculated by considering the *review rating score* (RRS) as suggested in [16] and [17]. This would mean that only the positive and negative polarities are taken into account, and are summed up using the following formula:

$$RRS^{(j)} = \left(\frac{P^{(j)}}{P^{(j)} + N^{(j)}} \cdot 4 \right) + 1, \tag{7}$$

where $P^{(j)}$ is the number of positive subjective phrases in review j, $N^{(j)}$ is the number of negative subjective phrases, and $1 \leq j \leq m$. As in [16], the new rating is scaled in the range of the corpus star rating (i.e., one to five stars).

Even a polarity ratio can be computed, too, where only the proportion between negative and positive polarities is taken into account.

Altogether, eight different models will be analysed and evaluated in the experiments that are introduced in the next section. They are summarised in Table III.

TABLE III Overview of prediction models.

Model	Hypothesis	Neutral polarities	Features normalised	Polarity average	RSS average	Polarity ratio
M1	h_{1_Θ}	✓				
M2	h_{1_Θ}	✓	✓			
M3	h_{2_Θ}					
M4	h_{2_Θ}		✓			
M5	h_{3_Θ}	✓		✓		
M6	h_{3_Θ}			✓		
M7	h_{3_Θ}				✓	
M8	h_{3_Θ}					✓

V. Experiments

Two different groups of experiments are considered for predicting the star ratings of mobile apps based on the expressed opinions from each review. All rely only on the polarity of the subjective phrases that are included in the annotated corpus.

The first group of experiments deals with assessing the importance of sentiment in the reviews. For example, whether to filter *neutral* phrases out from the corpus or not is investigated by applying different regression models, as introduced in the section above. Furthermore, filtering reviews

with no sentiment out (i.e., those that do not contain subjective phrases at all) is also analysed.

The second group of experiments makes use of other predictors, as suggested in [16] and [17], after considering the results of the first group of experiments.

Each individual experiment is run 10,000 times. A Monte Carlo cross-validation[4] is applied each time: on each iteration, the annotated reviews dataset is randomly partitioned into a 70% training dataset that is used to train the model in a supervised manner, and into a 30% testing dataset that is used to validate it.

A. Multi-variate linear regression-based predictors

Ganu et al. point out that neutral polarities do not add significant information to their experiments [16]. This could be also the case for the sentiment rating prediction of mobile apps that are used here. In order to investigate this, some of the regression models introduced in Section IV are trained and evaluated both with and without taking into account the neutral sentiments. Furthermore, they are also trained and evaluated with a reduced corpus that does not contain reviews that have no subjective phrases at all, i.e., reviews with no positive, neutral, or negative phrases are filtered out from the corpus. Concretely, a total of 77 reviews are filtered out.

The first experiment, experiment *E1*, considers the polarity count and evaluates the baseline regression model, i.e., hypothesis $h_{1_\Theta}(\mathbf{x})$ from Equation 3 and hypothesis $h_{2_\Theta}(\mathbf{x})$ from Equation 4. In other words, models *M1* and *M3* are evaluated, i.e., with and without neutral polarities.

The second experiment, experiment *E2*, uses the average-based hypothesis $h_{3_\Theta}(\mathbf{x})$ from Equation 6 for the training. Models *M5* and *M6* are evaluated, i.e., with and without neutral polarities.

The third experiment, experiment *E3*, considers the baseline model and the model without neutral polarities, i.e., hypotheses $h_{1_\Theta}(\mathbf{x})$ and $h_{2_\Theta}(\mathbf{x})$, both with normalised features. Models *M2* and *M4* are evaluated.

B. Univariate, average-based predictors

This group of experiments considers the RRS as defined in Equation 7.

First, an experiment *E4* with hypothesis $h_{3_\Theta}(\mathbf{x})$ is considered. In this case, model *M7* is evaluated.

A second experiment, experiment *E5*, also uses hypothesis $h_{3_\Theta}(\mathbf{x})$ but with the negative vs. positive polarities ratio, i.e., model *M8* is evaluated.

A third experiment, experiment *E7*, makes a *metadata-based prediction* (also similar to that proposed in [16]): given a new test review of an app, it predicts the rating by computing the average of all reviews available in the training set. A hypothesis like that from Equation 6 is considered and, with it, a new model *M9* is evaluated.

[4]The Monte Carlo cross-validation is a non-exhaustive cross-validation technique.

VI. Results and Discussion

The results show the averages of the mean squared errors (MSE) and the standard deviation σ for both the training and the test sets for each of the 10,000 iterations from the experiments. Together with these metrics, the value of the maximum error minus the minimum is also given.

Table IV shows the results for the first group of experiments, i.e., for those settings that evaluate not only the importance of neutral sentiment orientation but also whether reviews without subjective phrases should be included in the analysis or not.

The model that best predicts the star ratings is *M6* (see the last column of experiment *E2* in Table IV). This means that filtering both subjective phrases with neutral polarity *and* reviews with no sentiment orientation at all, fits much better the predictor (i.e., hypothesis $h_{3_\Theta}(\mathbf{x})$ from Equation 6) to the observed data.

In a second grade of importance are the best results that were obtained for experiments *E1* and *E3*. These are underlined. For our concrete corpus, it is not a good idea to normalise the model features: this does not improve the accuracy (see the second-last column of experiment *E3* in Table IV). Furthermore, models with more features profit from more data, as expected (see the first column of experiment *E1* in Table IV).

Figure 6 shows a visual comparison between the results of the first two experiments, *E1* and *E2*.

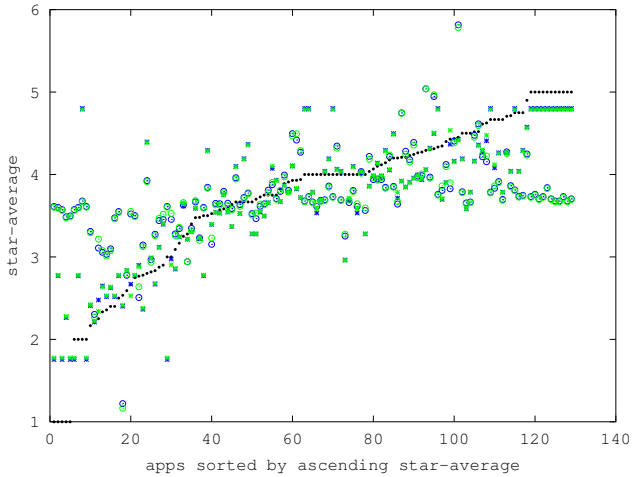


Fig. 6 Rating prediction for experiments *E1* and *E2*. Black asterisks: labels, blue (dark gray) circles: *E1* with neutral phrases, blue (dark gray) asterisks: *E2* with neutral phrases, green (light gray) circles: *E1* without neutral phrases, green (light gray) asterisks: *E2* without neutral phrases.

Since the hypothesis of the best model so far is h_{3_Θ}, then predicting the star rating for a new app given its review[5] would mean evaluating the hypothesis as follows:

$$h_{3_\Theta}(\mathbf{x}) = 1.0814 + 0.73538x_1,$$

[5]And after having classified the sentiment orientation of its subjective phrases.

TABLE IV Mobile apps rating prediction: Importance of sentiment in the reviews.

Experiments	with neutral phrases						without neutral phrases					
	with reviews with no subjective phrases			without reviews with no subjective phrases			with reviews with no subjective phrases			without reviews with no subjective phrases		
	MSE	σ	max-min	MSE	σ	max-min	MSE	σ	max-min	MSE	σ	max-min
E1: Linear regression with polarity count												
Training	0.60971	0.07354	0.51094	0.68953	0.08169	0.57188	0.60967	0.07364	0.50569	0.69099	0.08109	0.59013
Test	0.67642	0.17841	1.25400	0.75563	0.19703	1.57250	0.67660	0.17898	1.32320	0.75187	0.19591	1.49400
E2: Linear regression with polarity average												
Training	0.29072	0.04762	0.28186	0.26790	0.04754	0.25595	0.29048	0.04842	0.28524	**0.26720**	**0.04739**	**0.24608**
Test	0.31143	0.11380	0.72518	**0.28208**	0.11246	0.66817	0.31222	0.11589	0.69722	0.28359	**0.11206**	**0.59792**
E3: Linear regression with normalized polarity count												
Training	0.61055	0.07457	0.53986	0.68973	0.08230	0.60612	0.61063	0.07440	0.53547	0.68895	0.08144	0.57357
Test	0.67493	0.18186	1.48680	0.75434	0.19820	1.65960	0.67396	0.18094	1.35310	0.75660	0.19710	1.50860

where x_1 is the average of the positive and negative polarities of the review, and the intercept and the slope are the optimal parameters Θ that were found.

Table V shows the results for the second group of experiments.

TABLE V Mobile apps rating prediction: Other (average-based) predictors.

Experiments	with neutral phrases			without neutral phrases		
	MSE	σ	max-min	MSE	σ	max-min
E4: Linear regression with RRS						
Training	–	–	–	**0.23979**	**0.04484**	**0.21852**
Test	–	–	–	**0.25547**	**0.10604**	**0.50679**
E5: Linear regression with ratio neg/pos polarities						
Training	0.79105	0.08650	0.59050	0.87870	0.09371	0.65414
Test	0.82057	0.20284	1.55290	0.91346	0.21984	1.63780
E6: metadata-based prediction						
Training	–	–	–	–	–	–
Test	–	–	–	2.35960	0.08397	0.63069

If the review rating score is considered, i.e., model *M7*, then its results outperform all other predictions (see the final column of experiment *E4* in Table V).

Figure 7 shows a closer look when comparing the best models of both groups of experiments, i.e., *E2* and *E4*.

The predictions that are computed based on the review rating score are much closer to the star ratings given by the authors of the reviews, as Figure 8 clearly indicates (compared to those of Figure 2).

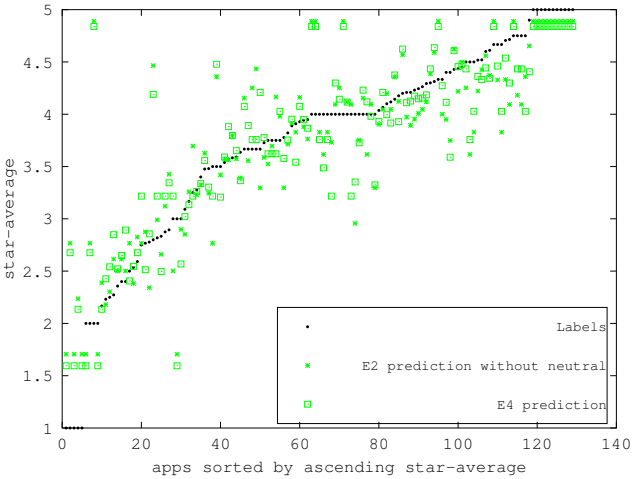


Fig. 7 Rating prediction for experiments *E2* and *E4*. Black dots: labels, green (light gray) asterisks: *E2* without neutral phrases, green (light gray) boxes: *E4* without neutral phrases.

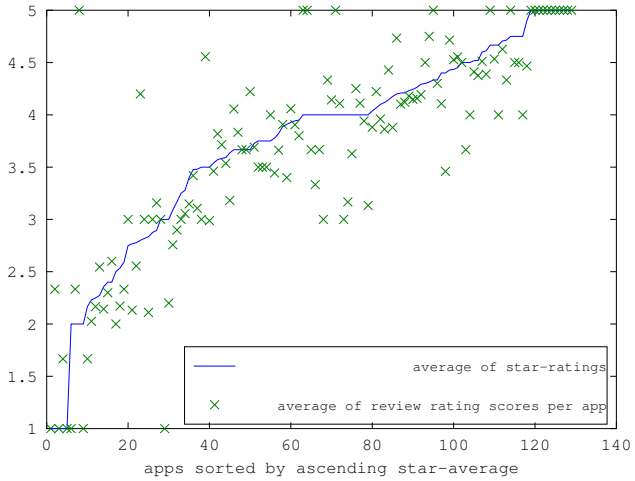


Fig. 8 Review rating scores per app.

VII. Conclusions

Textually-derived rating prediction can be performed well even when only phrase-level sentiment polarity is available. This is what the computational models introduced and evaluated in this paper have shown. Not all fine-granular opinions are of importance, however: filtering out subjective phrases with neutral sentiment and computing the overall sentiment of a review using the review rating score proposed in [16] and [17] provides the best star rating predictions for mobile apps' reviews written in German. Based on these results, new applications could suggest to customers how to

rate apps more consistently with the reviews they write, by considering their expressed opinions at the phrase level. Both customers and companies would benefit alike.

Further work will deal with the ideas that follow. Subjective phrases are aspect-oriented, i.e., the expressed opinions are probably related to features or aspects of a particular app. By extending the model to consider the aspects' relevance, an improvement in performance might be achieved. Furthermore, the phrase polarity is usually given in broad categories (i.e. positive, neutral, and negative). It could be interesting to analyse the *strengths* of the opinions [26], too. Moreover, it is our interest dealing with other types of models different than linear, multivariate regression ones.

ACKNOWLEDGEMENTS

We would like to thank Mario Sänger for insightful discussions and for providing us with the annotated customer mobile apps' reviews dataset.

REFERENCES

[1] A. Walz and R. Ganguly, *Apptentive 2015 Consumer Survey: The Mobile Marketer's Guide to App Store Ratings & Reviews*. Apptentive, 2015.

[2] Applause, "Listen to the Voice of Your Customers," n.d., available online at http://www.applause.com/resources#whitepapers, retrieved March 13, 2016.

[3] M. Galligan, "The right way to ask users to review your app," 2014, available online at https://medium.com/circa/the-right-way-to-ask-users-to-review-your-app-9a32fd604fca#.kud43shhq, retrieved March 13, 2016.

[4] A. Walz, "Dissecting the App Store Top Charts: The Anatomy of a Top App," 2015, available online at http://www.apptentive.com/blog/app-store-top-charts/, retrieved March 29, 2016.

[5] M. Smith, "Feedback and Loyalty on the Mobile Frontier: New Research From Apptentive and SurveyMonkey," 2016, available online at http://www.apptentive.com/blog/feedback-and-loyalty-on-the-mobile-frontier/, retrieved March 30, 2016.

[6] B. Pang and L. Lee, "Opinion Mining and Sentiment Analysis," *Foundations and Trends in Information Retrieval*, vol. 2, no. 1–2, pp. 1–135, 2008.

[7] B. Liu, *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers, 2012.

[8] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? Sentiment Classification using Machine Learning Techniques," in *Proceedings of the 43^{rd} Conference on Empirical Methods in Natural Language Processing, EMNLP'02*. Stroudsburg, PA, USA: Association for Computational Linguistics, 2002, pp. 79–86.

[9] B. Liu, "Sentiment Analysis and Subjectivity," in *Handbook of Natural Language Processing*, 2nd ed., N. Indurkhya and F. J. Damerau, Eds. Boca Raton, FL, USA: CRC Press, Taylor and Francis Group, 2010.

[10] B. Pang and L. Lee, "Seeing Stars: Exploiting Class Relationships for Sentiment Categorization with respect to Rating Scales," in *Proceedings of the 43^{rd} Annual Meeting of the Association for Computational Linguistics, ACL'05*. Stroudsburg, PA, USA: Association for Computational Linguistics, 2005, pp. 115–124.

[11] A. B. Goldberg and X. Zhu, "Seeing stars when there aren't many stars: Graph-based semi-supervised learning for sentiment categorization," in *Proceedings of the 1^{st} Workshop on Graph Based Methods for Natural Language Processing, TextGraphs-1'06*.

[12] D. Tang, B. Qin, T. Liu, and Y. Yang, "User Modeling with Neural Network for Review Rating Prediction," in *Proceedings of the 24^{th} International Joint Conference on Artificial Intelligence, IJCAI'15*, Q. Yang and M. Wooldridge, Eds. Palo Alto, CA, USA: AAAI Press, 2015, pp. 1340–1346.

[13] F. Li, N. Liu, H. Jin, K. Zhao, Q. Yang, and X. Zhu, "Incorporating Reviewer and Product Information for Review Rating Prediction," in *Proceedings of the 22^{nd} International Joint Conference on Artificial Intelligence, IJCAI'11*, T. Walsh, Ed., vol. 3. Menlo Park, CA, USA: AAAI Press, 2011, pp. 1820–1825.

[14] L. Qu, G. Ifrim, and G. Weikum, "The Bag-of-Opinions Method for Review Rating Prediction from Sparse Text Patterns," in *Proceedings of the 23^{rd} International Conference on Computational Linguistics, Coling'10*, C.-R. Huang and D. Jurafsky, Eds., vol. 2. Beijing, China: Tsinghua University Press, 2010, pp. 913–921.

[15] Y. Zhang, M. Zhang, Y. Liu, and S. Ma, "Boost Phrase-level Polarity Labelling with Review-level Sentiment Classification," *Computational Linguistics*, vol. 1, no. 1, pp. 1–25, 2006.

[16] G. Ganu, N. Elhadad, and A. Marian, "Beyond the Stars: Improving Rating Predictions Using Review Text Content," in *Proceedings of the 12^{th} International Workshop on the Web and Databases, WebDB'09*, 2009, pp. 1–6.

[17] G. Ganu, Y. Kakodkar, and A. Marian, "Improving the Quality of Predictions using Textual Information in Online User Reviews," *Information Systems*, vol. 38, no. 1, pp. 1–15, March 2013.

[18] N. Gupta, G. Di Fabbrizio, and P. Haffner, "Capturing the Stars: Predicting Ratings for Service and Product Reviews," in *Proceedings of the NAACL HLT 2010 Workshop on Semantic Search, SS'10*. Stroudsburg, PA, USA: Association for Computational Linguistics, 2010, pp. 36–43.

[19] S. O. Orimaye, S. M. Alhashmi, E. Siew, and S. J. Kang, "Review-Level Sentiment Classification with Sentence-Level Polarity Correction," *Computer Science, OALib Journal*, pp. 1–15, 2015.

[20] M. Sänger, "Aspektbasierte Meinungsanalyse von Bewertungen mobiler Applikationen," Master Thesis, Computer Science Dept., Humboldt-Universität zu Berlin, Berlin, Germany, December 2015.

[21] R. Klinger and P. Cimiano, "Bi-directional Inter-dependencies of Subjective Expressions and Targets and their Value for a Joint Model," in *Proceedings of the 51^{st} Annual Meeting of the Association for Computational Linguistics, ACL'13*, R. Navigli, J.-S. Chang, and S. Faralli, Eds., vol. 2. Sofia, Bulgaria: Association for Computational Linguistics, August 2013, pp. 848–854.

[22] ——, "Joint and Pipeline Probabilistic Models for Fine-grained Sentiment Analysis: Extracting Aspects, Subjective Phrases and their Relations," in *Proceedings of the IEEE 13^{th} International Conference on Data Mining Workshops, ICDMW'13*, W. Ding, T. Washio, H. Xiong, G. Karypis, B. Thuraisingham, D. Cook, and X. Wu, Eds. Dallas, TX, USA: IEEE Computer Society, December 2013, pp. 937–944.

[23] M. Sänger, U. Leser, S. Kemmerer, P. Adolphs, and R. Klinger, "SCARE – The Sentiment Corpus of App Reviews with Fine-grained Annotations in German," in *Proceedings of the 10^{th} International Conference on Language Resources and Evaluation, LREC'16*, N. Calzolari, K. Choukri, T. Declerck, M. Grobelnik, B. Maegaard, J. Mariani, A. Moreno, J. Odijk, and S. Piperidis, Eds. Paris, France: European Language Resources Association (ELRA), May 2016.

[24] R. Klinger and P. Cimiano, "The USAGE review corpus for fine grained multi lingual opinion analysis," in *Proceedings of the 9^{th} International Conference on Language Resources and Evaluation, LREC'14*, N. Calzolari, K. Choukri, T. Declerck, H. Loftsson, B. Maegaard, J. Mariani, A. Moreno, J. Odijk, and S. Piperidis, Eds. Reykjavik, Iceland: European Language Resources Association, May 2014, pp. 2211–2218.

[25] M. Sänger, private communication, 2016.

[26] T. Wilson, J. Wiebe, and R. Hwa, "Just how mad are you? Finding strong and weak opinion clauses," in *Proceedings of the 19^{th} National Conference on Artificial Intelligence, AAAI'04*, G. Ferguson and D. McGuinness, Eds. Menlo Park, California: AAAI Press, 2004, pp. 761–767.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 429–438
DOI: 10.15439/2016F224
ACSIS, Vol. 8. ISSN 2300-5963

Information Management for Travelers: Towards Better Route and Leisure Suggestion

Evgeny Pyshkin
Software Engineering Lab.
University of Aizu
Aizu-Wakamatsu, 965-8580, Japan
Email: pyshkin@icc.spbstu.ru

Alexander Baratynskiy, Alexander Chisler, Boris Skripal
Institute of Computer Science and Technology
Peter the Great St. Petersburg Polytechnic University
St. Petersburg, 195251, Russia
Emails: {baratynskiy, alexander.chisler, skipalboris}@gmail.com

***Abstract*—Contemporary travel information services are connected to huge amount of travel related data used for improving personalized suggestions. Such suggestions include finding better routes, access to amusement and educational amenities implemented as digital services, as well as the features for people collaboration, and for planning leisure time with respect to existing attractiveness evaluation algorithms under time-budget constraints. Much effort is required for supporting personalized itineraries construction in such a way which would leverage existing cultural and technological user experience. In this paper we analyze the underlying algorithms and major components being an implementation of the proposed model investigated with particular attention to annotated leisure walk route construction, traveler collaboration and travel meeting management. In sum, we make an effort to address a number of complex issues in the area of developing models, interfaces and algorithms required by modern travel services considered as an essential application of a human-centric computing multidisciplinary paradigm.**

I. Introduction

There is no exaggeration in saying that traveling is one of remarkable ways to discover the world. Among different obvious factors (which include journey time planning, accommodation management, equipment preparation, etc.) there are many aspects related to the careful forethought of a journey which would make it coming up well to the traveler's interests and expectations. To a significant extent, present day information services for travelers are about personalization, extending usage scenarios, improving suggestions and recommending journey options on the base of the information retrieved from various warehouses containing huge amount of travel related data.

The idea of our approach is to combine user experience orientation and rich facilities of the present day information retrieval, processing and presentation tools for the benefits of better travel planning, itinerary construction and traveler experience sharing. Indeed, with current level of computer-assisted tools and methods tourists expect more facilities than simply finding a fastest and cheapest way or getting a direction. They expect to be advised how to deal with many competitive factors of route construction, some of them (like sight attractiveness) aren't easily formalizable. The problems of developing better personalized services for travelers are within the scope of the emerging domain of urban computing [1], [2] where suggestion environments are rapidly developing. Thus, leisure could be suggested: we believe that advancing services for travelers is a perfect area to be upgraded by introducing better suggestion facilities.

In this work we don't pretend to cover all possible suggestion use cases supported by a big variety of existing applications. We partially address only a selection of scenarios representing five major aspects of travel planning and management:

1) Journey preparation and planning;
2) Itinerary construction automation and navigation;
3) Multimedia assistance automation;
4) Traveler interaction and collaboration;
5) Post-travel experience.

The listed aspects (which are not exhaustive, nor totally independent) and the corresponding scenarios are described in Section II.

II. Aspects and Use Cases

Let us examine our selection of travel planning and management aspects.

The first aspect is **journey preparation and planning**. Nowadays travelers still use but aren't totally satisfied with traditional paperback editions of numerous travel guides containing plenty of maps and city plan fragments. Voyagers consult different guides implemented as software applications (currently available on many mobile platforms and advertised on the web sites like *Musement*[1], *Izi.travel*[2] or *PocketGuide*[3]), they share their experience via social networks and traveler forums available on relevant web sites (like *TripAdvisor*[4]). Many applications exist for mobile devices supporting GPS-sensors, electronic maps, as well as various multimedia features. In Section III-A we provide a brief survey of existing web resources for travelers.

Instead of taking predefined packages, many people prepare their own itineraries with the help of information available on the web sites for travelers. Thus, the second aspect is **route construction automation and navigation**. While planning

[1] www.musement.com/
[2] https://izi.travel/en
[3] http://pocketguideapp.com
[4] http://tripadvisor.com

978-83-60810-90-3/$25.00

visiting some area for a certain period of time, tourists are unlikely able to visit each attraction. Effectively, they solve a kind of fuzzy optimization problem in order to select something that would be specifically interesting to them. Tourists select points of interest (POIs) depending on their significance from a certain point of view. Hence, one of possible application of traveler advisory systems is to navigate the selection process by implementing the criteria representing a tourist object attractiveness by some formal schema. In Section III-B we analyze possible approaches for such a formalization.

Creating personalized itineraries requires using rich facilities of **multimedia assistance automation**. This aspect is connected to different issues, including assisting visually impaired pedestrians [3], creating multimedia travel books integrated with electronic maps and accessible from mobile devices [4], creating audio guides for museums and cultural sights [5], [6], designing personalized recommendation systems in order to mitigate information overload [7], and so on.

The next aspect is **traveler interaction and collaboration**. Somehow, many current efforts are about delivering personalized solutions allowing travelers to leverage and to share their experience and/or to follow major scenarios we could learn as a result of travelers' experience study [8], [9].

Personalization and user collaboration is connected strongly to the possibility to leverage **post-travel experience** knowledge. Among existing solutions for travel reporting we can cite a couple of examples like *TripJournal*[5] and *TripCast*[6].

To sum up, we agree with the statement from [10] that the focus of many existing solutions is on creating a technology which wouldn't support only a kind of time-budget optimization problem but would allow travelers to develop their own *memorable experience* (the latter term being borrowed from [11]).

III. State-of-the-Art Approaches

This section contains a brief survey of state-of-the art approaches developed and used in the domain of traveling related information systems.

A. General Information Services for Travelers

Apart from research projects (examined in details in further sections), there are two big classes of present day electronic services for tourists: thematic web sites (which are sometimes linked to special software applications) and a huge variety of mobile applications. Among popular web based solutions there are the following major groups:

- Web sites for hotel, flight, restaurant searching and booking: they are kinds of aggregation bridges to many external services; sometimes they contain traveler forums and features for collecting and displaying user feedback and user impressions. Good examples are *TripAdvisor* and *Expedia*[7].
- Travel guides implemented as a collection of stories and suggestions which are sometimes focused not only on sites and attractions but also on interesting events and on sharing personal experience: particular implementations may contain event calendars, transport and place orientation, integration with electronic maps. Good examples are *Timeout*[8], *I.Know*[9] and *Japan Guide*[10].
- Individual trip planning or collaborative planning: journey agenda and route planning (like in *Travefy*[11] and *Tripomatic*[12]), transportation planning (like in *RouteRank*[13] and *Hyperdia*[14]).
- Journey and transport trackers, including flight trackers (*FlightTrack*[15], *GateGuru*[16]).
- Personalized itineraries construction examined in the subsequent sections of this paper.
- Platforms for sharing user in-travel and post-travel experience: good example is (*TravelDiaries*[17]). Travel diaries are often used by other travelers in time of preparation their future trips.
- Multimedia guides: good examples are audio guides used not only in traditional indoor museum environments but also for outdoor journeys (*Azbo*[18], *Izi.Travel*[19], *PocketGuide*[20]).
- Unusual models: good example is *Explorra*[21], an approach where a collection of attractions is accessed after a user selects a color.

Of course, the above mentioned groups exist rarely in its pure form, many features being shared among different solutions. Currently the focus is being shifted to lightweight (mobile) applications targeting the everyday life stories.

B. Itinerary Construction: Interfaces and Implementations

Among existing solutions for travel itinerary construction we have to mention several projects which are within the scope of our research.

The mobile application TAIS for guiding tourist activity described in [12] is focused on step-by-step itinerary construction in response to user actions and movements. There is an interesting feature of collecting user impressions about the visited POIs. The application generates recommendations collected on the base of other travelers' experience and evaluations. For each POI its detailed information includes a list of images associated with the attraction and its description. As

[5] www.trip-journal.com/
[6] https://tripcast.co/
[7] http://expedia.com
[8] http://www.timeout.com/
[9] http://iknow.travel/
[10] http://www.japan-guide.com
[11] http://www.travefy.com
[12] http://www.tripomatic.com
[13] http://www.routerank.com/en/
[14] http://www.hyperdia.com/
[15] https://www.mobiata.com/apps/flighttrack
[16] www.gateguru.com/
[17] www.traveldiariesapp.com/
[18] https://azboguide.com/en
[19] https://izi.travel/en
[20] http://pocketguideapp.com
[21] https://www.explorra.com/labs/travel-by-color

a routing service the authors use *OpenStreetMap* API, while *Yandex.Schedule* API is used for searching available public transportation routes, the latter being a very promising feature.

A tour planning system *Aurigo* combines a recommendation algorithm with interactive visualization for creating and managing personalized itineraries [10]. To a great extent, this project is in the same direction as the project of ours: the authors investigate a possible balance between an automated and purely manual approaches.

TripBuilder framework described in [13] is an implementation of an approach for planning personalized sightseeing tours in cities. The itineraries are being constructed after analysis of the geo-tagged photos collected from *Flickr*[22] and associated with the POIs collected from Wikipedia. The photos are considered as traces revealing the behaviors of tourists and as a source of spatio-temporal information about their sightseeing experience. The itinerary construction is modeled as an instance of the generalized maximum coverage problem with respect to visiting time-budget optimization and to further building of the itinerary as an instance of the traveling salesman problem. The approach was advanced in [14] by applying an algorithm for suggesting contextually relevant POIs on the base of user preferences and interests.

In our previous works (see [15], [16]) we described an approach to design an application which is not limited to obtaining the information about different attractions, it provides the ability of planning and constructing a travel itinerary in advance. That's why tourist guides may use it in their work. We expect that, by using the application, professional experts and amateur guides can prepare their tours both in automatic and manual modes.

In contrast to [17] we couldn't totally agree that "creating an efficient and economic trip plan is the most *annoying* (our emphasis) job for a backpack traveler". We believe that creating a customized itinerary is a very creative stage which is part automatic, part manual: for producing interesting high quality suggestions the contribution of *human* experts is extremely important.

Our application focuses a concept of an annotated travel itinerary. which is not a simple path (for instance, the shortest one) between two points on the map, but a route description which includes a set of POIs representing such attractions as architectural sights, museums, historical places, monuments, view points, etc. In order to create a relevant annotation, one should consider including such entities as texts, images (photos, drawing, replicas, diagrams, etc.), multimedia objects (audio or video clips), web links, notes, citations, dates, timeline connections, information about related people, places or events, to cite a few.

Figure 1 gives an idea of existing complexity of a domain specific ontology by an example of architectural points. Figure 2 follows a pattern proposed in [18] and shows major usage scenarios. The common-sense ontology provides a foundation for the information system architecture.

[22]http://www.flickr.com

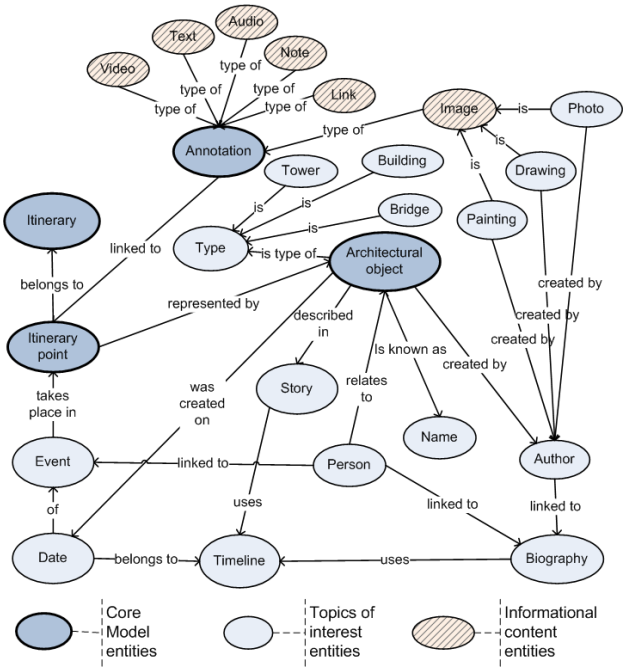


Fig. 1. An example of domain specific ontology fragment

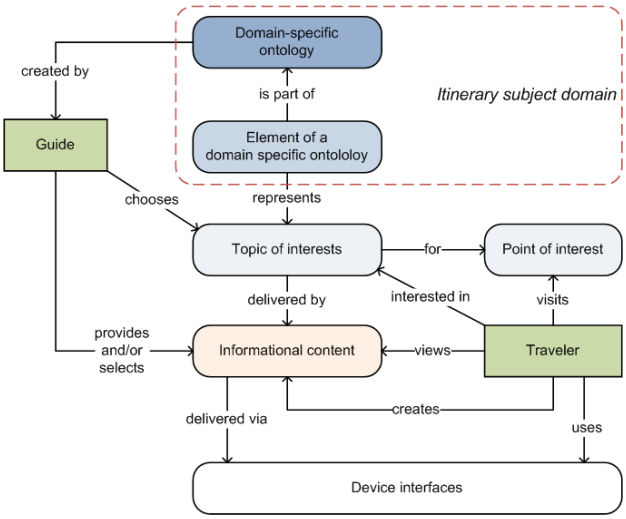


Fig. 2. Common sense ontology

C. Travel Route Automation: Formal Models and Algorithms

Itinerary planning automation is still an important aspect even for local itineraries for leisure walks. An obvious formal model representing the problem of generating an itinerary is finding an optimal way on an undirected or directed weighted graph. However the question is not about finding a shortest path. There are many competing factors that have to be taken into consideration:

- POI opening and closing times (time schedule);
- POI visual characteristics (viewpoints);

- POI purposes and their relation to the journey focus (important for thematically organized itineraries);
- POI importance (in a sense);
- POI types (historical objects, relaxation objects, sanitary facilities, transportation hubs, etc.);
- Dependency on weather, traffic, safety conditions, and similar issues.

Hence, common shortest path algorithms aren't enough. Hereafter we examine possible approaches that could be useful for automatic itinerary construction.

1) Using Genetic Algorithms: An adaptive genetic algorithm described in [19] is based on partitioning the work space by several blocks with unique numbers as Figure 3 shows.

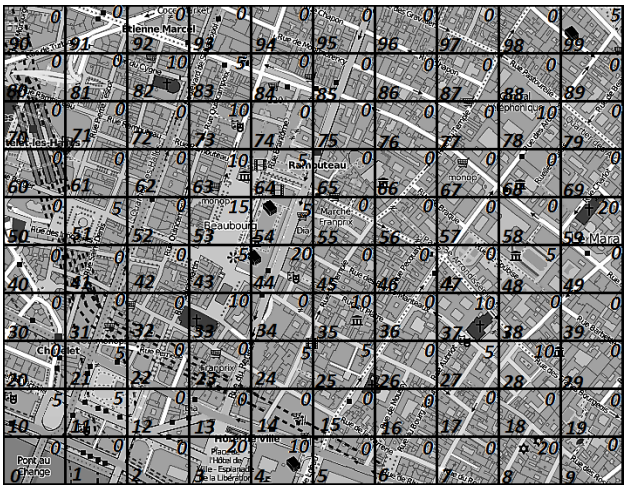


Fig. 3. Division the workspace into blocks

A path is a sequence of "visited" blocks, while a fitness function defined as an adopted version from [19] is as follows:

$$F = (1 + \frac{1}{\sqrt{n-1}}) \cdot D \cdot C \tag{1}$$

In equation (1) n is the number of free cells, D is the sum of linear distances between adjacent individuals, C is is the sum of object scores (shown on the upper right corners of each block in Figure 3) within the selected blocks.

This method allows planning routes quickly: on a grid of 10×10 cells with scale of population equal to 40 and length of each population equal to 20, a stable population is created after 40-th generation. However the fitting function ignores some important factors, for example, it ignores the walk duration.

2) Random Search Algorithms: In [20] the authors used a greedy randomized adaptive search procedure which could be explained as follows. The first step is to find some start solution (for example, we can select several POIs with the highest popularity and a shortest path between them). The second step is to calculate the value of F function that describes the effect of adding a POI to the route:

$$F = \frac{T_{beforeAdding}}{T_{afterAdding}} \cdot S \tag{2}$$

In equation (2) $T_{beforeAdding}$ is a route time before inclusion of current POI, $T_{afterAdding}$ is a time after inclusion of the POI into the best position (i.e. the position corresponding to the time which is as minimal as possible), while S is a score of current object.

The next steps are iterative: the idea is to find several points with the best F value which are not included to the major route, and to select a random point from them, so as to add it into the route. This process is being repeated until we decide to stop. There could be different stop situations: for example, we don't have more time to continue searching, or we don't have any POIs more. Figure 4 gives an illustration for one iteration.

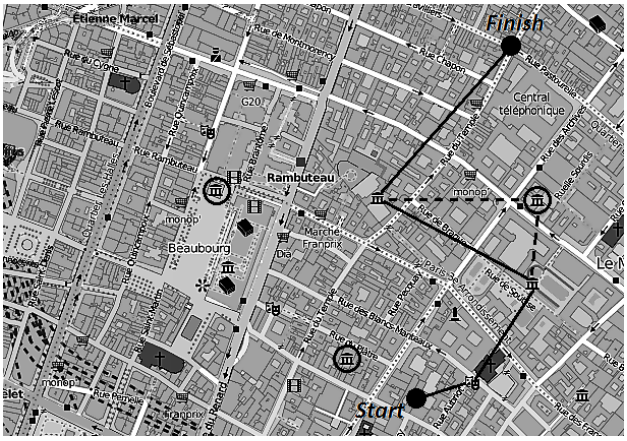


Fig. 4. One iteration of the route generation using GRASP

3) Ant Colony Optimization: The promising way is to use ant colony optimization (ACO) known since 1991 [21] and originally developed for solving a problem of finding an optimal path on a graph. A traditional ant colony algorithm is based on modeling the behavior of the ants dropping pheromones while deciding where to go depending on the intensiveness of pheromone trails. If an independent ant moves rather randomly, an ant discovering a previously laid pheromone trail follows this trail with higher probability. As a "side" effect, an ant following such previously laid trail reinforces the amount of pheromones on its way. There are ACO applications for many areas; in [22] there is a description of ACO application for a traveler salesman problem which is often a part of travel planning algorithms. An interesting ACO modification for tourist route planning was described in [23] as a formal foundation for a food tour recommendation system.

The solution can be enhanced by taking into consideration such important leisure walk properties as walk time, time required to visit a POI, nutrition conditions and other characteristics.

IV. A Prototype Software System

In this section we describe the software components developed to demonstrate the key ideas of our approach, namely: making suggestions, interface simplification, implicit actions preferred, a focus on the above mentioned selected aspects.

A. Component for Interactive Itinerary Construction Integrated with OpenStreetMap

According to an *annotated travel itinerary* concept presented in Section III-B, we developed an interface allowing to deal with a diversity of data to be visualized along the editable itinerary. Figure 5 shows a fragment of a POI description in terms of domain-specific ontology presented in Figure 1.

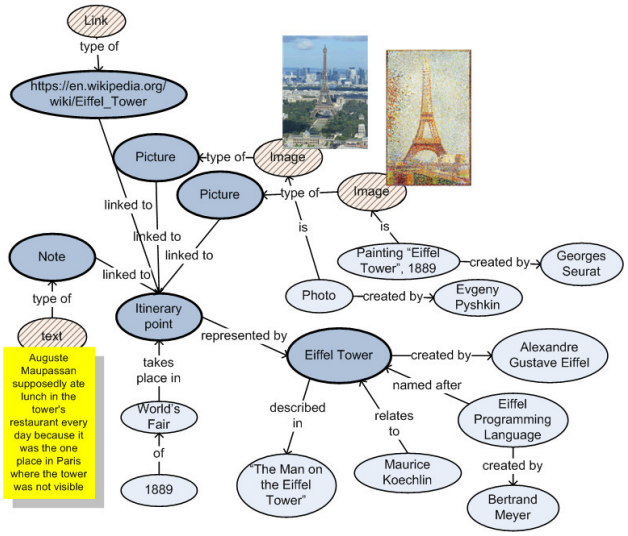


Fig. 5. A POI described in terms of domain-specific ontology (fragment)

Figure 6 shows a main panel of our prototype application for annotated itinerary construction together with some examples of POI annotations. As an electronic map service we use *OpenStreetMap* (OSM)[23] which is a collaborative project for free editable world maps creation. In our implementation a *JMapViewer* component is used.

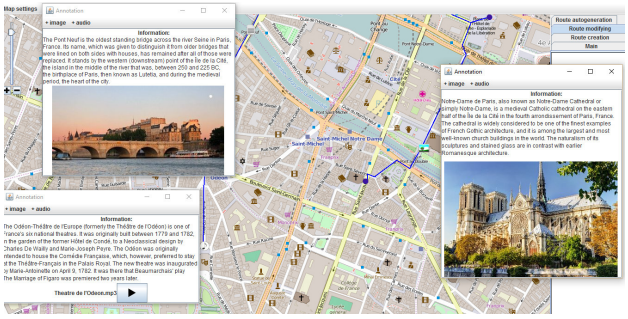


Fig. 6. Main frame window with an example of constructed itinerary

Users are able to select a search area on the map in order to allow suggesting the POIs for further consideration for including them to the itinerary. Figure 7 gives an idea of the adviser component user interface.

In sum, currently the system supports the following features:

[23]http://www.openstreetmap.org

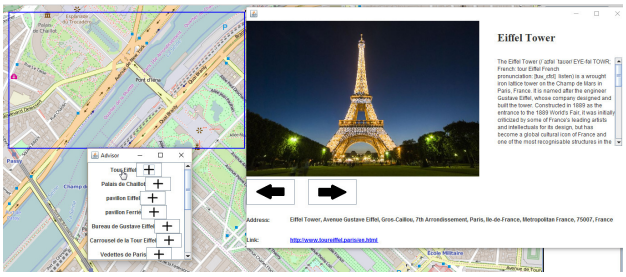


Fig. 7. Interactive adviser: attractions within the selected area

- Ability to create the route itself and to add the annotations containing text information, images, audio, links to web sites, etc.;
- Ability to modify the created tourist route;
- Ability to search POI-related information on the web;
- Automatic tourist route construction.

B. Automatic Itinerary Construction

For the leisure walk itinerary automation we use the following simplified scenario:

1) Define a time slot, select the departure and destination points. Define other constrains if required.
2) Explore potential locally accessible POIs for the current point on the route (initially the current point is set to the departure point).
3) Evaluate the POIs by using some formalized model for taking into account the degree of POI popularity, attractiveness, price to visit, location, etc.
4) Rank the POI and add the best ranked POIs to the part of the route in progress.

Steps 2–4 are repeated until the destination point is reached.

The standard task of tourist route generation is formalized by Souffriau [24] as follows:

Assume there are N POIs. For every POI: $x_{ij} = 1$, if there is a path between the POIs i and j, otherwise $x_{ij} = 0$. S_i is a i-th POI's score, i ranging from 1 (departure) to N (destination). Time t_{ij} corresponds to the shortest path from point i to point j. The total score S_{total} has to be maximized within the limit T_{max} time.

This model sets boundaries of the future route and determines criteria for the best tourist route which is the best selection of the tourist objects and the best path connecting these objects.

Due to the fact that the POI score depends on its place along the route, each potentially accessible POI's score has to be recalculated at each iteration. In order to explore POIs locally we use the geometric model shown in Figure 8.

This geometric model has the following parameters: *Start*, *Finish* – the arrival and departure points within the route area, a – the semi-major axis of the search area, b – the semi-minor axis of the search area, c – half of the focal distance of the search area, $S_{max}/2$ – half of the maximum distance that can be covered for the remaining time, α – search area angle of rotation, γ – semi-minor axis minimization coefficient.

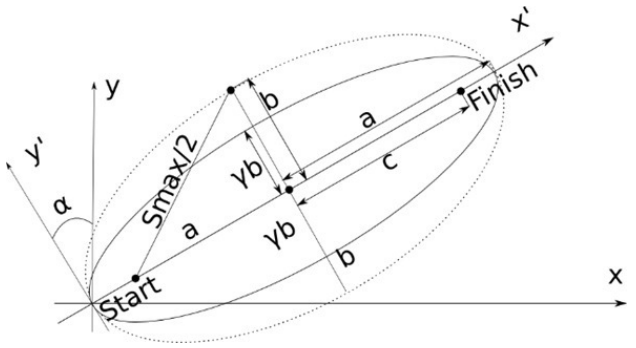


Fig. 8. Exploring local area

Unlike to the model used in the earlier mentioned *Aurigo* project, where the authors introduced a *Pop Radius* for exploring the POIs in the *circled* local area [10] (similar idea is also used in *ATIPS* project [25]), we use an elliptic model for the local POI exploration area. The elliptic form of the local region, where the potential POIs to visit are explored, allows us to use ellipse particularity: the sum of the distances to the two focal points is constant for every point on the curve. It means that if this sum is equal to the distance to the most distant point, a person can theoretically reach it within the given time and speed limits; each point outside the ellipse couldn't be reached. So, too distant POIs must be excluded from the short list used for further analysis. The POIs located close to the border of the ellipse may also be excluded from the search, otherwise it may happen that the resulting route is too sparse: a traveller will spend much time to walk in-between, instead of visiting the selected attractions.

In current implementation, we used an extension of a gradient descent algorithm for searching POIs to be included to the route. The extension is as follows: searching a new POI to be included occurs in the area with a maximum distance between two objects in the buffer route. This modification makes possible to reduce the maximum distance between the tourist objects and to exclude from the consideration hardly reachable objects. Then the selected objects are evaluated in order to find the object with the best score.

C. Implementing A Simplified Scenario for Meetings Halfway

Let us examine one interesting use case for suggestion oriented routing services which is a service targeting not the only user but the joint interests of several independent users.

Even in the case of a simplified scenario of arranging a meeting of two travelers, the task is not trivial: the involved persons have to consider air flight prices, flight schedule, tickets availability, accommodation and living costs. Huge arrays of data are potentially available via existing Internet services, but it is almost impossible to process them manually: an air ticket price and availability might change very quickly and dramatically in time, currency exchange rates are also not stable, there is a huge amount of data on flight connections, to cite a few.

We follow an idea similar to Google maps based *Meetways* application [26] developed since about 2008 [27]. One more example is *Geo Meetpoint* [28]. Probably the closest implementation is provided by *Meet Me in the Middle* application [29]. The latter has an easy-to-use interface and lightweight design fitting well searching for short distance routes, but there is limited support for cross-border points.

Informally, our model scenario could be described as a story of two friends trying to find a suitable meeting point:

Alex lives in St. Petersburg, Russia, while his girlfriend Tina lives in London, UK. They would like to meet each other. However, there are some constrains: they are students (so, they have rather limited funding), currently Alex has no UK visa (and it takes much time to get it), however they are open for any idea about the suitable place.

Saying more formally, assume $City_1$ is a city where Alex flies from, and $City_2$ is one where Tina flies from, while $City_3$ is a possible halfway meeting place. A set R_1 is a set of possible flights departing from the $City_1$ airports, while a set R_2 is a set of possible flight departing from $City_2$. Then an intersection X of R_1 and R_2 provides the selection of flights which correspond to the possible meeting places. In order to find the best choice we have to download the detailed flight information (prices and available dates) for all the flights departing from the $City_1$ and $City_2$ airports. Then we consider the combinations of all the destinations from the set X corresponding to the flights departing from $City_1$ or $City_2$. Such combinations give us an array of one-way flights. Then we add return flights in order to consider the full trip cost. Finally, the selection is sorted so as to discover the best options. Figure 9 illustrates the issue.

We organized the prototype implementation by introducing three components: web data access, data matching analysis and end user interface. First component is responsible for downloading data from available web sites providing access to flight data. Currently we use an *aviasales.ru* [30] open API allowing us to get flight information with dates, prices and possible directions from a selected city. Web data is parsed by a Python component. Data are extracted by using *Requests* library [31]; the response is received in JSON format [32]. Here is an example of such a response:


```
"success":true,
"data":[{
 "show_to_affiliates":true,
  "trip_class":0,
  "origin":"LED",
  "destination":"HKT",
  "depart_date":"2015-10-01",
  "return_date":"",
  "number_of_changes":1,
  "value":29127,
  "found_at":"2015-09-24T00:06:12+04:00",
  "distance":8015,
  "actual":true
}]
```

Figure 10 shows the user interface supporting four options: first departure city selection, second departure city selection, trip duration (in days) and desired departure date. Web-application is implemented by using *Flask* [33].

Fig. 9. A problem of finding a meeting point for two travelers

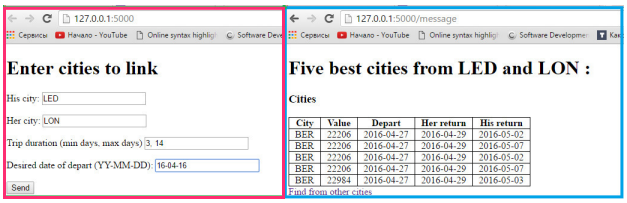


Fig. 10. A selection of flights which fit the user input

As a main result, currently the application yields a suggestion. Before any decision could be made, the application requires much more information than any user is able to handle. Thus, the implementation illustrates a concept of managed suggestion: we analyze plenty of flight related information and suggest the most suitable option. The solution can be extended so as to support many different use cases: meeting in a city, going to a cinema, following a music band in a tour, finding a place for a conference, arranging a business meeting, and many others.

V. Conclusion

In this paper we made an effort to touch a selection of aspects of today's information systems for travelers where there is a convergence of approaches being developed within the framework of human-centric, aware and urban computing. We designed the software components targeting better travel personalization and user collaboration with special emphasis on suggesting better itineraries both in local and global scope. Specifically, the following ongoing projects were presented:

- A visual *OpenStreeMap* based tool for constructing annotated travel itineraries: the tool is useful both for independent travelers and for guides preparing the journey for their clients;
- A tool for automatic of travel itinerary construction integrated with the above mentioned visual guide assistant;
- A *Meeting Point* application suggesting a suitable halfway meeting point for two travelers using air flights for reaching the meeting destination.

We believe that currently the process leading us towards better route and leisure suggestions is being rethought with more attention to the following aspects:

- More orientation on planning by collaboration;
- Deeper integration with external services including electronic maps, navigation systems, knowledge resources, libraries, etc.;
- Resolving usability issues and developing better intelligent user interfaces and HCI solutions for leveraging user experience and taking into account existing features of present day computers and portable smart devices;
- Planning experiments conducted by real users.

With regard to the content related improvements, we believe that automatic tourist route generation algorithms have to be extended in order to support an extremely useful feature like constructing the thematically linked and multi-day journeys. We think about deeper integration with multimedia features used during the suggested sightseeing walks. We also consider investigating the issue of using old maps available as images. A perspective to the ancient views accessible electronically will significantly extend the way to learn history while visiting tourist attractions all around the world.

Appendix

In research works it is rarely shown how information about existing relevant domain-specific solutions are being retrieved. Table I and Table II give a hint how the searching process was organized in our case. We included information about search queries, extended queries and a selection of Google results obtained while using Google Chrome in private view mode so as not to keep any search history information. Results mentioned as "Direct" mean that they are obtained directly as Google web search output. Results discovered after subsequent examination of "direct" resources are mentioned as "Indirect". We also noted the results relevance to the scope of current contribution. Some of explored models, approaches and algorithms provide good foundation for further investigations within the framework of our research.

Acknowledgment

The authors thank Matvei Pyshkin for his valuable contribution to the detailed survey of the solutions, products and services examined in our work.

References

[1] M. Andreea *et al.*, "The emerging technological trends in the tourism industry," *Annals-Economy Series*, pp. 73–76, 2014.
[2] Y. Zheng, L. Capra, O. Wolfson, and H. Yang, "Urban computing: concepts, methodologies, and applications," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 5, no. 3, p. 38, 2014.
[3] R. F. Joseph and A. A. Godbole, "An intelligent traveling companion for visually impaired pedestrian," in *Circuits, Systems, Communication and Information Technology Applications (CSCITA), 2014 International Conference on*, April 2014, pp. 283–288.

[4] V. Mladenovic, M. Lutovac, and M. Lutovac, "Electronic tour guide for android mobile platform with multimedia travel book," in *Telecommunications Forum (TELFOR), 2012 20th*, Nov 2012, pp. 1460–1463.

[5] R. Karimi, A. Nanopoulos, and L. Schmidt-Thieme, "Rfid-enhanced museum for interactive experience," in *Multimedia for cultural heritage*. Springer, 2012, pp. 192–205.

[6] J.-P. Gerval and Y. Le Ru, "Fusion of multimedia and mobile technology in audioguides for museums and exhibitions: from bluetooth push to web pull," 2011.

[7] Y.-M. Huang, C.-H. Liu, C.-Y. Lee, Y.-M. Huang *et al.*, "Designing a personalized guide recommendation system to mitigate information overload in museum learning."

[8] T. Y. Lim, "Designing the next generation of mobile tourism application based on situation awareness," in *Network of Ergonomics Societies Conference (SEANES), 2012 Southeast Asian*. IEEE, 2012, pp. 1–7.

[9] T.-D. Cao and N.-D. Tuan, "Improving travel information access with semantic search application on mobile environment," in *Proceedings of the 9th International Conference on Advances in Mobile Computing and Multimedia*, ser. MoMM '11. New York, NY, USA: ACM, 2011, pp. 95–102. [Online]. Available: http://doi.acm.org/10.1145/2095697.2095716

[10] A. Yahi, A. Chassang, L. Raynaud, H. Duthil, and D. H. P. Chau, "Aurigo: an interactive tour planner for personalized itineraries," in *Proceedings of the 20th International Conference on Intelligent User Interfaces*. ACM, 2015, pp. 275–285.

[11] V. W. S. Tung and J. B. Ritchie, "Exploring the essence of memorable tourism experiences," *Annals of Tourism Research*, vol. 38, no. 4, pp. 1367–1386, 2011.

[12] A. Smirnov, A. Kashevnik, N. Shilov, N. Teslya, and A. Shabaev, "Mobile application for guiding tourist activities: tourist assistant-tais," in *Open Innovations Association (FRUCT16), 2014 16th Conference of*. IEEE, 2014, pp. 95–100.

[13] I. Brilhante, J. A. Macedo, F. M. Nardini, R. Perego, and C. Renso, *Advances in Information Retrieval: 36th European Conference on IR Research, ECIR 2014, Amsterdam, The Netherlands, April 13-16, 2014. Proceedings*. Cham: Springer International Publishing, 2014, ch. TripBuilder: A Tool for Recommending Sightseeing Tours, pp. 771–774. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-06028-6_93

[14] ——, "Scaling up the mining of semantically-enriched trajectories: Tripbuilder at the world level," 2015, accessed on April 23, 2016. [Online]. Available: http://ceur-ws.org/Vol-1404/paper_12.pdf

[15] B. Skripal and E. Pyshkin, "Automated leisure walk route generation for an interactive travel planner," in *Proceedings of the International Workshop on Applications in Information Technology (IWAIT-2015)*, The University of Aizu. The University of Aizu Press, Oct 2015, pp. 29–32.

[16] A. Baratynskiy and E. Pyshkin, "Traveler guide assistant: Introducing an application for an openstreetmap based travel itinerary construction," in *Proceedings of the International Workshop on Applications in Information Technology (IWAIT-2015)*, The University of Aizu. The University of Aizu Press, Oct 2015, pp. 25–28.

[17] G. Chen, S. Wu, J. Zhou, and A. K. Tung, "Automatic itinerary planning for traveling services," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 26, no. 3, pp. 514–527, 2014.

[18] A. Garcia, O. Arbelaitz, M. T. Linaza, P. Vansteenwegen, and W. Souffriau, *Personalized tourist route generation*. Springer, 2010.

[19] S. Yu, J. Zhao, and C. Hu, "Route planning of stacker by improved genetic algorithm," *Automatic Control and Artificial Intelligence (ACAI 2012), International Conference on*, 2012.

[20] Z. Zabinsky, "Random search algorithms," *Department of Industrial and Systems Engineering, University of Washington, USA*, 2009.

[21] A. Colorni, M. Dorigo, V. Maniezzo *et al.*, "Distributed optimization by ant colonies," in *Proceedings of the first European conference on artificial life*, vol. 142. Paris, France, 1991, pp. 134–142.

[22] T. Stützle and M. Dorigo, "Aco algorithms for the traveling salesman problem," *Evolutionary Algorithms in Engineering and Computer Science*, pp. 163–183, 1999.

[23] K. Sriphaew and K. Sombatsricharoen, "Food tour recommendation using modified ant colony algorithm," *5th International Conference on Computing and Informatics, ICOCI 2015 11-13 August, 2015 Istanbul, Turkey*, 2015.

[24] W. Souffriau, *Automated Tourist Decision Support*. Katholieke Universiteit Leuven, 2010.

[25] H.-T. Chang, Y.-M. Chang, and M.-T. Tsai, "Atips: Automatic travel itinerary planning system for domestic areas," *Computational Intelligence and Neuroscience*, vol. 2016, 2015.

[26] "Meetways: Meet me in the middle," accessed: Dec 30, 2015. [Online]. Available: http://content.usatoday.com/communities/popcandy/post/2008/10/780903/1#.Voj79U_6LGA

[27] W. Matheson, "Why go the distance when you can go half?" Oct 2008, accessed: Mar 10, 2016. [Online]. Available: http://smallbusiness.chron.com/arrange-business-meeting-75187.html

[28] "Geo meetpoint," accessed: Mar 10, 2016. [Online]. Available: http://www.geomidpoint.com/meet/

[29] "Meet me in the middle," accessed: Mar 11, 2016. [Online]. Available: https://itunes.apple.com/us/app/meet-me-in-the-middle/id826982528?mt=8

[30] "Aviasales api," accessed: Dec 29, 2016. [Online]. Available: https://www.aviasales.ru/API

[31] "Requests: Http for humans," accessed: Mar 15, 2016. [Online]. Available: http://docs.python-requests.org/en/master/

[32] "18.2.json encoder and decoder," accessed: Mar 15, 2016. [Online]. Available: https://docs.python.org/2/library/json.html

[33] "Flask web development, one drop at a time," accessed: Apr 28, 2016. [Online]. Available: http://flask.pocoo.org/

[34] A. Rikitianskii, M. Harvey, and F. Crestani, *Advances in Information Retrieval: 36th European Conference on IR Research, ECIR 2014, Amsterdam, The Netherlands, April 13-16, 2014. Proceedings*. Cham: Springer International Publishing, 2014, ch. A Personalised Recommendation System for Context-Aware Suggestions, pp. 63–74. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-06028-6_6

[35] J. Duffy, "The best travel apps of 2015," *PC*, August 2015.

[36] R. Anacleto, L. Figueiredo, A. Almeida, and P. Novais, "Mobile application to provide personalized sightseeing tours," *Journal of Network and Computer Applications*, vol. 41, pp. 56 – 64, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1084804513002105

[37] M. De Choudhury, M. Feldman, S. Amer-Yahia, N. Golbandi, R. Lempel, and C. Yu, "Automatic construction of travel itineraries using social breadcrumbs," in *Proceedings of the 21st ACM conference on Hypertext and hypermedia*. ACM, 2010, pp. 35–44.

[38] R. Rajeswari and J. M. Mannan, "Efficient multiuser itinerary planning for travelling services using fkm-clustering algorithm," 2015.

[39] D. Batchelor, "Collaborative travel and tourism: the best way to predict the future is to invent it," January 2012, accessed: Apr 23, 2016. [Online]. Available: http://blogamadeus.com/18/01/collaborative-travel-the-best-way-to-predict-the-future-is-to-invent-it/

[40] A. Henri, "Five best travel planning apps," November 2013, accessed: Apr 23, 2016. [Online]. Available: http://lifehacker.com/five-best-travel-planning-apps-1470002139

[41] S. Dunstall, M. E. Horn, P. Kilby, M. Krishnamoorthy, B. Owens, D. Sier, and S. Thiebaux, "An automated itinerary planning system for holiday travel," *Information Technology & Tourism*, vol. 6, no. 3, pp. 195–210, 2003.

[42] S. B. Roy, G. Das, S. Amer-Yahia, and C. Yu, "Interactive itinerary planning," in *Data Engineering (ICDE), 2011 IEEE 27th International Conference on*. IEEE, 2011, pp. 15–26.

[43] A. Gionis, T. Lappas, K. Pelechrinis, and E. Terzi, "Customized tour recommendations in urban areas," in *Proceedings of the 7th ACM international conference on Web search and data mining*. ACM, 2014, pp. 313–322.

[44] X. Li, "Multi-day and multi-stay travel planning using geo-tagged photos," in *Proceedings of the Second ACM SIGSPATIAL International Workshop on Crowdsourced and Volunteered Geographic Information*. ACM, 2013, pp. 1–8.

[45] A. Majid, L. Chen, H. T. Mirza, I. Hussain, and G. Chen, "A system for mining interesting tourist locations and travel sequences from public geo-tagged photos," *Data & Knowledge Engineering*, vol. 95, pp. 66–86, 2015.

[46] L.-Y. Wei, Y. Zheng, and W.-C. Peng, "Constructing popular routes from uncertain trajectories," in *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2012, pp. 195–203.

TABLE I
QUERIES AND RESULTS (PART 1)

Query	*Finding*			
	Link	*Description*	*Direct or Indirect*	*Relevance*
sightseeing tools	http://link.springer.com/chapter/10.1007/978-3-319-06028-6_93	"TripBuilder: A Tool for Recommending Sightseeing Tools" (Article, 2014) [13]	Direct	Yes
	http://ceur-ws.org/Vol-1404/paper_12.pdf	"Scaling up the Mining of Semantically-enriched Trajectories: TripBuilder at the World Level" (Article, 2015) [14]	Indirect	Yes
	http://link.springer.com/chapter/10.1007/978-3-319-06028-6_6	"A Personalised Recommendation System for Context-Aware Suggestions" (Article, 2014) [34]	Indirect	Yes
	https://www.londontoolkit.com/mnu/london_tours.htm	Travel company web site	Direct	No
	http://googlesightseeing.com/tools/	Virtual tours (not affiliated with Google)	Direct	No
sightseeing applications	http://www.tripomatic.com	Tripomatic Trip Planner and Sightseeing Travel Guide with Offline Maps	Direct	Yes
	https://itunes.apple.com/us/app/id554726752?mt=8	izi.TRAVEL - sightseeing, museum and landmark audio tour guide app for travelers	Direct	Yes
	http://www.pcmag.com/article2/0,2817,2422244,00.asp	"The Best Travel Apps of 2015" (PC News, Article, 2015) [35]	Direct	Yes
	http://www.sciencedirect.com/science/article/pii/S1084804513002105	Mobile application to provide personalized sightseeing tours (Article, 2014) [36]	Direct	Yes
	http://www.tourpal.com/	City audio guides	Direct	Partially
	https://play.google.com/store/apps/details?id=com.touristeye	Tourist Eye App form Lonely Planet	Direct	Partially
visual travel guide	http://www.visualtravelguide.com/	Photo collection	Direct	No
	http://www.dk.com/us/travel/	DK publishing travel guides	Direct	No
	https://www.virtualtourist.com/	Telling stories forum	Direct	No
	https://www.explorra.com/labs/travel-by-color	Discover travel destinations by color	Direct	Partially
travel itinerary construction	http://www.openu.ac.il/personal_sites/moran-feldman/	"Automatic Construction of Travel Itineraries using Social Breadcrumbs" (Article, 2010) [37]	Direct	Yes
	https://www.irjet.net/archives/V2/i2/Irjet-v2i221.pdf	"Efficient Multiuser Itinerary Planning for Travelling Services Using FKM-Clustering Algorithm" (Article, 2015) [38]	Direct	Yes
	http://kspt.ftk.spbstu.ru/media/files/2015/iwait-2015/proceedings/4.pdf	"Traveler Guide Assistant: Introducing an Application for an OpenStreetMap Based Travel Itinerary Construction" (Our previous paper) [16]	Direct	Yes
travel diary	www.traveldiariesapp.com/	Creating user travel diaries (pictures, texts, simple maps)	Direct	Partially
	https://play.google.com/store/apps/details?id=ch.robera.android.traveldiary	Travel diary app for Android	Direct	Partially
travel planning	https://www.triphobo.com/	Trip time scheduler	Direct	Yes
	https://www.tripit.com/	Travel scheduler	Direct	Partially

TABLE II
QUERIES AND RESULTS (PART 2)

Query	*Finding*			
	Link	*Description*	*Direct or Indirect*	*Relevance*
collaborative traveling	http://www.collaborativeconsumption.com/2014/06/25/	"Collaborative Economy Services: Changing the Way We Travel" (Internet article on economics and organization, not technology related)	Direct	No
	http://blogamadeus.com/18/01/	"Collaborative travel and tourism: the best way to predict the future is to invent it" (Blog article, 2012) [39]	Direct	Yes
collaborative traveling tools	https://travefy.com/	Product for travel agents and for travelers for planning collaborative itineraries	Direct	Partially
	http://toomanyadapters.com/7-collaboration-tools-travelling-entrepreneurs/	Entrepreneurship related article	Direct	No
	http://lifehacker.com/five-best-travel-planning-apps-1470002139	Internet review on travel planning [40]	Direct	Yes
	https://www.planapple.com/	Simple application for travel planning	Direct	Partially
travel tracking	https://www.tripit.com/destinations/the-ultimate-travel-tracker/	Distance tracking and mileage over time calculation	Direct	No
	http://www.trip-journal.com/	Mobile app for trip tracking, recording, documenting and sharing	Direct	Yes
	https://trackmytour.com/	Mobile app for creating online maps of your journey for friends and family to follow along	Direct	Yes
travel itinerary automation	http://triplantica.com/ru	Visual travel scheduler with e-map integration	Direct	Yes
	https://www.researchgate.net/publication/220542946	"An Automated Itinerary Planning System for Holiday Travel" (Article, 2003) [41]	Direct	Yes
	http://www.computer.org/csdl/trans/tk/2014/03/ttk2014030514-abs.html	"Automatic Itinerary Planning for Traveling Services" (Article, 2014) [17]	Direct	Yes
	http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=5767920	"Interactive Itinerary Planning" (Article, 2011) [42]	Indirect	Yes
	http://dl.acm.org/citation.cfm?id=2559893	"Customized tour recommendations in urban areas" (Article, 2014) [43]	Indirect	Yes
	http://dl.acm.org/citation.cfm?id=2534733	"Multi-day and multi-stay travel planning using geo-tagged photos" (Article, 2013) [44]	Indirect	Yes
	http://www.sciencedirect.com/science/article/pii/S0169023X14000962	"A system for mining interesting tourist locations and travel sequences from public geo-tagged photos" (Article, 2014) [45]	Indirect	Yes
	http://dl.acm.org/citation.cfm?id=2339562	"Constructing popular routes from uncertain trajectories" (Article, 2012) [46]	Indirect	Yes
	http://kspt.ftk.spbstu.ru/media/files/2015/iwait-2015/proceedings/5.pdf	"Automated leisure walk route generation for an interactive travel planner" (Our previous paper) [15]	Direct	Yes

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 439–445
DOI: 10.15439/2016F221
ACSIS, Vol. 8. ISSN 2300-5963

Semantic Knowledge Extraction from Research Documents

Rishabh Upadhyay, Akihiro Fujii
Department of Applied Informatics,
Hosei University, Tokyo, Japan
Email: uhrishabh@gmail.com, fujii@hosei.ac.jp

***Abstract*—In this paper, we designed a knowledge supporting software system in which sentences and keywords are extracted from large scale document database. This system consists of semantic representation scheme for natural language processing of the document database. Documents originally in a form of PDF are broken into triple-store data after pre-processing. The semantic representation is a hyper-graph which consists of collections of binary relations of 'triples'. According to a certain rule based on user's interests, the system identify sentences and words of interests. The relationship of those extracted sentences is visualized in the form of network graph. A user can introduce new rules to extract additional Knowledge from the Database or paper. For practical example, we choose a set of research papers related IoT for the case study purpose. Applying several rules concerning authors' indicated keywords as well as the system's specified discourse words, significant knowledge are extracted from the papers.**

***Index Terms*—Knowledge extraction; Semantics; Ontology; Discourse; Science and technology foresight**

I. Introduction

KNOWLEDGE extraction can be defined as the creation of information from structured or unstructured data. The general purpose of Knowledge discovery is to "extract implicit, previously unknown, and potentially useful information from data" [1]. Due to continuous growth of electronic articles or documents, automated knowledge extraction techniques become more and more important. The extraction of useful information from unstructured or semi-structured articles is attracting attention [3-7]. Knowledge mining can be characterized as concerned with developing and integrating a wide range of data analysis methods that are able to derive directly or incrementally new knowledge from large or small volumes of data from structured or unstructured sources using relevant prior knowledge [2]. Text mining tools in this context have the capability to analyze large or small quantities of natural language text and detect lexical and linguistic usage patterns [8]. The extracted information also should be machine understandable as well as human understandable in terms of open-data perspective.

This paper proposes a new method for Knowledge extraction or mining based on the integration of Semantics Technology (ST), Natural language processing (NLP) and Information extraction (IE). ST and NLP are significant topics in recent years. Knowledge extraction works in iterative manner, starting with a small set of rules which are tested on the available corpora or dataset and extended until the desired recall value is reached. The process of extracting knowledge is guided by certain rules inputting to the system which will define the knowledge according to the interests of a particular user. Semantic technology is based on RDF model. NLP concerns with the correction of the sentences and text which is obtained after IE phase. In this paper, we explore the benefit and application that can be achieved by the integration of these three areas for knowledge mining.

As we assume, one of the application of this system would be foresight scenario building based on the results where experts are working on writing or discussing technology trend of a certain field of research. Those experts are not necessary knowledgeable about technical issue written in a research paper, so that extracting fragments of knowledge and facts should be provided to the user compactly and easily in a limited time period.

In section 2, we give an introduction and background to Text mining as well as Technology forecasting. In section 3, introduce our model that we have used for knowledge extraction or mining. Details of the application of our system are introduced in section 4. Then in section 5 we examine the result acquired by the proposed model. Related works is presented in section 6. Conclusion and future works are introduced in section 7.

II. Background

A. Text Mining

Research in text mining has been carried out since the mid- 80s when the Prof Don Swanson, realized that, by combining information slice or fragments from seemingly unrelated medical articles, it was possible to deduce new hypotheses [13]. "Text mining" is used to describe the

978-83-60810-90-3/$25.00

application of data mining techniques to automated discovery of useful or interesting knowledge from unstructured text such as email, text documents etc. [9]. Several techniques have been proposed for text mining including association mining [10 and 11], decision tree [12], Machine learning, conceptual structure and rule induction methods. In addition, Information extraction and Natural language processing are the fundamental methods for Text Mining [23].

In Information extraction, natural language texts are mapped into predefine, structured representation, or templates, which, when it is filled, represent an extract of key information from the original text [15]. So the IE task is defined by its input and output. The work of [16] and [17] present text mining using IE. The Application of text mining can be extended to various sectors where text documents exist. For example, Crime detection can profit by the identification of similarities between various crimes. Some of the past researches in the field of Text mining or knowledge extractions are as follows:

Rajman and his colleagues [24] presented an approach for knowledge extraction using Text Mining technique. They have presented two example of information that can be extracted from text collections- probabilistic association of keywords and prototypical document instances. They have given the importance of the Natural language processing tools for such knowledge extraction. So his was the base for our method.

Alani and his team [25] have provided an updated for Artequakt System. This system uses Natural Language processing tools to automatically extract knowledge about artists from documents using predefined ontology. Steps for knowledge extraction are as follows: Document Retrieval, Entity Recognition and Extraction procedure. In knowledge extraction procedure, consists of Syntactical analysis, Semantic Analysis and Relation extraction. They have produced acceptable results.

Peter Clark and Phil Harrison [26] worked on knowledge extraction by making database of "tuples" and thus capturing the simple word knowledge. And then using it in improving parsing and the plausibility assessment of paraphrase rules used in textual entailment.

Parikh [27] proposed an approach to learn a semantic parser for extracting nested event structures with or without annotated text. The idea behind this method is to model the annotations as latent variables and incorporate prior that matched with Semantics parses of the events.

B. Technology Forecasting and foresight:

Technology forecasting is used widely by the private sector and by governments for applications ranging from predicting product development or a competitor's technical capabilities to the creation of scenarios for predicting the impact of future technologies [19]. It is "the prediction of the invention, timing, characteristic, performance, technique or process serving some useful purpose". Detailed account of achievements and failures of the technology forecasting over the four decades is given by Cuhls [19]. Johnston [20] proposed five stages in the chronology of foresight, technology forecasting and futurism leading to technology foresight, which can be used for wide understanding of the economic and social processes that shape technology. Foresight can be referred as systematic process of reflection and vision building on the future among a group of stakeholders. Foresight is nowadays referred as an element in a continuous policy learning process that is contributing to a more effective mode of policy making [21]. In a European research group, foresight is described as "...a systematic, participatory, future intelligence-gathering and medium-term vision-building process aimed at present-day decisions and mobilizing joint actions" [22].

III. RELATED WORKS

Their have been many research in the field of literature mining or extracting knowledge from research documents. But most of them are related to Biomedical or medicinal field. Our approach was related to the field of Technologies. QasemiZadeh [33] have presented an approach for structure mining from Scientific or research papers. He has only processed using Language processing, but we have combined three main fields to extract knowledge. Cimiano et. al. [34] gave a survey of current methods in ontology constructions and also discussed relation between ontology and Natural language processing.

Mima et. al. [35] gave a method for knowledge management using ontology-based similarity calculation. We have also used ontology for extracting knowledge but apart from ontology we have also given emphasis on Natural language processing and a bit of Information extraction (early stage). Mima have also not presented any information regarding evaluation of the system. Hahm et. al. [36] presented an approach for knowledge extraction from web data. Triple store was produced using web data. But in our method we have given more emphasis on research document and producing triple with it and then ontologies is applied on the produced triple datasets of line and sentences.

IV. PROPOSED SYSTEM

Proposed system uses combination of semantics of sentences and natural language processing technique over the sentences. It also provides visualization of the result. We do not expect fully machine processing results from the system. In a sense that after some processing by inference rule and getting sentences which might be significant, user creativity is required to understand what is written in the document. In this sense, our method is hybrid with software processing and expert knowledge in the area. The following Fig.1 describes the proposed model of our system.

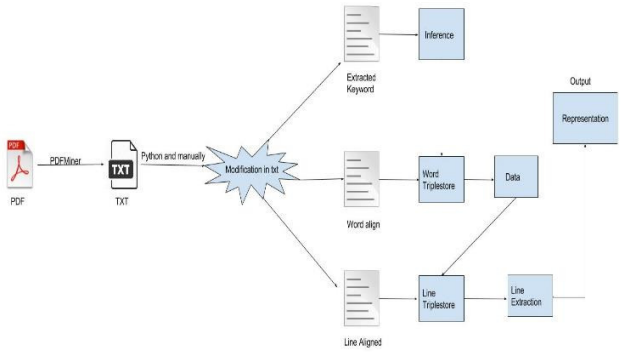


Fig. 1 Pipeline of Proposed System

Our model consists of the following Steps which is given below:

1. Extracting information from PDF file to text file.
2. Pre-processing of data.
3. Extracting the Keywords.
4. Extraction of Discourse words
5. Building a simple Triple-store (Word and Line).
6. Inference Rule.
7. Visualizing the data.

1) Extracting information from PDF file to Text file:

We made a dataset of IOT research papers. We had about 69 research papers dataset. We used some paper out of those papers. The paper were in pdf format. So to work on them we have to extract information from the pdf files. So to extract information we used Python Library "PDFMiner" [29].

2) Pre-processing of data:

After extracting information from *pdf* file into text file pre-processing was done. The extracted documents had problems, as extracted information was on a single line. So to work on them the modification of the extracted information is required. So the modification includes removing the noise (image), make proper alignment of the sentences, checking the extracted data.

So we removed the noise which is caused by extracting the image from the pdf file. As image will contains least information in the form of text so we ignored the text extracted by image. We align each sentence on the separate line and also checked the percent of the extracted data. Most of the file had some errors so we corrected it manually.

3) Extracting the Keywords:

As this system will be extracting the important knowledge from the research papers, rules should be created according to the important Keywords. So, the easiest way was to extract the Keywords from the research papers.

We worked on the research paper with the standard structure such as Title, Abstract, Keywords then Main text. So the extraction of Keywords from the research paper takes places only for specific structure. So we automated the process of extracting the Keyword using python's regular expression. After getting keywords the frequency of the extracted words is obtained. Because those frequent words will be used by the inference rule to create new rules for extracting useful sentences from documents.

4) Extraction using Discourse words:

Discourse words are the words which give "important message" that helps us to understand or comprehend a text. This discourse words ranges from Numbering words to Adding words, linking words to Contrasting words (however) etc. These words are used everywhere from articles to research papers. So we emphasis on these words to represent knowledge or message from the research articles. We went through 5-6 paper to get the list of the discourse words. Those words were used in making new rules for the extraction of the message from the articles. Then after creating the rules for those 5-6 papers we then used those words to create the rules for other research documents to check, if these words can be generalized.

5) Building a simple Triple-store:

Till now we were doing the pre-processing of the data and collecting words for making new rules. In this section we will discuss about the schema to analyze the sentence and word data. We focused on two triple-stores that is, sentence and words separately.

There are many semantics toolkit available. We have referred Python code [28]. We choose this programming language because of its simplicity and flexibility towards various toolkits in the field of semantics and Artificial intelligence.

So from the above toolkit we have produced a dataset of sentences and words. In fact, this dataset is of three type formats that is why single data it is known as "triple". The three type formats are Subject, Predicate and Object.

We have maintained two triple-stores. The format of both the Triple-Store is shown below.

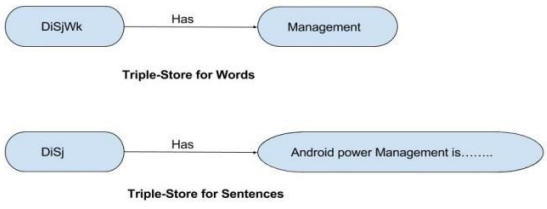


Fig.2 RDF graph example

Above Fig.2 shows a triple syntax called an RDF graph. RDF, an abbreviation for "Resource Description Framework," is a concept adopted in defining knowledge structure. Knowledge fragments are given in a syntax consisting of three elements: the subject, the predicate, and the object.

Let's consider the triple store for words for the above example, the subject is "*DiSjWk*", Predicate is "*Has*" and Object is "Management". So the subject *DiSjWk* stands for *i* th Document, *j* th Sentence and *k* th word. We have chosen the predicate to be "*Has*". The Object will be the extracted word from the paper.

For the sentences triple-Store, We have subject "*DiSj*", Predicate "*Has*" and Object will be the Sentence extracted from the document. Subject *Di* stands for *i*th Document and *Sj* stands for *j*th sentence. This Triple-Store will be base for our Experiments.

So this triple-Store will be used for extracting the knowledge or useful terms from the documents. For extracting the so called "knowledge", we will use Inference rule, which is introduced in next step.

6) Inference Rule:

Inference is the process of deriving new information from the information you already have [28]. So the "information" and which rule to apply to extract the information will vary depending on the context. As we have explained the structure of the Knowledge fragments that are given in a three elements structure. So to describe ontologies, logical expression is configured. So the process to configure the logical expression uses a syntax called predicate. Ontologies are written in OWL. Ontology is defined as the explicit and abstract model representation of already defined finite sets of terms and concepts, involved in knowledge management, knowledge engineering, and intelligent information integration [30]. In simple word, ontology is the collection and classification of terms to recognize their semantic relevance. OWL stands for Web Ontology Writing Language; its standardization has been conducted by W3C. To describe a knowledge structure in a predicate logic, a set of elements that meet a certain condition is constructed, such as "*If A, then B*". After construction, the resulting set is Fundamental to knowledge processing in the semantic data processing. Knowledge processing based on the predicate logic takes the form of generating answers from the collection of the knowledge fragments, such as *"If A is true, then B is satisfied*" and "*If B, then C,*" to queries such as "*Is Z true, if A holds?*". This process is referred to as the reasoning mechanism.

The basic pattern of inference is simply to query the information relevant to the rule and then apply the transformation that turns these bindings into a simple triple data added to the main Triple-Store, with appropriate Subject, predicate and Objects. After getting this new triple data, we use this information to process the knowledge from whole tripe-store, so Fig.3 gives the insight of our Triple data sets.

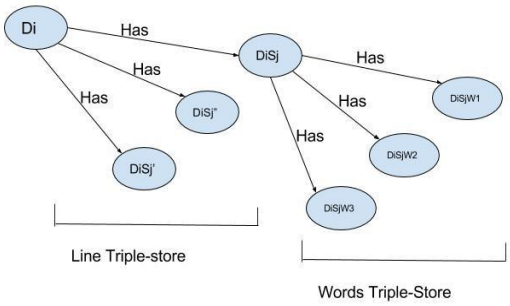


Fig.3 Triple Data Example

We have data sets with line and word Triple store. The knowledge processing or extraction of line data set uses the words triple store. We use the rule on word triple store, added the new triple data using inference rule and then we extract the knowledge from the line data sets. The steps are shown in the Fig.4 below.

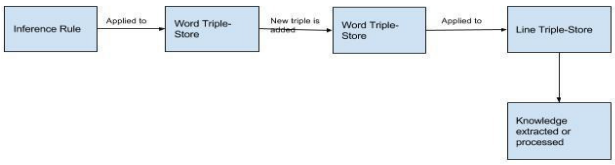


Fig.4 Pipeline of the Inference System

7) Visualization:

So, using the above process we extracted the useful and meaningful line from whole documents. Then we emphasized on visualization of the extracted knowledge. For visualization, we created new triple dataset to make a relationship between words and line. This dataset will show which word is in which line. Subject will be word, Predicate will be "Has" and Object will be sentences. So using this we got a third triple store that will consist of words and sentences together. So we will use this RDF file for visualization of the knowledge.

We used *GraphViz* graph drawing package [31]. This library has support extrinsic data in its graph drawing algorithm. It offers one of the best solution for drawing directed cyclic graphs which is state-of-the-art in graph drawing. *GraphViz* application *neato* [32] is a straightforward method for rapidly visualizing undirected graphs. The graphs are generated by using:

"neato-Teps undirected.gv > undirected.eps",

Where undirected.gv is a file that contains the code of the graph, –Teps specifies an Encapsulated Postscript output format. So using the above given library we visualized our knowledge.

V. APPLICATION

Natural language processing (NLP) is one of main field of artificial intelligence study. Extraction of meaningful information from vast amount of document database is an example of the features that NLP possibly contribute well in performing scenario building practices.

On the other hand, Science and Technology Foresight (STF) is an activity to identify important research areas or directions as well as to search key features and driving forces for those researches. For STF, there are numerous papers and books on methodologies. The scenario development is one of main measures for such future-oriented work. However, even in scenario development alone, there is wide range of possible practices depending on purpose of outcome, participant's knowledge etc.

For the scenario building method in STF, the creativity of those who participating the workshop is the most important. Usually such an activity is held in limited time scale. In order to stimulate the creativity of participant in short period of times at scenario building workshop, for example, quick identification of important sentences of the area of interests from large number of research papers is quite helpful. At the same time results of extractions could be provided with visual imaging with of data processing and graphic visualization software. With the help of such a system, human experts may exercise more efficiently during scenario building workshop.

In this context, we have designed an artificial intelligence support for the scenario building activities in this paper. We applied semantic information about structure of sentences in research papers and support human user quick visualization of extracting of sentences and key words for the purpose.

By using algorithm or method, we have achieved two aspects to a certain extent. Scenario writing is one of the example of such system could support efficiently. In the process, firstly some key driver is necessary. We may use current algorithm to get meaningful sentences that may help understanding scenario drivers. Then our algorithm also could produce different sentences with slightly different angle of interests.

The extracted sentences and words might help the scenario writer to write down case scenario.

VI. EXPERIMENTAL RESULTS.

This section show the example of using Hybrid method. We have a dataset of 69 research papers which we developed using varius sources. This dataset consist of paper which was pubished in IEEE journals and conferences. For results we have given emphasis on two methods, that are:

- Extracting the knowledge using Keywords from the paper.
- Extracting the knowledge using discourse Keywords.

As the dataset consist of the research paper which were in PDF format. So the first step is to extract the information from the PDF document so that it would be easier to work on them. PDFminner was used for this process.Then the prepocessing of the Information takes place. In prerocessing the unwanted extracted information was removed. While using PDFminner, the images in the document was extracted but contains many noise or unwanted characters. So preprocessing emphasis on removing the unwanted characters and also aligning the line using python. Then after the preprocessing we give importance on extarcting the keyword from all the documents and then we can choose the most frequent words from the document. These words were used directly from the word mentioned in the keywords section. As the keyword section consists of the words which are important acording to the context. So all this processing and counting of the word is done automatically using python. The Table1 shown below consists of frequent term in all the 69 documents:

Table 1 Keywords Examples of the 69 papers

Internet of Things	42
Wireless-network sensors	22
RFID	13
Social network	8
System security	8
Energy Efficiency	7
Service Management	7
Enterprise systems	6
Learning technology	5
6lowpan	4

After this the formation of tripe-store takes place. The dataset consists of 69 documents, with about 21K lines and 300K words before filtering. After filtering we got about 200K words and 21K lines. So two triple-store are created one for lines and one for words. Subject is choosen such that it have some relationship between two triple-store. After the making of triple-store and all process now the important step is to extract the knowledge using INFERENCE rule. This consist of the SPARQL type query. So the user makes important rule ,through this rule the knowledge is extracted. The main step in this method is INFERENCE rule. This rule can be created by user according to their preference. So this

rule will help to extract the knowledge from the Huge database rather then reading all the documents. To ease the understanding we also given importance to visualise the knowledge.

The above method was related to our first method that we mentioned in the start of this section.

In next method, we gave attention to the discourse keywords in the documents. Discourse keywords is the keywords which talks and give information about the text. After going through set of papers we came up with some Discourse keywords which are listed in Table2.

Table 2 Examples of Discourse Keywords

consists	become	aimed	Instead
useful	capable	using	Provide
method	propose	enhance	application
future	explore	aspects	Discover
objectives	focused	pedagogy	Crucial
different	various	integration	Import
promote	reflect	classified	Need

So new rule is generated using the discourse term which is mentioned above. We used this words in two form first on training set then on test set. Using training documents we got those words, then we used these words in some other documents to check the effectiveness of the words.

Here, we chose one of the 69 papers as a case study example. Document [37] is a servey paper about IoT applications of RFID(Radio Frequency IDentifier). We extracted some useful sentences which can be considered as significant knowledge descriptions about the application. Some example sentences are given below with the associated discourse term as results from the system.

"aimed" -- "this paper is aimed at drawing a landscape of the current research on rfid sensing from the perspective of iot for personal healthcare

"application" -- "thanks to the recent advantages in biomaterial engineering, a rich variety of tattoo-like thin surface electronic devices have been experimen- ted in the last three years

"based" -- "the body-centric networks investigated so far are mostly based on active devices"

"bring" -- "data processing and human behavior analysis the interactions between people and their habitat bring precious information on behavioral parameters concerning the activity rate during the different phases of the day and of the night"

"useful" -- "air monitoring is also useful in common spaces with the purpose to avoid the inhalation of anesthetic gases"

Althought those are not whole result but only the fragments of whole content of the paper, sentences indicates condenced informations about several issues discussed in the paper. We have evaluated usefulness of extracted sentences over 10 documents out of 69 papers. In each paper, there are around 10 discourse terms, but most of them were same as before and after that the inference rules were introduced based on the network visualization in the process of extraction.

In the Fig.5 we have presented both the methods i.e. Keywords and Discourse words. So first we have extracted Keywords from both the sources and then we have used the semantic analysis on it to extract the knowledge from it. The output is in the form of graph which is easy to understand. One example of the extracted knowledge is given below which is extracted by using keyword "Pedagogy".

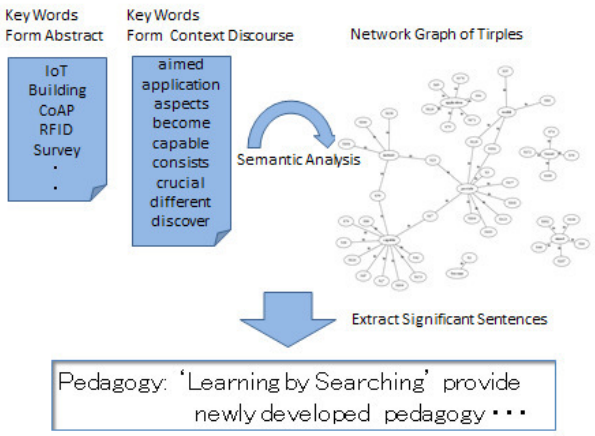


Fig.5 Example Visualization of Konwledge

Apart from this to check the extracted information we calculated the precision of the information. So after using this Hybrid method, which range from Information extraction to Semantics for knowledge extraction, the precision of the model was good.

VII. Conculsion And Future work.

In this article, we have discussed practical sentence extraction procesure and supporting system which we intended to call knowledge extraction system. Since processed data is always stored in a form of triples, resulted dataset is always fully machine readable in every stage of cyclic extraction and cleaning of data. The system assumes human exparts support in selecting so called discourse keywords. Such characteristics is useful and practical in the situations where exparts need to aquire a certain level of knowledge in a research area such as Science and Technology Foresight activities. As we have also shown, that the above introduced or obtainted discourse words can be used in any research documents and on the basis of that words useful information can be obtainted.

There are two directions for enhancing the system. One is to inroduce more sophisticated inference rules over sentences. Purhaps it will come with NLP technique with looking at grammatical structure of sentences. In addition to this, the system can try to extract discourse words from the extracted lines (using Keywords) rather than those words which are mentioned by us. Another direction goes toward practical utilizations. Big data analysis is currently one of the most needed technology in IT related services. The availability of software based data processing is very important aspect of reusing and deepining knowledges obtaind in a stage of processing.

REFERENCES

[1] W. Frawley and G. Piatetsky-Shapiro and C. Matheus, Knowledge Discovery in Databases: An Overview. AI Magazine, 1992,213-228.

[2] Michalski, R.S.: Knowledge Mining: A Proposed New Direction, In: Invited talk at the Sanken Symposium on Data Mining and Semantic Web, Osaka University, Japan, March 10-11, 2003.

[3] Jérôme Darmont, chair. Proceedings of the 15th international conference on extraction and knowledge management, Luxembourg, 2015.

[4] Jerzy Grzymala-Busse, Ingo Schwab, Maria Pia di Buono, editor. Proceedings of the second on Big Data, Small Data, Linked Data and Open Data (ALLDATA 2016) workshop on Knowledge Extraction and Semantic Annotation. Portugal, 2016.

[5] International World Wide Web Conferences (WWW 2015) Second workshop on Knowledge Extraction from Text, Italy, 2015.

[6] XIV Conference of the Spanish Association for Artificial Intelligence (CAEPIA 2011) workshop on Knowledge Extraction and Exploitation from semi-Structured Online Sources, Spain, 2011.

[7] 1st international Workshop on Knowledge Extraction and Consolidation from Social Media collocated with the 11th International Semantic Web Conference (ISWC), USA, 2012.

[8] F. Sebastiani, "Machine learning in Automated Text Categorization," ACM Computing Surveys, vol. 1, no. 34, pp. 1–47, 2002.

[9] J. Han and M. Kamber. Data Mining: Concepts and Techniques. Morgan Kaufmann, San Francisco, 2000.

[10] Dion H. Goh and Rebecca P. Ang (2007), "An introduction to association rule mining: An application in counseling and help seeking behavior of adolescents", Journal of Behavior Research Methods39 (2), Singapore, 259-266.

[11] Pak Chung Wong, Paul Whitney and Jim Thomas," Visualizing Association Rules for Text Mining", ", International Conference, Pacific Northwest National Laboratory, USA, 1-5.

[12] C. Apte and F. Damerau and S. M. Weiss and Chid Apte and Fred Damerau and Sholom Weiss," Text Mining with Decision Trees and Decision Rules", In Proceedings of the Conference on Automated Learning and Discorery, Workshop 6: Learning from Text and the Web, 1998.

[13] J. Nightingal, "Digging for data that can change our world," the Guardian, Jan 2006.

[14] Grishman R. (1997), "Information Extraction: Techniques and Challenges", International Summer School, SCIE-97.

[15] Wilks Yorick (1997), "Information Extraction as a Core Language Technology", International Summer School, SCIE-97.

[16] H. Karanikas, C. Tjortjis, and B. Theodoulidis, "An approach to text mining using information extraction," in Proceedings of Workshop of Knowledge Management: Theory and Applications in Principles of Data Mining and Knowledge Discovery 4th European Conference, 2000.

[17] U. Nahm and R. Mooney, "Text mining with information extraction," in Proceedings of the AAAI 2002 Spring Symposium on Mining Answers from Texts and Knowledge Bases, 2002.

[18] Committee on Forecasting Future Disruptive Technologies; Air Force Studies Board; Division on Engineering and Physical Sciences; National Research Council. "Persistent Forecasting of Disruptive Technologies",2009

[19] Cuhls K," From Forecasting to Foresight Processes – New Participative Foresight Activities in Germany", Journal of Forecasting, 23, pp 93–111 European Foresight Monitoring Network, available at http://www.efmn.info/.

[20] Johnston R," The State and Contribution of Foresight: New Challenges". In Proceedings of the Workshop on the Role of Foresight in the Selection of Research Policy Priorities' IPTS, Seville.

[21] Weber, M., 'Foresight and Adaptive Planning as Complementary Elements in Anticipatory Policy-making: A Conceptual and Methodological Approach' In: Jan-Peter Voß, Dierk Bauknecht, René Kemp (eds.) Reflexive Governance For Sustainable Development Edward Elgar, pp. 189-22.

[22] FOREN 2001: A Practical Guide to Regional Foresight. FOREN network, European Commission Research Directorate General, STRATA programme.

[23] S. Jusoh and H. M. Alfawareh, "Techniques Techniques, Applications and Challenging Issue in Text Mining." IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 6, No 2, November 2012.

[24] Martin Rajman and Romaric Besancon,"Text mining- Knowledge extraction from unstructured textual data".In:Proceedings of the 6th Conference of the International Federation of Classification Societies, Rome, 1998.

[25] Alani, Harith, Kim, Sanghee, Millard, David E., Weal, Mark J., Lewis, Paul H., Hall, Wendy and Shadbolt, Nigel R," Automatic Extraction of Knowledge from Web Documents", Wendy; Lewis, Paul H. and Shadbolt, Nigel R. In, *2nd International Semantic Web Conference - Workshop on Human Language Technology for the Semantic Web abd Web Services, Sanibel Island, Florida, USA,20 - 23 Oct 2003.*

[26] Peter Clark and Phil Harrison,"Large-Scale Extraction and Use of Knowledge From Text", In: Proceedings of the fifth international conference on Knowledge capture(K-CAP '09),USA, 2009.

[27] Ankur P. Parikh; Hoifung Poon; Kristina Toutanova," Grounded Semantic Parsing for Complex Knowledge Extraction",In: Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies,2015.

[28] Toby Segaran, 'Semantic Web Programming', O'reilly, 2009.

[29] Shinyama, Y. (2010) PDFMiner: Python PDF parser and analyzer. Retrieved on 11 June 2015 from:http://www.unixuser.org/~euske/python/pdfminer/.

[30] D.Fensel, "Ontologies: Silver Bullet for Knowledge Management and e-Commerce", Springer Verlag, Berlin, 2000.

[31] J. Ellson, E. R. Gansner, L. Koutsofios, S. C. North, and G. Woodhull. Graphviz — open source graph drawing tools. In P. Mutzel, M. Junger ¨, and S. Leipert, editors, Proc. 9th Int. Symp. Graph Drawing (GD 2001), number 2265 in Lecture Notes in Computer Science, LNCS, pages 483–484. Springer-Verlag, 2002.

[32] Stephen C. North, "Drawing graphs with NEATO", NEATO User manual, April 26, 2004.

[33] Behrang QasemiZadeh, "Towards Technology Structure Mining from Scientific Literature", 9th International Semantic Web Conference, ISWC 2010, Shanghai, China, November 7-11, 2010.

[34] Cimiano, P., Buitelaar, P., Völker, J.: Ontology construction. In: Indurkhya, N., Damerau, F.J. (eds.) Handbook of Natural Language Processing, 2nd edn., pp. 577–605.

[35] Mima, H., Ananiadou, **S**. & Matsushima, K. (2004) Design and Implementation of a Terminology-based literature mining and knowledge structuring system, in Proceedings of international workshop of Computational Terminology, CompuTerm, Coling, Geneva , Switzerland.

[36] Younggyun Hahm, Hee-Geun Yoon, Se-Young Park, Seong-Bae Park, Jungwon Cha, Dosam Hwang, Key-Sun Choi, Towards Ontology-based Knowledge Extraction from Web Data with Lexicalization of Ontology for Korean QA System, Submitted to NLIWoD, 2014.

[37] S. Amendola, R. Lodato, S. Manzari, C. Occhiuzzi, and G. Marrocco "RFID Technology for IoT-Based Personal Healthcare in Smart Spaces", IEEE INTERNET OF THINGS JOURNAL, VOL. 1, NO. 2, APRIL 2014.

6th International Workshop on Dealing with Spatial and Temporal Uncertainty and Imprecision

MANY fields of research concern data that carry a spatial or temporal component: data related to a location or time, which often are prone to uncertainty and/or imprecision. As both spatial and temporal data tend to be an additional aspect of the data and as both have similar issues with resolution, relative notions and intuitive notions, they are often considered together. The combination of uncertainty and/or imprecision, an increasing amount of data and the need for a higher level of reasoning on the data requires innovative methods.

The first iteration of this event concerns estimating, handling and representing uncertainty within the model, but also new ways of modelling and processing uncertain spatial or temporal data. The main focus is on novel technologies, such as possibilistic and fuzzy methods, to represent, limit or estimate the uncertainty or imprecision; but also on algorithms that involve such methods to process the data accordingly.

TOPICS

- uncertain spatial data
- spatial data analysis
- fuzzy set theory in spatial data
- use of artificial intelligence in spatial data
- fuzzy topology
- natural language in spatial data
- uncertain temporal data
- temporal data analysis
- fuzzy set theory in temporal data
- use of artificial intelligence in temporal data
- uncertainty in time series
- natural language in temporal data

EVENT CHAIRS

- **Verstraete, Jörg,** Systems Research Institute PAS, Poland

PROGRAM COMMITTEE

- **Billiet, Christophe,** Ghent University, Belgium
- **Bun, Rostyslav,** Lviv Polytechnic National University, Ukraine
- **Danylo, Olha,** IIASA, Austria
- **De Tré, Guy,** Ghent University, Belgium
- **Dyczkowski, Krzysztof,** Adam Mickiewicz University, Poland
- **Gągolewski, Marek,** Systems Research Institute - Polish Academy of Sciences, Poland
- **Horabik, Joanna,** Systems Research Institute - Polish Academy of Sciences, Poland
- **Kaczmarek, Katarzyna,** Systems Research Institute - Polish Academy of Sciences, Poland
- **Kruse, Rudolf,** Otto Von Guericke University of Magdeburg, Germany
- **Lesiv, Mirosława,** IIASA, Austria
- **Marrara, Stefania,** Information Retrieval Lab DISCo - University of Milano, Italy
- **Petry, Fred,** Naval Research Laboratory, United States
- **Tainio, Marko,** Cambridge University, United Kingdom
- **Van de weghe, Nico,** Ghent University, Belgium
- **Wilke, Gwendolin,** Lucerne University of Applied Science and Arts, Switzerland

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 449–457
DOI: 10.15439/2016F539
ACSIS, Vol. 8. ISSN 2300-5963

Uncertainty of Spatial Disaggregation Procedures: Conditional Autoregressive Versus Geostatistical Models

Joanna Horabik-Pyzel
Systems Research Institute
of Polish Academy of Sciences
ul. Newelska 6, 01-447 Warsaw, Poland
Email: joanna.horabik@ibspan.waw.pl

Zbigniew Nahorski
Systems Research Institute PAS
and Warsaw School of Information Technology
ul. Newelska 6, 01-447 Warsaw, Poland
Email: zbigniew.nahorski@ibspan.waw.pl

***Abstract*—Consider the problem of allocation of spatially correlated gridded data to finer spatial scale, conditionally on covariate information observable in a fine grid. Spatial dependence of the process can be captured with the conditional autoregressive structure, suitable for gridded (areal level) data. Also geostatistical methods, particularly empirical universal kriging, can be used for this purpose. In this study, we compare prediction results as well as prediction standard errors for two disaggregation procedures, based on the inventory of agricultural ammonia emissions reported in Pomeranian Voivodeship of Poland.**

I. Introduction

IN MANY environmental and epidemiological applications, one has to deal with spatial variables observed at different resolutions. The change of support problem is encountered, for example, in a development of high-resolution inventories of greenhouse gases [1], [2] or ammonia emissions [3].

The choice of a relevant model capturing spatial correlation depends on a type of data, but also it can depend on a size of dataset and computational efficiency. In principle, the model suitable for areal data is based on Markov random fields, in particular the commonly used conditional autoregressive structure. However, the point-referenced data can be aggregated to the area level, and modelled the same way [4]. On the other hand, the geostatistical approach, designed for continuous spatial processes, can be used to model a process over a gridded domain.

In this paper, we aim to explore uncertainty underlying a particular procedure of areal data disaggregation from a coarse to fine grid. The setting assumes knowledge on (i) a variable of interest in a *coarse* grid, and (ii) some related variables (proxy data) in a *fine* grid. The task is accomplished with two alternative approaches to modelling spatial data: the one based on conditional autoregressive model, and the other, using geostatistical methods. These models represent two general classes commonly used in spatial statistics. Within geostatistical approach, empirical universal kriging was applied for a prediction of unknown values in a fine grid. For each model we compare the prediction standard errors against actual residuals (their absolute values), as resulting from an empirical study of ammonia emission inventory. In addition, we analyse the effect of the level of disaggregation.

II. Motivating Data Set

A. Inventory of ammonia emissions

The analysed dataset concerns ammonia (NH_3) emission inventory in a region of Poland. Ammonia is emitted mainly (up to $80-90\%$) by agricultural sources such as livestock production and fertilized fields [5], [6]. High concentrations of ammonia can lead to acidification of soils [7], forest decline, and eutrophication of waterways [6]. All of these lead to loss of plant biodiversity [8]. Moreover, ammonia emissions are recognized for their importance in contributing to fine particulate matter [9], hence their spatial distribution is of great importance.

However, agricultural emission sources cannot be measured directly, and spatial emission patterns need to be assessed otherwise. This issue was addressed, for instance in [3], where agricultural and land cover data were used to disaggregate the national NH_3 emission totals across Great Britain. This was accomplished employing a spatially weighted redistribution of emission sources, with weights based on respective landcover classes. It was demonstrated in [10] that this type of straightforward, linear approaches to spatial allocation can be substantially improved by introducing a spatial random effect modelled with a conditional autoregressive structure.

B. Data description

The dataset comprises the gridded inventory of ammonia emissions from fertilization (in tonnes per year), reported in Pomeranian Voivodeship of Poland. The inventory grid cells are of regular 5km×5km size, and the whole of cadastral survey compiles $n = 800$ cells, denoted $\boldsymbol{y} = (y_1, \ldots, y_n)^T$; see Fig. 1.

It should be noted, that the considered variable y of ammonia refers to a total amount of emissions over a grid cell; it is called an *extensive* variable [11]. This should be distinguished from *intensive* variables, e.g. emission concentrations or proportions over a geographic region.

978-83-60810-90-3/$25.00

For explanatory information we use the CORINE Land Cover Map for this region, available from the European Environment Agency [12]. Specifically, for each single grid cell we calculate area (in m^2) of those land use classes that are related to ammonia emissions. The following CORINE classes were considered (for reference, the CORINE class numbers are given in brackets):

- Non-irrigated arable land (211), denoted $\boldsymbol{x}_1$;
- Fruit tree and berry plantations (222), denoted $\boldsymbol{x}_2$;
- Pastures (231), denoted $\boldsymbol{x}_3$;
- Complex cultivation patterns (242), denoted $\boldsymbol{x}_4$;
- Principally agriculture, with natural vegetation (243), denoted $\boldsymbol{x}_5$.

Only land use data are used as explanatory information. Also, it should be pointed out that our modelling approach includes both a regression component as well as a spatial correlation component, and the resulting regression coefficients are not the same as typical emission coefficients, specific for the listed land use classes.

Performance of a disaggregation framework depends on various factors. Among others, it highly depends on the extent of disaggregation; note, that this is connected with a preservation of the correlation across spatial scales. An impact of this feature is evaluated in the study. We test the disaggregation from 10km×10km and 15km×15km (coarse) grids to a 5km×5km (fine) grid. To examine performance of the disaggregation procedure, first we aggregate the original fine grid emissions into respective coarse grid cells. Next, we fit respective model and predict ammonia emissions for a 5km fine grid. Finally, we check obtained results with the original inventory emissions of a 5km grid. Thus, our simulation study tests the cases of a fourfold and ninefold disaggregation. The aggregated values of the two coarse grids as well as the actual inventory data in the fine grid are shown in Fig. 1.

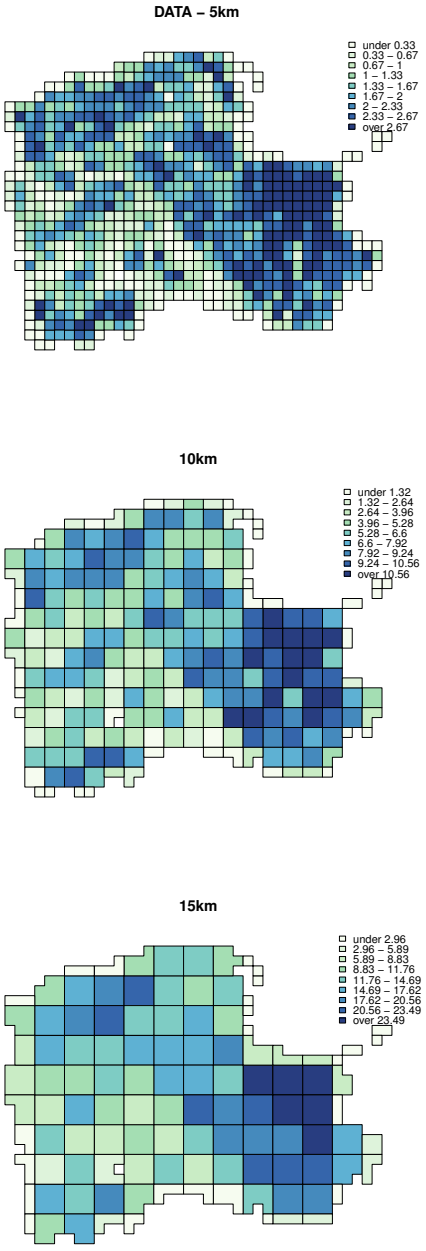


Fig. 1. Ammonia emissions (in tonnes/year): inventory data in 5km grid, and aggregated values in 10km and 15km grids

III. Disaggregation Model Based on Conditional Autoregressive Structure

In this section we present an approach for areal to areal data realignment, where the residual covariance structure is modelled with the conditional autoregressive (CAR) specification [13], [14]. This class of models is used in the case of areal data, and it introduces spatial association through a neighbourhood structure.

A. The model

1) Fine grid: We begin with the model specification in a fine grid. Let $\boldsymbol{Y} = \{Y_i\}_{i=1}^n$ denote random variables associated with missing values of interest $\boldsymbol{y} = \{y_i\}_{i=1}^n$ defined at each cell i, $i = 1, ..., n$ of a fine grid. Assume that random variables Y_i follow a Gaussian distribution with respective mean and variance, $Y_i|\mu_i, \sigma_Y^2 \sim \mathcal{N}\left(\mu_i, \sigma_Y^2\right)$. Given the values μ_i and σ_Y^2, the random variables Y_i are independent.

The values $\boldsymbol{\mu} = \{\mu_i\}_{i=1}^n$ represent the underlying mean process, and the (missing) observations in a fine grid are related to this process through a measurement error of variance σ_Y^2. The model for the underlying mean process is formulated as a sum of regression component with available covariates, and a spatially varying random effect.

The approach to modeling μ_i expresses an assumption that available covariates explain part of the spatial pattern, and the remaining part is captured through a spatial dependence, introduced as the conditional autoregressive CAR model. The CAR scheme follows an assumption of similar random effects in adjacent cells, and it is given through the specification of

full conditional distribution functions of μ_i for $i = 1, \ldots, n$ [15], [16]

$$\mu_i|\mu_{j,j\neq i} \sim \mathcal{N}\left(\boldsymbol{x}_i^T\boldsymbol{\beta} + \rho\sum_{\substack{j=1\\ j\neq i}}^{n}\frac{w_{ij}}{w_{i+}}\left(\mu_j - \boldsymbol{x}_j^T\boldsymbol{\beta}\right), \frac{\tau^2}{w_{i+}}\right), \tag{1}$$

where $\boldsymbol{\mu}_{-i}$ denotes all elements in $\boldsymbol{\mu}$ but μ_i; w_{ij} are the adjacency weights ($w_{ij} = 1$ if j is a neighbour of i and 0 otherwise, also $w_{ii} = 0$); w_{i+} is the number of neighbours of area i; $\boldsymbol{x}_i^T\boldsymbol{\beta}$ is a regression component with explanatory covariates for area i and a respective vector of regression coefficients.; and τ^2 is a variance parameter.

The conditionals (1) yield the following joint distribution of the process $\boldsymbol{\mu}$, see e.g. [15]

$$\boldsymbol{\mu} \sim \mathcal{N}_n\left(\boldsymbol{X}\boldsymbol{\beta}, \tau^2\left(\boldsymbol{D} - \rho\boldsymbol{W}\right)^{-1}\right), \tag{2}$$

where $\boldsymbol{X}$ is a design matrix with vectors $\boldsymbol{x}_i$; $\boldsymbol{D}$ is an $n \times n$ diagonal matrix with w_{i+} on the diagonal; and $\boldsymbol{W}$ is an $n \times n$ matrix with adjacency weights w_{ij} . Equivalently, we can write (2) as

$$\boldsymbol{\mu} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}_n\left(\boldsymbol{0}, \boldsymbol{N}\right), \tag{3}$$

denoting also $\boldsymbol{N} = \tau^2\left(\boldsymbol{D} - \rho\boldsymbol{W}\right)^{-1}$.

2) Coarse grid: The model for a coarse grid (aggregated) observed data is obtained by multiplication of the mean process (3) with an $N \times n$ *aggregation matrix* $\boldsymbol{C}$, where N is a number of observations in a coarse grid

$$\boldsymbol{C}\boldsymbol{\mu} = \boldsymbol{C}\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{C}\boldsymbol{\epsilon}, \quad \boldsymbol{C}\boldsymbol{\epsilon} \sim \mathcal{N}_N\left(\boldsymbol{0}, \boldsymbol{C}\boldsymbol{N}\boldsymbol{C}^T\right). \tag{4}$$

The matrix $\boldsymbol{C}$ consists of 0's and 1's, indicating which cells have to be aligned together. The random variable $\boldsymbol{\lambda} = \boldsymbol{C}\boldsymbol{\mu}$ is treated as the mean process for variables $\boldsymbol{Z} = \{Z_i\}_{i=1}^N$ associated with observations $\boldsymbol{z} = \{z_i\}_{i=1}^N$ of the aggregated model

$$\boldsymbol{Z}|\boldsymbol{\lambda} \sim \mathcal{N}_N\left(\boldsymbol{\lambda}, \sigma_Z^2\boldsymbol{I}_N\right), \tag{5}$$

where $\boldsymbol{I}_N$ is the $N \times N$ identity matrix. Also at this level, the underlying process $\boldsymbol{\lambda}$ is related to $\boldsymbol{Z}$ through a measurement error with variance σ_Z^2.

B. Maximum likelihood estimation

The parameters $\boldsymbol{\beta}, \sigma_Z^2, \tau^2$ and ρ are estimated with the maximum likelihood method based on the joint unconditional distribution of $\boldsymbol{Z}$

$$\boldsymbol{Z} \sim \mathcal{N}_N\left(\boldsymbol{C}\boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{M} + \boldsymbol{C}\boldsymbol{N}\boldsymbol{C}^T\right), \tag{6}$$

where $\boldsymbol{M} = \sigma_Z^2\boldsymbol{I}_N$. Next, the log likelihood function associated with (6) is formulated

$$L(\cdot) = -\frac{1}{2}\log\left|\boldsymbol{M} + \boldsymbol{C}\boldsymbol{N}\boldsymbol{C}^T\right| - \frac{N}{2}\log\left(2\pi\right)$$
$$-\frac{1}{2}\left(\boldsymbol{z} - \boldsymbol{C}\boldsymbol{X}\boldsymbol{\beta}\right)^T\left(\boldsymbol{M} + \boldsymbol{C}\boldsymbol{N}\boldsymbol{C}^T\right)^{-1}\left(\boldsymbol{z} - \boldsymbol{C}\boldsymbol{X}\boldsymbol{\beta}\right),$$

where $|\cdot|$ denotes the determinant. The analytical derivation is limited to the regression coefficients $\boldsymbol{\beta}$, and further maximisation of the profile log likelihood is performed numerically. The standard errors of parameter estimators for this model have been developed by means of the expected and observed Fisher information matrices, details of which are provided in [17].

C. Prediction in a fine grid

Regarding the missing values in a fine grid, the underlying mean process is of our primary interest. The predictors optimal in terms of the minimum mean squared error are given by $E\left(\boldsymbol{\mu}|\boldsymbol{z}\right)$. The joint distribution of $\left(\boldsymbol{\mu}, \boldsymbol{Z}\right)$ is

$$\begin{bmatrix}\boldsymbol{\mu}\\ \boldsymbol{Z}\end{bmatrix} \sim \mathcal{N}_{n+N}\left(\begin{bmatrix}\boldsymbol{X}\boldsymbol{\beta}\\ \boldsymbol{C}\boldsymbol{X}\boldsymbol{\beta}\end{bmatrix}, \begin{bmatrix}\boldsymbol{N} & \boldsymbol{N}\boldsymbol{C}^T\\ \boldsymbol{C}\boldsymbol{N} & \boldsymbol{M} + \boldsymbol{C}\boldsymbol{N}\boldsymbol{C}^T\end{bmatrix}\right).$$

The above distribution allows for full inference, yielding both the predictor

$$\widehat{E\left(\boldsymbol{\mu}|\boldsymbol{z}\right)} = \boldsymbol{X}\widehat{\boldsymbol{\beta}} + \widehat{\boldsymbol{N}}\boldsymbol{C}^T\left(\widehat{\boldsymbol{M}} + \boldsymbol{C}\widehat{\boldsymbol{N}}\boldsymbol{C}^T\right)^{-1}\left[\boldsymbol{z} - \boldsymbol{C}\boldsymbol{X}\widehat{\boldsymbol{\beta}}\right] \tag{7}$$

and its variance

$$\widehat{Var\left(\boldsymbol{\mu}|\boldsymbol{z}\right)} = \widehat{\boldsymbol{N}} - \widehat{\boldsymbol{N}}\boldsymbol{C}^T\left(\widehat{\boldsymbol{M}} + \boldsymbol{C}\widehat{\boldsymbol{N}}\boldsymbol{C}^T\right)^{-1}\boldsymbol{C}\widehat{\boldsymbol{N}}. \tag{8}$$

IV. Geostatistical Approach

In this section, we briefly review geostatistical approach, which is dedicated to modelling point-referenced data over a continuous domain. It specifies the process through a covariance function.

In the application considered, ammonia emission $Y(\boldsymbol{s})$ is the variable of area type. The point level data are obtained by dividing this variable by area (in km^2) of respective grid cell. This is a kind of approximation which expresses emissions over a unit area, or (roughly) *emission intensity*. Thus, a geostatistical model is applied to the modified process $Y_A\left(\boldsymbol{s}\right) = Y(\boldsymbol{s})/A$, where A stands for area of a 5km grid cell. We observe $\boldsymbol{Y}_A = \left(Y_A(\boldsymbol{s}_1), \ldots, Y_A(\boldsymbol{s}_n)\right)^T$ in a fine grid, and we wish to predict the variable $Y_A(\boldsymbol{s}_0)$ at a location $\boldsymbol{s}_0 \in D$ where it has not been observed, i.e. in centroids of a coarse grid.

Gaussian geostatistical models are based on two assumptions: second order stationarity and isotropy. Second order stationarity means that the process mean is constant and its covariance function depends only on the difference between locations. The process is isotropic if, additionally, the covariance depends only on distance (not direction) between two locations. Once these assumptions are met, spatial process can be modelled with parametric covariance functions. The exponential covariance function, applied in this study, is defined as,

$$cov\left(h\right) = \begin{cases}\sigma^2\exp\left(-\phi h\right) & \text{if } h > 0\\ \tau_{nug}^2 + \sigma^2 & \text{if } h = 0,\end{cases} \tag{9}$$

where h denotes the Euclidean distance between two points, τ_{nug}^2 represents the nugget effect, σ^2 is the partial sill, and ϕ denotes the effective range of the covariance. Furthermore, denote $\boldsymbol{K} = cov\left(\boldsymbol{Y}_A, \boldsymbol{Y}_A^T\right)$, $\boldsymbol{k} = cov\left\{\boldsymbol{Y}_A, Y_A(\boldsymbol{s}_0)\right\}$, and $k_0 = cov\left\{Y_A(\boldsymbol{s}_0), Y_A(\boldsymbol{s}_0)\right\}$.

With land use information available as covariates denoted $\boldsymbol{X} = (\boldsymbol{x}(\boldsymbol{s}_1), \ldots, \boldsymbol{x}(\boldsymbol{s}_n))^T$ and $\boldsymbol{x}(\boldsymbol{s}_0)$, universal kriging [18] was applied for a prediction in a fine grid. The model for a random field $Y_A(\boldsymbol{s})$ has a linear mean function and the error process $\epsilon(\boldsymbol{s})$

$$Y_A(\boldsymbol{s}) = \boldsymbol{x}(\boldsymbol{s})^T\boldsymbol{\beta} + \epsilon(\boldsymbol{s}), \tag{10}$$

where $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)^T$ is a vector of p unknown coefficients, and $\epsilon(\boldsymbol{s})$ is a zero-mean Gaussian process with the exponential covariance function given by (9).

The geostatistical prediction problem is formulated as follows. We seek a predictor $\hat{Y}_A(\boldsymbol{s}_0)$ that minimises the mean squared prediction error among the predictors satisfying two properties:

1) linearity: $\hat{Y}_A(\boldsymbol{s}_0) = \boldsymbol{\lambda}^T\boldsymbol{Y}_A$
2) unbiasedness: $E\hat{Y}_A(\boldsymbol{s}_0) = E\left(\boldsymbol{\lambda}^T\boldsymbol{Y}_A\right) = EY_A(\boldsymbol{s}_0)$ for all $\boldsymbol{\beta} \in \mathbb{R}^p$

We obtain minimisation of $E\left\{Y_A(\boldsymbol{s}_0) - \boldsymbol{\lambda}^T\boldsymbol{Y}_A\right\}^2$ subject to

$$\boldsymbol{X}^T\boldsymbol{\lambda} = \boldsymbol{x}(\boldsymbol{s}_0). \tag{11}$$

This constraint optimisation task can be solved with the method of Lagrange multipliers, see e.g. [19], [15]. Provided that matrices $\boldsymbol{K}$ and $\boldsymbol{X}^T\boldsymbol{K}^{-1}\boldsymbol{X}$ are invertible, it yields

$$\begin{aligned}\boldsymbol{\lambda} = &\left\{\boldsymbol{K}^{-1} - \boldsymbol{K}^{-1}\boldsymbol{X}\left(\boldsymbol{X}^T\boldsymbol{K}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{K}^{-1}\right\}\boldsymbol{k} \\ &+ \boldsymbol{K}^{-1}\boldsymbol{X}\left(\boldsymbol{X}^T\boldsymbol{K}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{x}\left(\boldsymbol{s}_0\right)\end{aligned} \tag{12}$$

and the best linear unbiased predictor (BLUP) of $Y_A(\boldsymbol{s}_0)$ becomes

$$\begin{aligned}\hat{Y}_A(\boldsymbol{s}_0) =& \boldsymbol{\lambda}^T\boldsymbol{Y}_A \\ =& \left\{\boldsymbol{k} + \boldsymbol{X}\left(\boldsymbol{X}^T\boldsymbol{K}^{-1}\boldsymbol{X}\right)^{-1}\left[\boldsymbol{x}\left(\boldsymbol{s}_0\right) - \boldsymbol{X}^T\boldsymbol{K}^{-1}\boldsymbol{k}\right]\right\}^T \\ &\times \boldsymbol{K}^{-1}\boldsymbol{Y}_A.\end{aligned} \tag{13}$$

The resulting mean squared error of the BLUP, called also the kriging variance, is given by

$$k_0 - \boldsymbol{k}^T\boldsymbol{K}^{-1}\boldsymbol{k} + \boldsymbol{\gamma}^T\left(\boldsymbol{X}^T\boldsymbol{K}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{\gamma}, \tag{14}$$

where $\boldsymbol{\gamma} = \boldsymbol{x}(\boldsymbol{s}_0) - \boldsymbol{X}^T\boldsymbol{K}^{-1}\boldsymbol{k}$.

Estimation of parameters has been performed using the `geoR` package from the `R` software [20].

V. Results

A. Fourfold disaggregation

This subsection presents the model testing results for disaggregation from a 10km grid.

Table I displays the maximum likelihood estimates and standard errors for all parameters. Also the statistical significance of regression coefficients is reported with the t-statistic and respective *p*-values. It should be stressed, that estimation of parameters has been performed for *emission values* in the case of CAR models, and for *emission intensity* in the case of geostatistical models (denoted GEOST). Therefore, parameter estimates are comparable only within the same class of models. As regards GEOST models, due to an optimisation procedure [20], the standard errors are available only for the ratio τ^2_{nug}/σ^2_Z.

From a visual comparison of the 5km maps with predicted values of ammonia emissions (not shown), the differences with respect to the original data cannot be easily distinguished. Instead, Fig. 2 presents scatterplots of predicted values y_i^* against observations y_i. This suggests that CAR model gives better results than GEOST. In general, CAR model provides very accurate predictions, although it tends to overestimate significantly some of small values.

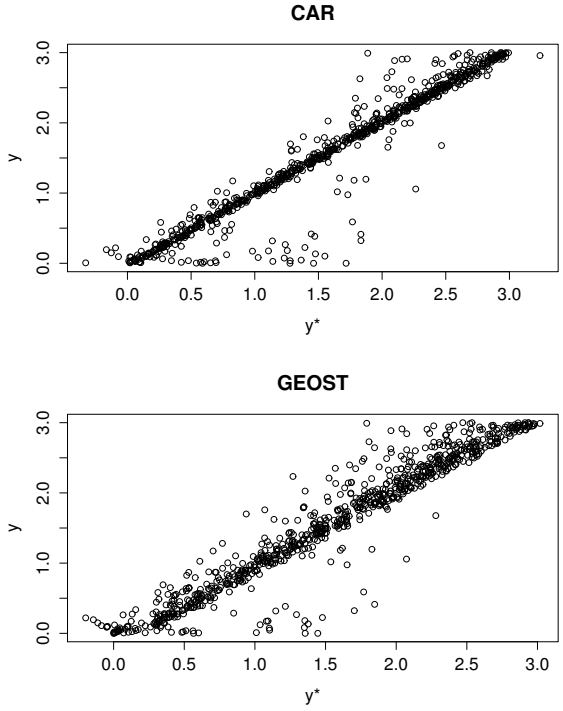


Fig. 2. Predicted (y^*) versus observed (y) values (in tonnes/year); disaggregation from 10km grid

Model residuals $(d_i = y_i - y_i^*)$ are further summarised in Table II (the upper panel). The quantitative comparison confirms that CAR model outperforms the geostatistical one, both in terms of the mean squared error (*mse*) as well as the sample correlation coefficient (*r*). Still, the highest overestimate, i.e. min(d_i) = -1.717, is reported for CAR model.

Next, the prediction standard error was calculated following the formula (8) for CAR model, and the formula (14) for geostatistical model. Since in the present case study the correct values of emissions predicted in 5km grid are known, we are in a position to compare the prediction error with actual residuals (more precisely, with their absolute values). In Fig. 3 these values are presented on maps for both disaggregation procedures. For CAR model, it is noticeable that the prediction error does not reflect diversification of actual residuals very ac-

TABLE I
MAXIMUM LIKELIHOOD ESTIMATES FOR COARSE GRIDS

	Est.	**Std.Err.**	**t-statistic**	***p*-value**	**Est.**	**Std.Err.**	**t-statistic**	***p*-value**
CAR models								
	10km				**15km**			
β_1	1.13e-07	3.26e-09	34.66	2.99e-59	1.12e-07	3.95e-09	28.26	6.37e-51
β_2	2.56e-07	1.94e-07	1.31	0.09	-	-	-	-
β_3	9.77e-08	1.19e-08	8.20	3.34e-13	1.07e-07	1.84e-08	5.83	3.11e-08
β_4	1.18e-07	2.13e-08	5.51	1.27e-07	1.24e-07	2.77e-08	4.49	9.02e-06
β_5	1.27e-07	1.32e-08	9.57	2.92e-16	1.27e-07	1.74e-08	7.31	2.84e-11
σ_Z^2	0.334	0.073	-	-	2.339	0.424	-	-
τ^2	0.536	0.082	-	-	0.214	0.088	-	-
ρ	0.948	9.98e-04	-	-	0.966	4.91e-04	-	-
GEOST models								
	10km				**15km**			
β_1	9.72e-08	5.83e-09	16.68	1.80e-31	9.21e-08	8.75e-09	10.53	2.19e-18
β_2	-	-	-	-	-	-	-	-
β_3	8.06e-08	1.62e-08	4.96	1.36e-06	-	-	-	-
β_4	9.53e-08	3.65e-08	2.61	0.005	1.21e-07	5.69e-08	2.12	0.018
β_5	1.12e-07	2.30e-08	4.88	1.91e-06	1.12e-07	3.79e-08	2.96	0.001
σ_Z^2	2.04e-03	-	-	-	4.50e-04	-	-	-
τ_{nug}^2	9.92e-05	0.07	-	-	9.84e-05	0.285	-	-
ϕ	205.01	298.41	-	-	61.02	61.39	-	-

TABLE II
ANALYSIS OF RESIDUALS

	mse	**min(d_i)**	**max(d_i)**	***r***
10km grid				
CAR	0.064	-1.717	1.104	0.961
GEOST	0.077	-1.444	1.200	0.956
15km grid				
CAR	0.136	-2.428	0.646	0.915
GEOST	0.144	-1.914	1.519	0.913

curately, and the highest values of residuals are underestimated (compare the scales of both maps). Otherwise, the prediction standard errors seem to provide a reasonable assessment. On the other hand, the prediction standard error for GEOST model is significantly underestimated, as seen from the map scales. Note, that for both disaggregation methods, the highest residuals are reported on the border of the domain; this fact is known in spatial statistics as the *edge effect*.

In addition, Fig. 4(a) presents the differences between the prediction standard errors of the models and absolute values of actual residuals. The empirical cumulative distributions of these differences confirm that the geostatistical model underestimates the prediction standard errors, much more than the CAR model does.

B. Ninefold disaggregation

Next, the results of disaggregation from a 15km grid are presented. In Table I we can see, for respective models, the increase of variances σ_Z^2 when turning from a 10km to 15km disaggregation.

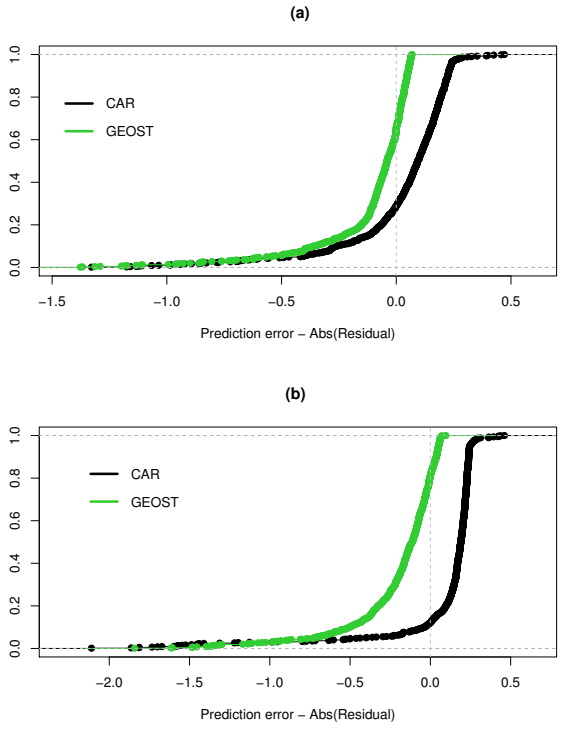


Fig. 4. Empirical cumulative distributions of the differences between the prediction standard errors and absolute values of residuals (in tonnes/year) for (a) 10km and (b) 15km disaggregation

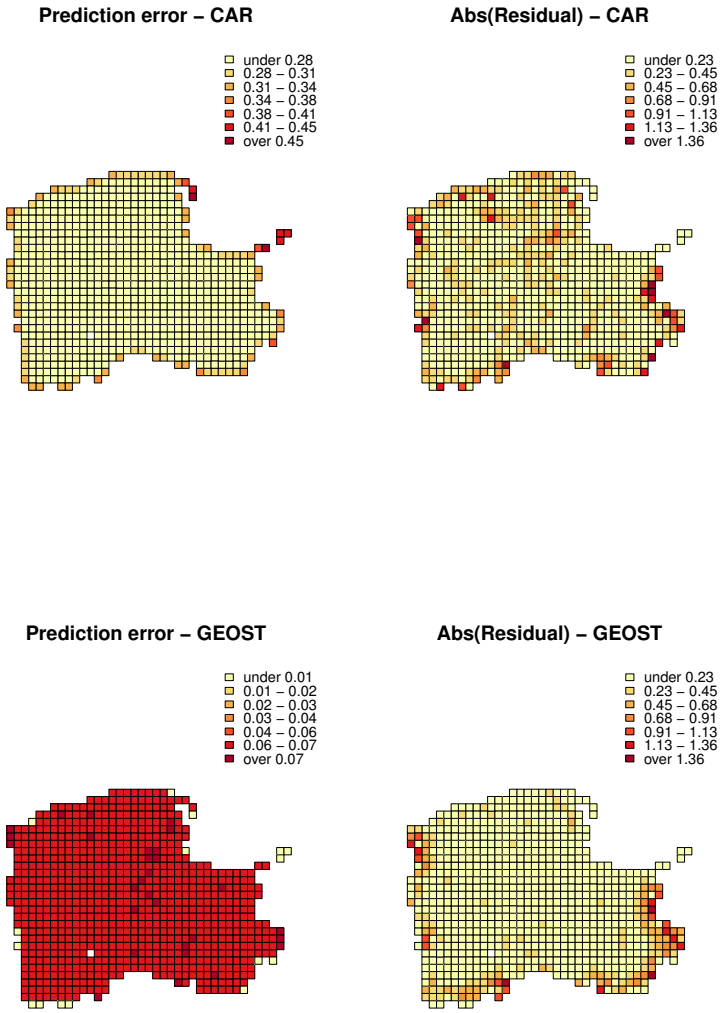


Fig. 3. Prediction standard errors and absolute values of residuals (in tonnes/year) for CAR (upper panel) and GEOST (lower panel) models; disaggregation from 10km grid. Note that the maps are drawn in different scales.

The scatterplot in Fig. 5 reveals important differences between the two methods. CAR model generally provides more accurate predictions but heavily overestimates numerous values. Predictions from GEOST are less accurate, and the model also tends to overestimate low values and underestimate high ones. Overall, Table II shows that both approaches provide comparable quality of predictions, as summarised by the mean squared error and correlation coefficient. Again, the highest overestimate, i.e. $\min(d_i)$ = -2.428, is noted for CAR model.

For the case of ninefold disaggregation, the prediction standard errors and absolute values of residuals are depicted in Fig. 6. For CAR model, this comparison provides quite good picture, although the highest values of residuals are still underestimated. Apart from this, the model uncertainty is reflected rather well. This is not the case for GEOST model. Firstly, a regular pattern on the map of prediction standard error is completely different from the actual residuals. It can be attributed to inherent features of the geostatistical method which provides the lowest prediction error at observed locations. Secondly, the prediction error for this procedure is evidently underestimated, similarly as for the case of 10km disaggregation.

Respective cumulative distributions of the differences be-

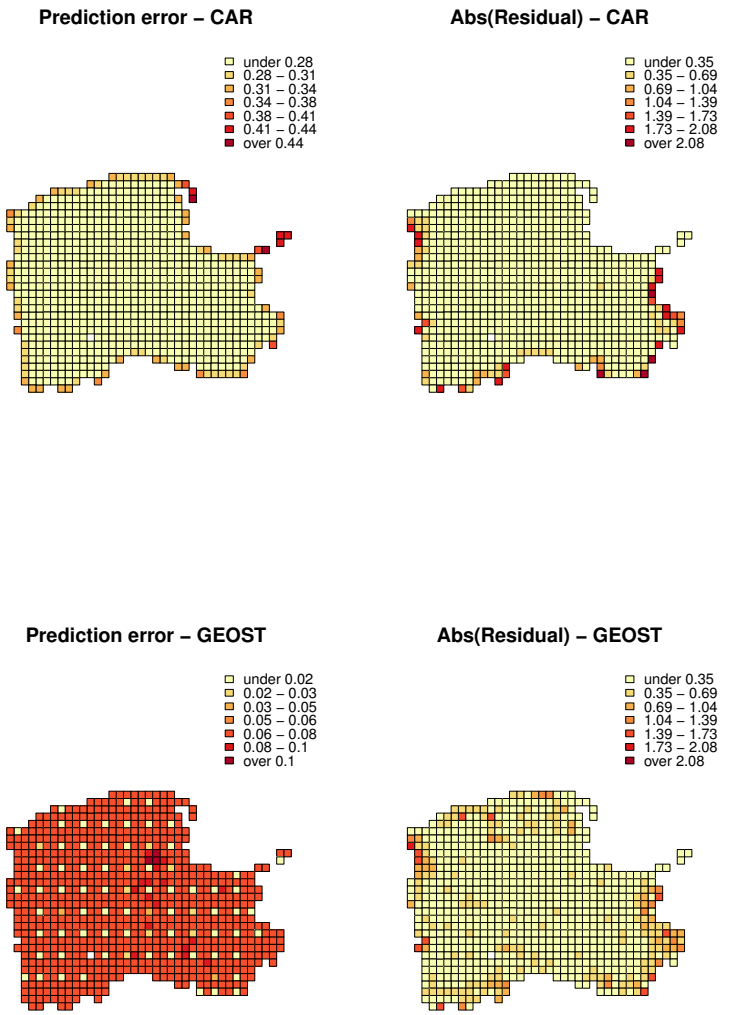


Fig. 6. Prediction standard errors and absolute values of residuals (in tonnes/year) for CAR (upper panel) and GEOST (lower panel) models; disaggregation from 15km grid. Note that the maps are drawn in different scales.

tween the prediction standard errors and absolute values of residuals, presented in Fig. 4(b), illustrate that CAR model provides rather higher estimates of error than the actual residuals, while GEOST underestimates the model error. Compared with the results for 10km disaggregation, we note that accuracy of uncertainty assessment improved for CAR model, and it declined for GEOST.

VI. Concluding Remarks

The major objective of this paper was to study uncertainty of two procedures for spatial allocation from a coarse to fine grid. For a particular disaggregation setting with proxy data available in a fine grid, we analysed the approach based on the conditional autoregressive structure, and the one based on the geostatistical methods.

For disaggregations from 10km and 15km grids to a 5km grid, both methods provided very good predictions, with r = 0.96 and 0.91, respectively. The mean squared error of predictions was approximately 20% lower for CAR model in the case of fourfold disaggregation, and 5% in the case of ninefold disaggregation.

As regards the geostatistical method, despite its good predictive performance, this disaggregation procedure failed to properly assess uncertainty of the model. It should be noted

TABLE III
PROS AND CONS OF THE DISAGGREGATION METHODS

CAR model	GEOST model
ADVANTAGES	
- Very good predictive performance - Reliable assessment of prediction error - Accuracy of uncertainty assessment remains high also when increasing a degree of disaggregation. - The method is well suited for areal data.	- Very good predictive performance - Prediction with universal kriging is a well known, popular method, and thus easy to implement for practitioners. - Wide availability of dedicated software
DISADVANTAGES	
- CAR structure is less popular among practitioners, and usually one needs to develop their own codes.	- Poor assessment of model uncertainty - The method is dedicated to point-referenced data, and application for areal data requires some additional manipulations.

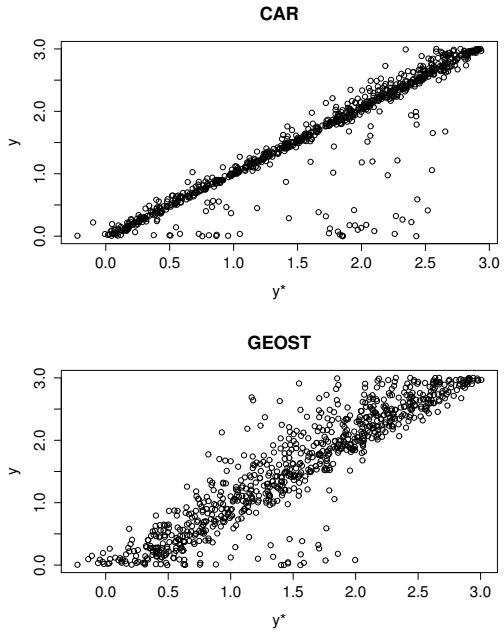


Fig. 5. Predicted (y^*) versus observed (y) values (in tonnes/year); disaggregation from 15km grid

that in our study the covariance function was unknown and respective parameters were estimated, which resulted in *empirical* universal kriging procedure. In such a case, the predictor is no longer a linear function of the data. In [16] the authors note that empirical kriging variance tends to underestimate the actual prediction error variance of the empirical universal kriging predictor because it does not account for additional error due to parameter estimation. CAR model provided a reliable assessment of prediction error. In this particular case study, this might result also from the fact that CAR structure is dedicated to areal data, like the analysed dataset of ammonia emissions.

When increasing a degree of disaggregation, obviously the quality of predictions decreases, but the accuracy of uncertainty assessment improved for CAR model. For the geostatistical approach presented, it is generally poor. Nevertheless, universal kriging is a popular, widespread technique, built into numerous software tools, which facilitates application of this disaggregation approach.

To summarise, Table III lists advantages and disadvantages of both methods for the considered case of areal data disaggregation.

ACKNOWLEDGMENT

The authors greatefully acknowledge the staff of *Ekometria sp. z o.o.*, in particular Wojciech Trapp and Małgorzata Paciorek, for provision of data.

REFERENCES

[1] K. Boychuk and R. Bun, "Regional spatial inventories (cadastres) of GHG emissions in the Energy sector: Accounting for uncertainty," *Climatic Change*, vol. 124, pp. 561–574, 2014. doi: 10.1007/s10584-013-1040-9. [Online]. Available: http://dx.doi.org/10.1007/s10584-013-1040-9

[2] R. Bun, K. Hamal, M. Gusti, and A. Bun, "Spatial GHG inventory at the regional level: accounting for uncertainty," *Climatic Change*, vol. 103, no. 1-2, pp. 227–244, 2010. doi: 10.1007/s10584-010-9907-5. [Online]. Available: http://dx.doi.org/10.1007/s10584-010-9907-5

[3] U. Dragosits, M. Sutton, C. Place, and A. Bayley, "Modelling the spatial distribution of agricultural ammonia emissions in the UK," *Environ. Pollut.*, vol. 102, no. S1, pp. 195–203, 1998. doi: 10.1016/S0269-7491(98)80033-X. [Online]. Available: http://dx.doi.org/10.1016/S0269-7491(98)80033-X

[4] H. Song, M. Fuentes, and S. Ghosh, "A comparative study of Gaussian geostatistical models and Gaussian Markov random field models," *J. Multivariate Anal.*, vol. 99, pp. 1681–1697, 2008. doi: 10.1016/j.jmva.2008.01.012. [Online]. Available: http://dx.doi.org/10.1016/j.jmva.2008.01.012

[5] T. Misselbrook, T. Van Der Weerden, B. Pain, S. Jarvis, B. Chambers, K. Smith, V. Phillips, and T. Demmers, "Ammonia emission factors for UK agriculture," *Atmos. Environ.*, vol. 34, pp. 871–880, 2000. doi: 10.1016/S1352-2310(99)00350-7. [Online]. Available: http://dx.doi.org/10.1016/S1352-2310(99)00350-7

[6] G. Velthof, C. van Bruggenb, C. Groenesteinc, B. de Haand, M. Hoogeveene, and J. Huijsmans, "A model for inventory of ammonia emissions from agriculture in the Netherlands," *Atmos. Environ.*, vol. 46, pp. 248–255, 2012. doi: 10.1016/j.atmosenv.2011.09.075. [Online]. Available: http://dx.doi.org/10.1016/j.atmosenv.2011.09.075

[7] P. Barak, B. Jobe, A. Krueger, L. Peterson, and D. Laird, "Effects of long-term soil acidification due to nitrogen fertilizer inputs in Wisconsin," *Plant Soil*, vol. 197, pp. 61–69, 1997. doi: 10.1023/A:1004297607070. [Online]. Available: http://dx.doi.org/10.1023/A:1004297607070

[8] R. Bobbink, M. Hornung, and J. Roelofs, "The effects of air-borne nitrogen pollutants on species diversity in natural and semi-natural European vegetation," *J. Ecol.*, vol. 86, pp. 717–738, 1998. doi: 10.1046/j.1365-2745.1998.8650717.x. [Online]. Available: http://dx.doi.org/10.1046/j.1365-2745.1998.8650717.x

[9] J. Erisman and M. Schaap, "The need for ammonia abatement with respect to secondary PM reductions in Europe," *Environ. Pollut.*, vol. 129, pp. 159–163, 2004. doi: 10.1016/j.envpol.2003.08.042. [Online]. Available: http://dx.doi.org/10.1016/j.envpol.2003.08.042

[10] J. Horabik and Z. Nahorski, "Improving resolution of a spatial inventory with a statistical inference approach," *Climatic Change*, vol. 124, no. 3, pp. 575–589, 2014. doi: 10.1007/s10584-013-1029-4. [Online]. Available: http://dx.doi.org/10.1007/s10584-013-1029-4

[11] P. Legendre and L. Legendre, *Numerical Ecology*, ser. Developments in Environmental Modelling. Elsevier Science, 2012. ISBN 9780444538697

[12] European Environment Agency, "Corine Land Cover 2000," http://www.eea.europa.eu/data-and-maps/data, 2000.

[13] J. Besag, "Spatial interaction and the statistical analysis of lattice systems (with discussion)," *J. Roy. Stat. Soc. B*, vol. 36, pp. 192–236, 1974. [Online]. Available: http://www.jstor.org/stable/2984812

[14] ——, "Statistical analysis of non-lattice data," *J. Roy. Stat. Soc. D-Sta.*, vol. 24, no. 3, pp. 179–195, 1975. doi: 10.2307/2987782. [Online]. Available: http://dx.doi.org/10.2307/2987782

[15] N. Cressie, *Statistics for spatial data*, ser. Wiley series in probability and mathematical statistics: Applied probability and statistics. J. Wiley, 1993. ISBN 9780471002550. [Online]. Available: http://books.google.pl/books?id=4SdRAAAAMAAJ

[16] A. Gelfand, P. Diggle, P. Guttorp, and M. Fuentes, *Handbook of Spatial Statistics*, ser. Chapman & Hall/CRC Handbooks of Modern Statistical Methods. Taylor & Francis, 2010. ISBN 9781420072884

[17] J. Horabik and Z. Nahorski, "The Cramer-Rao lower bound for the estimated parameters in a spatial disaggregation model for areal data," in *Intelligent Systems 2014*, P. Angelov, K. Atanassov, L. Doukovska, M. Hadjiski, V. Jotsov, and J. Kacprzyk, Eds. Springer International Publishing, 2015. doi: 10.1007/978-3-319-11313-5. ISBN 978-3-319-11312-8 pp. 661–668.

[18] S. Banerjee, B. Carlin, and A. Gelfand, *Hierarchical Modeling and Analysis for Spatial Data*, ser. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis, 2004. ISBN 9780203487808. [Online]. Available: http://books.google.pl/books?id=YqpZKTp-Wh0C

[19] M. Stein, *Interpolation of Spatial Data. Some Theory for Kriging*, ser. Springer Series in Statistics. Springer, New York, 1999.

[20] P. Ribeiro Jr. and P. Diggle, "geoR: a package for geostatistical analysis," *R-NEWS*, vol. 1, no. 2, pp. 15–18, 2001. [Online]. Available: http://cran.R-project.org/doc/Rnews

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 459–465
DOI: 10.15439/2016F318
ACSIS, Vol. 8. ISSN 2300-5963

Estimation of Temporal Uncertainty Structure of Greenhouse Gas Inventories for Selected EU Countries

Jolanta Jarnicka
Systems Research Institute, Polish Academy of Sciences
Newelska 6, 01-447 Warszawa, Poland;
and International Institute for Applied Systems Analysis
Schlossplatz 1, 2361 Laxenburg, Austria
Email: Jolanta.Jarnicka@ibspan.waw.pl

Zbigniew Nahorski
Systems Research Institute, Polish Academy of Sciences
Newelska 6, 01-447 Warszawa, Poland;
and Warsaw School of Information Technology
Newelska 6, 01-447 Warszawa, Poland
Email: Zbigniew.Nahorski@ibspan.waw.pl

***Abstract*—The paper addresses the problem of uncertainty in the greenhouse gas emission inventories, by proposing an alternative method for assessing uncertainty and its evolution over time. To estimate the inventory accuracy, the revisions published in consecutive years are used. These revisions are considered nonstationary time series. We describe evolution by time-dependent models, used to analyze data from the National Inventory Reports published annually up to 2015, for selected EU countries. We present a parametric model and a procedure for estimating parameters, along with the results obtained.**

I. Introduction

ACCORDING to the United Nations Framework Convention on Climate Change (UNFCCC) and its Kyoto Protocol, each of the cosignatories is obliged to provide annual data on greenhouse gas (GHG) inventory. These data are given in the National Inventory Reports (NIR), prepared either according to the 2000 IPCC report 'Good Practice Guidance and Uncertainty Management in National Greenhouse Inventories'[1], or later to the 2006 'IPCC Guidelines for National Greenhouse Gas Inventories' [2], describing in detail how uncertainty analysis should be conducted. Each NIR report contains data from a given year and revisions of past data. Data for previous years are revised given more precise information. This means that, revisions made in different years use different knowledge, and hence uncertainties in different revisions change.

In general, uncertainty associated with GHG inventory can be classified as scientific uncertainty (when the the actual emission and/or removal process is insufficiently understood) and estimation uncertainty (mostly structural, connected with activity data, emission factors, and other parameters), present whether GHG emissions are quantified. Three tier's are described for categorizing both emissions factors and activity data. A tier represents a level of methodological complexity. Tier 1 is the basic method, while Tier's 2 and 3 are each more demanding in terms of complexity and data requirements. Two of these approaches i.e. the error propagation (Tier 1) and the Monte Carlo approach (Tier 2) are recommended to assess uncertainty. The first one is much easier to calculate, while the second is considered more accurate. Since the use of the given approach is only suggested, most countries use only one of these approaches, or change the method of uncertainty assessment in consecutive reports, which makes it difficult to compare the estimates and its changes over time. In particular, it may happen that the alleged reduction in uncertainty is in fact connected with the different method of its assessment. The goal of this paper is to present an alternative, data-driven method of uncertainty assessment.

The problem of uncertainty analysis from the report data is not new, and has been dealt with for several years. Various databases were analyzed, including the IPCC data from the National Inventory Reports, but in most papers, all revision data were studied independently, e.g. in [3]. The question of how to analyze temporal evolution of the accuracy of emission inventories from several revisions was first formulated in [5], where uncertainty estimates were calculated for Austria using an algebraic approach, based on available data from different revisions year by year. Some conclusions from the results obtained there, were also presented in [4]. A similar year by year approach was presented in [8], using the Austrian NIR data as well.

This paper presents a different, revisions oriented analysis. We are interested in all consecutive yearly revisions, and differences between them, rather than in examining each of them separately. Intuitively this means that, by analyzing consecutive revisions and therefore errors and inaccuracies associated with them, which are different for each revision, we want to capture the structure of the uncertainty and its evolution over time. The method proposed combines a nonparametric regression technique using smoothing splines, as presented in [3], with a parametric model. This two-step semiparametric approach enables prior preparation of the data to which the model is fitted. Using the spline, is aimed at smoothing the data, i.e. at de-trending of the time series reported, and this in turn results in much better modelling.

978-83-60810-90-3/$25.00

This paper continues considerations outlined in [6] and carried out in [7], where we discussed some parametric model, applied to the NIR data published up to the year 2007. With the analysis carried out on longer samples, i.e. based on the NIR data published up to 2015, we managed to significantly improve the model, and get more representative results.

In Section II we present the idea of interpreting the data and propose a parametric model, that describes the uncertainty structure. Section III contains the results of fitting the model to the data on CO_2 emission from the National Inventory Reports for selected EU countries, along with the uncertainty assessment. Conclusions are given in Section IV.

II. Data and Model

We analyze data from the National Inventory Reports for selected EU countries. To consider a model, the data must first be interpreted in a manner which allows the extraction of uncertainty.

A. Data interpretation

Let $E^n_{y_j,i}$ denote the inventory data for the country i, in the year n, $n = 1, \ldots, N_j$, revised in the year y_j, $j = 1, \ldots, J$, where y_J is the last year, when the last revision was made. The index j enumerates the revisions. For a given country i, all the inventory data form a table (Table I), in which each row contains revision data reported in the year y_j i.e. a time series indexed by n, and each column contains emission inventory for the year n, recalculated in consecutive yearly revisions up to n, (a time series indexed by y_j). The analysis will be conducted for rows of that table, i.e. investigating emission inventories from consecutive yearly revisions.

For a given country i, we model any revision data to be composed of the 'real' emission, which we call the 'deterministic' fraction and the 'stochastic' fraction, related to our lack of knowledge and imprecision of observation of the real emission. We assume that the uncertainty is related to the stochastic part of the model.

For the most recently revised data, there is

$$E^n_{y_J,i} = D^n_{y_J,i} + S^n_{y_J,i}, \qquad S^n_{y_J,i} \sim \mathcal{N}(0, \sigma_{y_J,i}),$$

where E stands for the emission inventory, D for its deterministic fraction, S for the stochastic fraction, and n is the year, for which the revised data were recalculated.

Now, the data revised in the year y_j, where $j = 1, \ldots, J-1$ are modeled as having the same deterministic fraction. Thus they follow the same type of decomposition

$$E^n_{y_j,i} = D^n_{y_J,i} + S^n_{y_j,i}, \quad \text{with} \quad S^n_{y_j,i} \sim \mathcal{N}\big(0, \sigma^n_{y_j,i}\big), \quad (1)$$

where the standard deviations $\sigma^n_{y_j,i}$ are of the form

$$\sigma^n_{y_j,i} = \sqrt{\sigma^2_{y_J,i} + \alpha_{j,i}\,(y_J - y_j)^2}, \quad \alpha_{j,i} > 0. \quad (2)$$

Parameters $\alpha_{j,i}$, associated with the stochastic fraction $S^n_{y_j,i}$, can be estimated from the data together with $\sigma^2_{y_J,i}$. They describe a shift of the precision level and depend on the difference between the revision year y_j, $j = 1, \ldots, J-1$ and the most recent revision year y_J, due to the learning. The deterministic fraction $D^n_{y_J,i}$ is found, using smoothing spline, as presented in [3]. Applying this nonparametric approach to the most recently revised data $E^n_{y_J,i}$, gives not only the estimate of the deterministic fraction, but also an estimate of the variance $\sigma^2_{y_J,i}$. Considering dependence of the obtained results on spline determination, all results given in the paper are conditioned on the splines, though it is not additionally stressed in the sequel.

TABLE I
Indexing the data

$$\begin{array}{cccccccc}
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\cdots & E^n_{y_j,i} & E^{n+1}_{y_j,i} & \cdots & E^{y_j}_{y_j,i} & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\cdots & E^n_{y_J,i} & E^{n+1}_{y_J,i} & \cdots & E^{y_j}_{y_J,i} & E^{y_j+1}_{y_J,i} & \cdots & E^{y_J}_{y_J,i}
\end{array}$$

B. Model and Parameters

Given the smoothing spline $\mathrm{Sp}_{y_J,i}$, we consider it the estimate of $D^n_{y_J,i}$ and use it, along with the emission data $E^n_{y_j,i}$, for $j = 1, \ldots, J-1$ to estimate uncertainty described by $S^n_{y_J,i}$. The method is based on analysis of the differences between the revisions $E^n_{y_j,i}$, and the smoothing spline $\mathrm{Sp}_{y_J,i}$

$$v^n_{y_j,i} = E^n_{y_j,i} - \mathrm{Sp}_{y_J,i},$$

where $j = 1, \ldots, J-1$, $n = 1, \ldots, N_j$. Following (1) – (2), we assume that, for a fixed country i

$$v^n_{y_j} \sim \mathcal{N}\Big(0, \sigma^n_{y_j}\Big), \quad (3)$$

where differences $v^n_{y_j}$ are independent and

$$\sigma^n_{y_j} = \sqrt{\sigma^2_{y_J} + \alpha_j (y_J - y_j)^2} \quad (4)$$

Parameters α_j in (4) can be obtained as Maximum Likelihood estimators. Due to (3) the log-likelihood function, with parameter $\sigma^n_{y_j}$ is of the form

$$\ln L(\sigma^n_{y_j}) = -N_j \ln\sqrt{2\pi} - \frac{1}{2} N_j \ln(\sigma^n_{y_j}) - \frac{1}{2(\sigma^n_{y_j})^2} \sum (v^n_{y_j})^2$$

Substituting (4), gives

$$\ln L(\alpha_j) = -N_j \ln\sqrt{2\pi} - \frac{1}{2} N_j \ln((\sigma^n_{y_J})^2 + \alpha_j (y_J - y_j)^2) - \frac{\sum (v^n_{y_j})^2}{2((\sigma^n_{y_J})^2 + \alpha_j (y_J - y_j)^2)}$$

Then

$$\frac{d \ln L(\alpha_j)}{d\alpha_j} = -\frac{N_j (y_J - y_j)^2}{2((\sigma^n_{y_J})^2 + \alpha_j (y_J - y_j)^2)} + \frac{1}{2} \sum (v^n_{y_j})^2 \frac{(y_J - y_j)^2}{((\sigma^n_{y_J})^2 + \alpha_j (y_J - y_j)^2)^2}$$

Applying the necessary condition of extreme, we get the ML estimator of α_j, $j = 1, \ldots, J-1$

$$\widehat{\alpha}_j = \frac{1}{(y_J - y_j)^2} \left(\frac{1}{N_J} \sum (v^n_{y_j})^2 - (\sigma^n_{y_J})^2 \right). \tag{5}$$

Having obtained (5), we take

$$\alpha_j = \beta (y_J - y_j)^\gamma, \text{ where } \alpha_j > 0 \tag{6}$$

which leads us to the following **model**

$$v^n_{y_j} \sim \mathcal{N}\Big(0, \sigma^n_{y_j}\Big), \text{ where } \sigma^n_{y_j} = \sqrt{\sigma^2_{y_J} + \beta (y_J - y_j)^{\gamma+2}}. \tag{7}$$

Parameters β and γ in (7) are to be estimated by the Least Squares method, fitting (6) to the sequence $\widehat{\alpha}_j$. We put $\widetilde{\alpha}_j = \ln \widehat{\alpha}_j$, and $\widetilde{\beta} = \ln \beta$, which brings (6) to the following regression model

$$\widetilde{\alpha}_j = \widetilde{\beta} + \gamma \ln (y_J - y_j). \tag{8}$$

Parameters $\sigma^n_{y_j}$, $j = 1, \ldots, J-1$ are now obtained from (7). Dividing $\sigma^n_{y_j}$ by the smoothing spline Sp_{y_J} gives the relative uncertainty estimates of the form

$$\widehat{u}_j = \frac{\sigma^n_{y_j}}{\mathrm{Sp}_{y_J}}, \quad j = 1, \ldots, J-1. \tag{9}$$

The procedure for a fixed country i, consists of two steps.

Procedure 2.1: Assessing uncertainty.

Step 1 For the most recently revised data $E^n_{y_J}$

- find the smoothing spline Sp_{y_J}
- estimate the variance $\sigma^2_{y_J}$
- calculate the differences $v^n_{y_j} = E^n_{y_j} - \mathrm{Sp}_{y_J}, \quad j = 1, \ldots, J-1.$

Step 2 To fit the parameters in model (3) – (4)

- find $\widehat{\alpha}_j$, $j = 1, \ldots, J-1$, using (5)
- estimate $\widetilde{\beta}$ and γ in regression model (8)
- find $\sigma^n_{y_j}$, $j = 1, \ldots, J-1$, due to (7)
- find relative uncertainty estimates $\widehat{u}_j$, $j = 1, \ldots, J-1$, using (9).

III. Uncertainty assessment

We analyze UNFCCC data on CO_2 emission (in Gg) without land-use, land-use change and forestry (LULUCF), published yearly in the National Inventory Reports up to the year 2015 [9]. Calculation of emission estimates based on the measurements collected takes approximately two years, so the data reported in 2015 originate from the year 2013.

To illustrate various features of uncertainty structure we consider the data for six EU countries: Austria, Belgium, UK, Denmark, Ireland, and Finland. Each of them started to report data on GHG emission before the agreed year 2003 (on y_J=2001), conducting a test phase and providing data on emission since 1999 (Austria, UK, Ireland, and Finland) and since 2000 (Belgium and Denmark). This means we analyze the data on CO_2 emission excluding LULUCF for the year $y_J = 2013$, and all the earlier revisions, down to 1999 or 2000.

The smoothing splines Sp_{y_J}, built for the most recently revised data, i.e. a time series $E^n_{y_J}$, where $n = 1990, \ldots, 2013$, for each of the countries considered are depicted in Fig. 1. They clearly evidence sudden year-to-year changes in the inventories that are interpreted as results of errors with respect to the curves obtained by smoothing.

One can notice that, the spline fit is not the same for all countries. It seems to be much better for countries whose emissions are depicted in figures in the left-hand column of Fig. 1. Since the purpose of the spline was to extract data on the 'real' emission, this can be explained by different levels of uncertainty, reported by these countries for the year 2013. Total uncertainty, reported for 2013 in the Austrian NIR was equal 4.27%, the one for the UK 4%, and 3.45% for Ireland. In the case of Belgium and Denmark the spline obtained significantly smooths the data reported, which can be interpreted in terms of higher uncertainty – the total uncertainty reported for 2013 was equal 5.53% and 5.2% for Belgium and Denmark respectively. Fig. 1(d) can be considered a good example illustrating the problem of uncertainty assessment. Finnish total uncertainty reported for 2013 is the highest of them, estimated for 6%, although the fit may suggest much lower value. This is connected with the fact, that the uncertainty assessment in the Finnish NIR was obtained using Tier 2 (Monte Carlo approach), while the remaining ones are calculated using Tier 1. Just like in this case, the results obtained through different approaches are often difficult to compare.

Fig. 1 demonstrates also some similarities in the monotonic behaviour of emissions for the countries analyzed. It's easy to see a decreasing trend starting from 2005 (it is visible even in the case of oscillating Belgian and Danish emissions, heavily smoothed by the spline). More interesting, however, is a significant drop in emissions in 2009, associated with the economic crisis.

The estimates for variances $\sigma^2_{y_J}$, where $y_J = 2013$, calculated when building the smoothing spline are equal 2715127.62, 8091277.78, 121449538.0, 15182144.8, 650202.035, and 5940914.85 for Austria, Belgium, UK, Denmark, Ireland, and Finland respectively.

Having built the smoothing spline for $y_J = 2013$, we subtracted it from all the earlier revisions $E^n_{y_j}$, where $y_j = 1999, \ldots, 2012$ in the case of Austria, the UK, Ireland, and Finland, and $y_j = 2000, \ldots, 2012$ for Belgium and Denmark. In each case, the revision data represented a time series $E^n_{y_j}$, for $n = 1990, \ldots, y_j$.

The assumptions in model (7) were checked, by performing statistical tests. The differences obtained, were tested for normality using the Shapiro-Wilk test (considered the most reliable normality test). Since in some cases we had to deal with small samples, the results were confirmed by the Lilliefors test (modification of the Kolmogorov-Smirnov test with unknown parameters). Moreover the differences were tested for significance of true population mean, using two-sided t-test, and taking significance level 0.05.

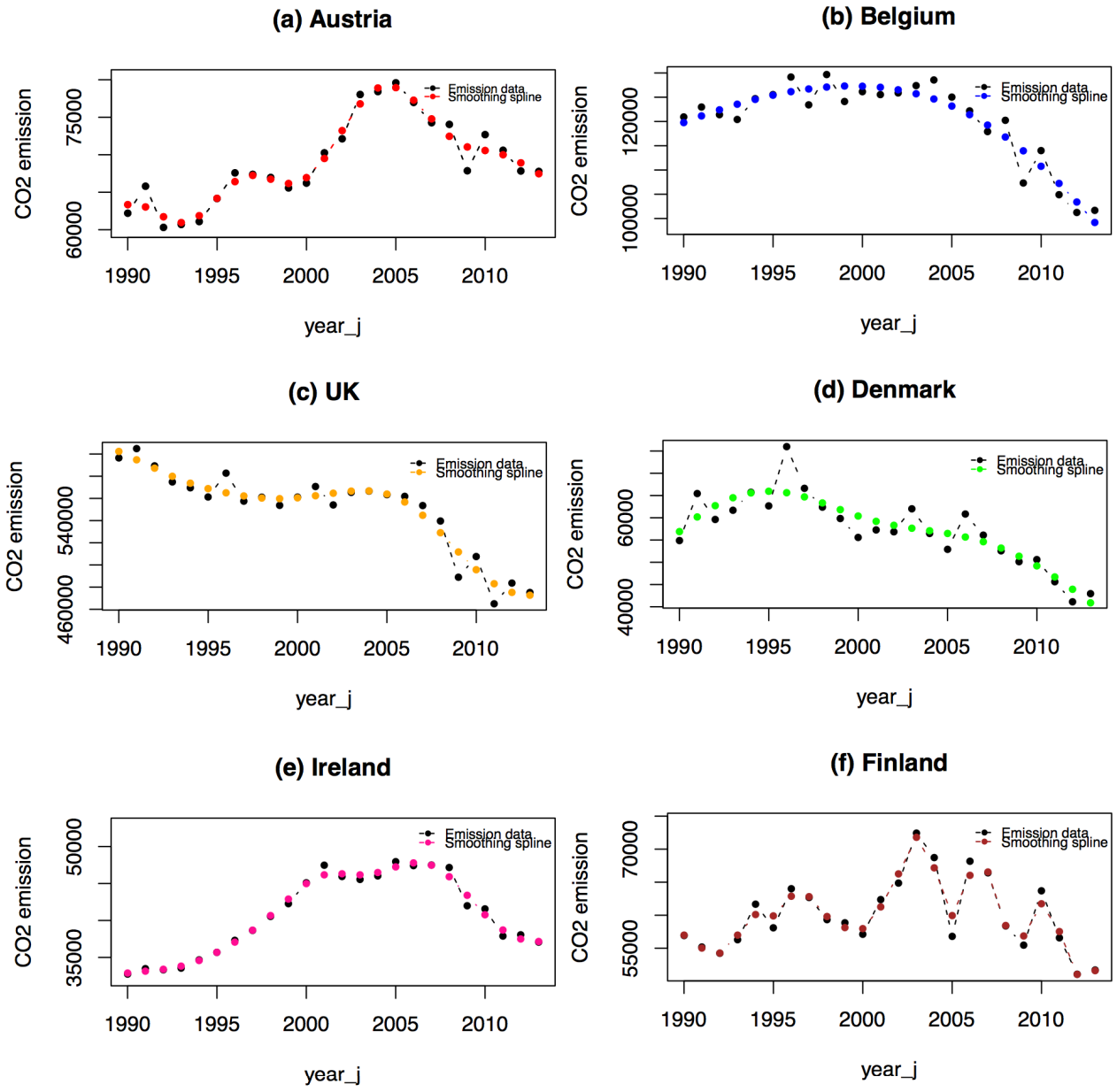


Fig. 1. Smoothing spline built for the most recent NIR data ($y_J = 2013$) on CO_2 emission [Gg CO_2], published in 2015

.

In most cases, there was no statistical evidence against the null hypothesis on normality of the data analyzed, and hence the alternative hypothesis was clearly rejected. The test failed in the case of the most initial revisions, in particular those provided in the test phase, i.e. concerning emission data on the years 1999 – 2000, which can partly be explained by the fact that the method of calculating emission revisions was being developed at that time. The t-tests performed on these differences, for which the normality assumption was met, showed that in most cases true population mean, is statistically insignificant and can be assumed zero. This means that, the assumptions taken in the model considered, were reasonable.

Following procedure 2.1, we found coefficients α_j, as ML estimators (5) and fitted parameters in regression model (8). The model fitted has been verified. All the parameters estimated turned out to be significant – the null hypotheses considering them insignificant were rejected, since p-values in the case of $\widetilde{\beta}$ ranged from $\exp(-7)$ to $\exp(-13)$ and for γ from $\exp(-4)$ to $\exp(-8)$, the same as p-values in the F-test.

Coefficient of determination R^2, indicating the goodness of fit of the model considered, was equal 84% for Austria, 77% for Belgium, 86% for the UK, 94% for Denmark, 81% for Ireland, and 97% for Finland.

The estimates of $\gamma + 2$ and $\widetilde{\beta}$ are given in Table II, along

with the results of model (8) validation.

The main result of the paper – relative uncertainties u_j, calculated due to (9) based on $\sigma^n_{y_j}$, $j = 1, \ldots, J-1$ in (7) for each of the six countries considered, are given in Table III and depicted in Fig. 3.

The temporal evolution of standard deviations $\sigma^n_{y_j}$, $j = 1, \ldots, J-1$ in (7), for all countries analyzed is shown in Fig. 2. It can be observed that, they are decreasing rather slowly in time. In turn, the corresponding relative uncertainty estimates $\widehat{u}_j$, $j = 1, \ldots, J-1$, presented in Fig. 3, for some countries, like Austria, Uk, and Ireland decrease, some others, like those for Denmark or Belgium increase slowly until 2004 – 2005. An exception is Finland, for which nonmonotonic and slightly oscillating values of u_j are shown. However, it can be observed that the uncertatinty estimates for all of the countries analyzed in the paper grow quite rapidly in later years (in particular those for Denmark and Ireland). This is connected with noticeable decrease of emissions for all countries in the years 2005 – 2012, much quicker than the slow decrease of standard deviations $\sigma^n_{y_j}$. Note that the values of $\sigma^n_{y_j}$ estimated for the UK, are much higher than for other countries, due to much higher emissions. This did not prevent us, however, in getting a good uncertainty assessment, also for that country.

It is worth stressing that, the uncertainty estimates obtained due to (9) agree quite well with the official uncertainty assessments provided in the National Inventory Reports (see Fig.4). Uncertainty assessments have become part of the 2000 IPCC Good Practice Guidance [1]. Next to the emission data, parties are expected to provide assessment of total uncertainty level of reported emission and trend uncertainty, using Tier 1 analysis (error propagation), along with the uncertainty assessment for each greenhouse gas, and each of the key IPCC categories. It is also suggested to give the results at the Tier 2 level (Monte Carlo simulation), if available. The advantage of using Tier 2 methodology is that uncertainties are taken into account and the ranking shows where uncertainties can be reduced. In the 2006 IPCC guidelines [2] it is suggested that, good practice reporting should include analysis of both Tier 1 and Tier 2.

Since applying given approach is only suggested, most countries use only one of them, i.e. Tier 1, which is easier to calculate, although considered to be less accurate. Tier 2 approach is used by Austria (starting with 2005), for Finland (for the years 2001-2005 Tier 1 analysis was not conducted, the uncertainty assessments published in the National Inventory Reports was based on Tier 2), and the UK (2003), while Belgium provides the uncertainty assessments obtained using Monte Carlo approach only for Flanders.

We compared the resulting uncertainty estimates $\widehat{u}_j$, with the reported Tier 1 trend uncertainty, and the uncertainty of CO_2 (Fig. 4). In the case of Finland, the trend uncertainty reported for 2001-2005 and 2011-2012 was calculated using Tier 2 approach, therefore in Fig. 4 both methods are considered. For the convenience of the reader, we set the same range on the vertical axis for all figures in Fig. 4, which allows for better comparison of the estimated uncertainties.

It can be seen that, the proposed uncertainty estimates correspond to those reported in the National Inventory Reports. In the case of Austria and the UK, they almost coincide with the official CO_2 uncertainty assessments, but after 2006, you may notice a slight difference in both ratings – the relative uncertainties u_j are then slightly higher than the reported ones. For Belgium and Denmark the values of u_j are pretty close to the uncertainty assessments for CO_2 and show similar monotonic behaviour. The estimates obtained for Ireland and Finland agree rather with the CO_2 uncertainty assessment, however the general monotonic behaviour is comparable with the trend uncertainty reported.

Assessments of trend uncertainty, reported since 2003 and 2004, are somewhat higher than the determined values of u_j, which may be partly explained by the fact that, they relate to total GHG emission.

The results obtained using the method proposed can therefore be considered independent confirmation of the official uncertainty estimates calculated according to [1] and [2].

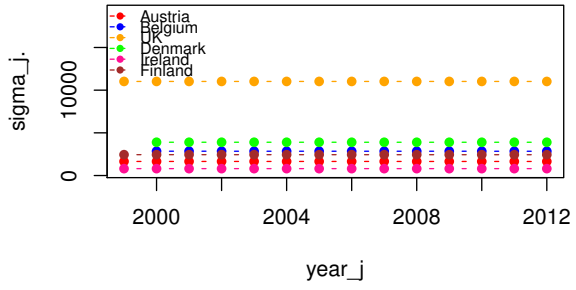


Fig. 2. Estimates of standard deviations $\sigma^n_{y_j}$ in model (7), for Austria, Belgium, UK, Denmark, Ireland, and Finland.

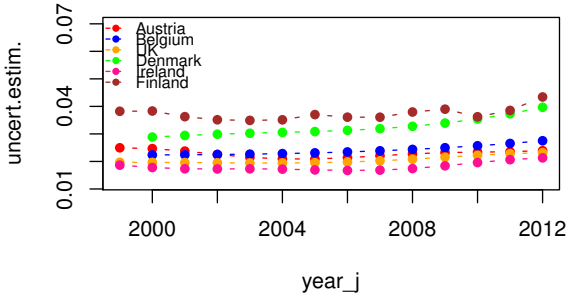


Fig. 3. Uncertainty assessment by means of $\widehat{u}_j$, for Austria, Belgium, UK, Denmark, Ireland, and Finland.

TABLE II
ESTIMATES OF $\gamma + 2$ AND $\widetilde{\beta}$ IN (8).

Model	Austria	Belgium	UK	Denmark	Ireland	Finland
$\gamma + 2$	0.60	0.91	0.59	0.62	0.48	1.02
$H_0 : \gamma = 0$ against $H_1 : \gamma \neq 0$	p-value					
	$8.48e-05$	$7.35e-05$	$6.33e-07$	$4.78e-08$	$7.75e-05$	$4.93e-11$
	reject H_0; γ significant					
$\widetilde{\beta}$	7.63	7.50	8.08	8.40	6.74	7.67
$H_0 : \widetilde{\beta} = 0$ against $H_1 : \widetilde{\beta} \neq 0$	p-value					
	$1.53e-09$	$1.60e-10$	0.00078	$1.60e-13$	$2.24e-08$	$< 2e-16$
	reject H_0; $\widetilde{\beta}$ significant					
F-test	$8.48e-05$	$7.35e-05$	0.00078	$4.78e-08$	$7.747e-05$	$4.926e-11$
R^2	0.843	0.773	0.856	0.933	0.814	0.968

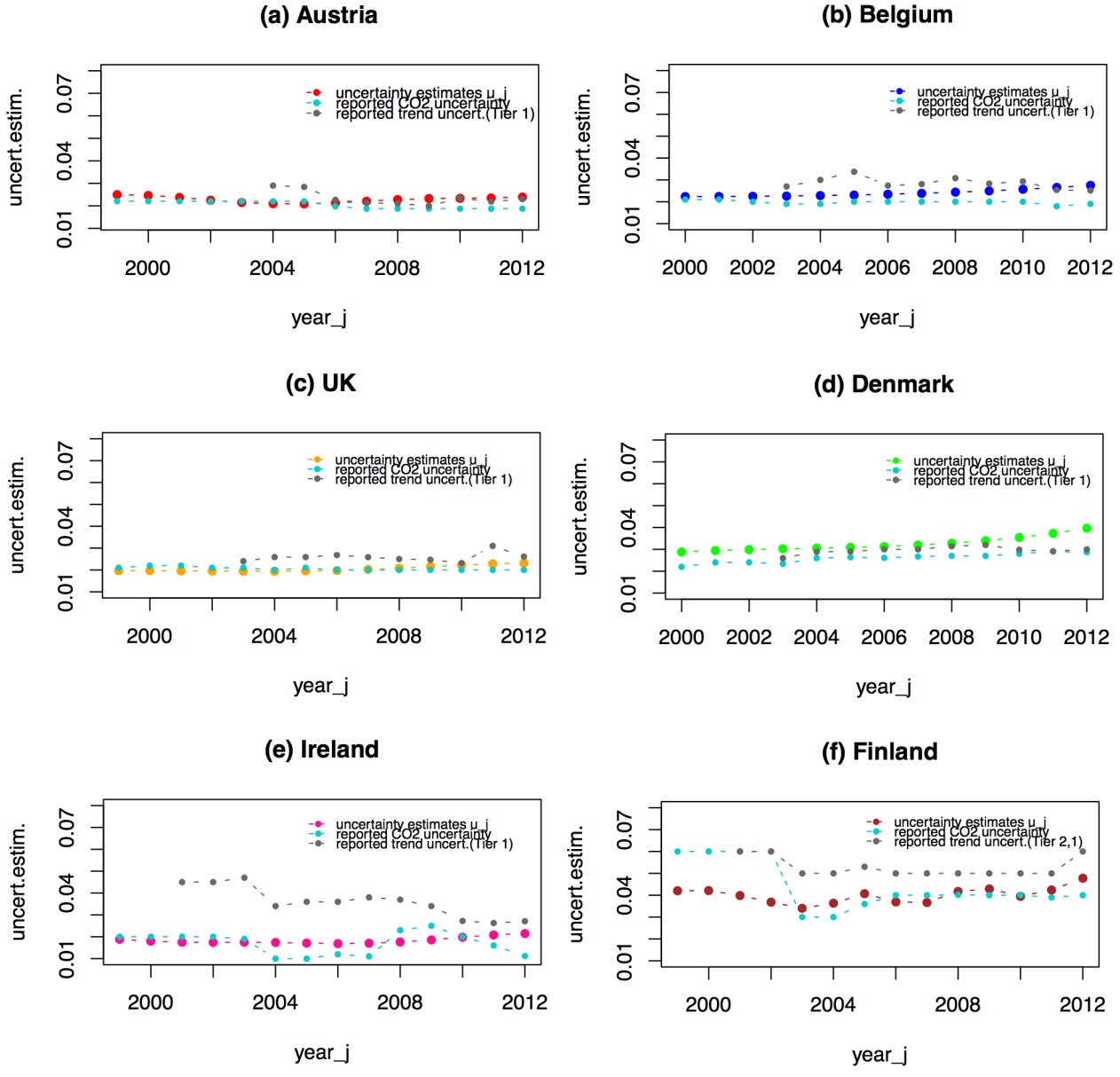


Fig. 4. Uncertainty assessment by means of relative values $\widehat{u}_j$ compared with the NIR uncertainty reported.

TABLE III
UNCERTAINTY ESTIMATES.

y_j	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Austria	0.0249	0.0247	0.0237	0.0225	0.0215	0.0209	0.0209	0.0214	0.0221	0.0228	0.0232	0.0234	0.0235	0.0239
Belgium	-	0.0224	0.0224	0.0225	0.0226	0.0228	0.0231	0.0234	0.0239	0.0244	0.0250	0.0257	0.0265	0.0275
UK	0.0197	0.0197	0.0196	0.0195	0.0195	0.0194	0.0195	0.0198	0.0202	0.0208	0.0215	0.0222	0.0228	0.0232
Denmark	-	0.0317	0.0324	0.0329	0.0333	0.0336	0.0339	0.0344	0.0351	0.0360	0.0373	0.0390	0.0411	0.0436
Ireland	0.0188	0.0179	0.0175	0.0174	0.0175	0.0173	0.0171	0.0169	0.0170	0.0175	0.0186	0.0197	0.0208	0.0215
Finland	0.0420	0.0421	0.0398	0.0368	0.0340	0.0363	0.0407	0.0369	0.0366	0.0417	0.0428	0.0395	0.0423	0.0477

IV. CONCLUSIONS

The paper deals with estimating uncertainty of the GHG inventory prepared by the countries within the Kyoto Protocol from the reported data. As opposed to conventional way of calculation of inventory error variance by processing the estimates of the activity and emission coefficients of all atom emission sources, recommended in the guidelines, the method presented uses solely official inventory data, submitted by parties to IPCC, according to common inventory protocol. The uncertainty estimates are obtained under very mild assumptions on smoothness of consecutive in time emission values for a given party. A simple model of uncertainty evolution in time is assumed. The statistical methods are used to estimate model parameters and then to calculate the uncertainty estimates. Hence, the results are obtained regardless of estimates given by parties and confirm independently the values calculated according to the methods recommended by IPCC.

The method gives quite smooth temporary curves for uncertainty estimates evolution. This effect can be induced by a model structure which was designed to catch the main directions of uncertainty evolution rather than very accurate estimation of local changes. An advantage of such model is, that the estimates obtained from our statistical estimation resemble the uncertainty estimates reported by countries as to the similar smoothness of both kinds of curves. They are also close in values to the estimates reported by countries.

The method presented is general enough to be applied to inventory data provided by other countries. Being a statistical method, its accuracy depends a lot on the number of available data points. The six countries analyzed in this paper were chosen due to their relatively long inventory sequences. It is intended to extend the calculation for other countries. Studying common properties of these models hopefully enables obtaining reliable results also for countries with shorter reported samples.

REFERENCES

[1] "Good Practice Guidance and Uncertainty Management in National Greenhouse Inventories," IPCC, 2000, available at http://www.ipcc-nggip.iges.or.jp/public/gp/english/

[2] "IPCC Guidelines for National Greenhouse Gas Inventories", IPCC, 2006, available at http://www.ipcc-nggip.iges.or.jp/public/2006gl/

[3] Z. Nahorski and W. Jęda, "Processing national CO_2 inventory emission data and their total uncertainty estimates," *Water, Air, and Soil Pollution: Focus*, vol. 7, pp. 513–527, 2007. DOI: 10.1007/s11267-006-9114-6.

[4] G. Marland, K. Hamal and M. Jonas, "How uncertain are estimates of CO_2 emissions?" *J. Industrial Ecology,* vol. 13, 2009, pp. 4–7. DOI: 10.1111/j.1530-9290.2009.00108.x

[5] K. Hamal, "Reporting GHG emissions: Change in uncertainty and its relevance for detection of emission changes," *Interim Report IR-10-003,* IIASA, Laxenburg, 2010.

[6] Z. Nahorski and J. Jarnicka, "Modeling uncertainty structure of greenhouse gases inventories", Report RB/11/2010, SRI PAS, Warsaw, 2010, unpublished.

[7] J. Jarnicka and Z. Nahorski, "A method for estimating time evolution of precision and accuracy of greenhouse gases inventories from revised reports," *in Proc. 4th Intl Workshop on Uncertainty in Atmospheric Emissions,* Kraków, Poland, 2015, pp. 97–102, available at http://www.ibspan.waw.pl/unws2015/index.php?go=publications

[8] P. Żebrowski, M. Jonas, E. Rovenskaya, "Assessing the improvement of greenhouse gases inventories: can we capture diagnostic learning?" *in Proc. 4th Intl Workshop on Uncertainty in Atmospheric Emissions,* Kraków, Poland, 2015, pp. 90–96, available at: http://www.ibspan.waw.pl/unws2015/index.php?go=publications

[9] "National Inventory Report 2003-2016 under UNFCCC Treaty", available at: http://unfccc.int/national_reports/annex_i_ghg_inventories /national_inventories_submissions/items/8812.php.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 467–474
DOI: 10.15439/2016F412
ACSIS, Vol. 8. ISSN 2300-5963

Underwater Acoustic Communications in Time-Varying Dispersive Channels

Iwona Kochańska
Gdańsk University of Technology
ul. Narutowicza 11/12,
82-233 Gdańsk, Poland
Email:
iwona.kochanska@eti.pg.gda.pl

Jan Schmidt
Gdańsk University of Technology
ul. Narutowicza 11/12,
82-233 Gdańsk, Poland
Email:
jan.schmidt@eti.pg.gda.pl

Mariusz Rudnicki
Gdańsk University of Technology
ul. Narutowicza 11/12,
82-233 Gdańsk, Poland
Email:
mariusz.rudnicki@eti.pg.gda.pl

***Abstract*—Underwater acoustic communication (UAC) system designers tend to transmit as much information as possible, per unit of time, at as low as possible error rate. However, the bit rate achieved in UAC systems is much lower than for wire or radio-communication systems. This is due to disadvantageous properties of the UAC channels, namely the sea and inland waters. Estimation of UAC channel transmission properties is possible within a limited bandwidth and temporal resolution. Thus, the UAC physical layer of data transmission is designed on the basis of roughly estimated channel parameters, or assuming the worst possible conditions. The paper presents the methodology of adapting UAC signaling schemes to tough underwater propagation conditions, through an example of two communication systems designed and developed at the Gdansk University of Technology.**

I. Introduction

THE Department of Marine Electronic Systems, the Faculty of Electronics, Telecommunications and Informatics Gdańsk University of Technology has a long and rich tradition in the processing of acoustic signals used in underwater systems. The main areas of interest are hydrolocation systems, namely sonar systems [1-6], and UAC systems [7-9]. The latter are particularly vulnerable to interference, due to tough propagation conditions, causing a time dispersion and time variability in the acoustic signals transmitting the information.

Due to a wide range in the transmission properties of UAC channels, there are, in the world, only few standards used to define very slow communication. In deep-water channels, transmission rates of up to 100 kbps can be achieved, while the same research centers offer much slower standards in shallow water channels, where reliable communication at a speed of 40-80 bps is a significant achievement. Such differences in UAC systems performances is due, *inter alia,* to the large temporal uncertainty of underwater channel transmission characteristics [10].

Furthermore, depending on the communications system in question, there are different requirements for speed and the dependability of transmissions in acoustic links. It may be a case of: a) slow but reliable transmission with autonomous underwater vehicle's (AUV) control signals, b) slightly faster but still generally reliable measurement of data transmission from an underwater monitoring system or, c) maximum speed video transmissions from underwater cameras. Therefore, there is no one, single, typical, UAC problem. The physical layer of the data transmission should be adapted to the specific propagation conditions of the particular channel.

II. Transmission properties of the UAC channel

A. Propagation conditions

The range of the UAC system is determined mainly by the absolute value of absorption attenuation, and varies in proportion to the square of the frequency of the system (except for the band between 0.5 to 5 kHz and 200 to 1000 kHz, in which it grows more slowly). Excessive differences of attenuation, due to growth in range has a limiting effect on the bandwidth of the system and reduces its throughput.

Phenomenon which strongly impacts transmission properties of the UAC channel consists of reflections from the sea-bottom and the water's surface, as well as other objects present in water. This causes multipath propagation, and, in consequence, reception of both direct and delayed signals. This phenomenon also goes hand-in-hand with strong refraction, which is caused by a significant change in sound velocity as a function of depth. Both multipath propagation as well as refraction produce time dispersion in the transmitted signal, that can be measured as a parameter, known as multipath delay spread τ_M. Moreover, due to multipath propagation phenomenon, selective fading of transmitted signal spectrum is observed. A maximum bandwidth not affected with selective fading is expressed as coherence bandwidth B_C.

The movement of the UAC system's transmitter and receiver causes the Doppler Effect, resulting in the time-domain scaling of a natural broadband communication signal. This phenomenon also has a significant impact on the communication system's performance. Because of the relatively low velocity of propagation of acoustic waves in water, which is approximately 1500 m/s, the relative Doppler shift is approx. 200,000 times higher than in case of radio-communication systems. The impact of the Doppler Effect on the received signal in UAC channels is expressed as a scal-

ing factor or – for narrowband signals – a maximum Doppler spread ν_M.

UAC channel propagation conditions can change over-time. Depending on the phenomena under consideration, the variability of transmission properties can be of a range of several months (i.e. seasons), several days and hours (i.e. tides, times of day), minutes (i.e. internal waves), a few seconds (i.e. surface waves) as well as the order of milliseconds (that is to say: reflections, scattering) [7]. In designing the UAC physical layer it is essential to determine the time and frequency ranges over which the channel can be considered as stationary. This is defined by coherence time T_C parameter [11].

B. UAC channel transmission properties measurement

The time-varying impulse response (TV-IR) $h(t,\tau)$ of the UAC channel is modeled as tapped delay line (Fig. 1). It is defined in the domain of two time variables: observation time t and delay τ. The values of t indicate the moments of subsequent IR measurements while τ denotes the position on the time axis of successive samples of the IR in a single observation.

TV-IR is the basis for computation of transmission characteristics. If TV-IR fulfills the wide sense stationary uncorrelated scattering (WSSUS) assumption, it can describe the channel statistically, providing information on transmission parameters: multipath delay spread τ_M, Doppler spread ν_M, coherence time T_C and coherence bandwidth B_C [11]. The accuracy of instantaneous transmission parameters depends on the accuracy of channel impulse response estimation [12].

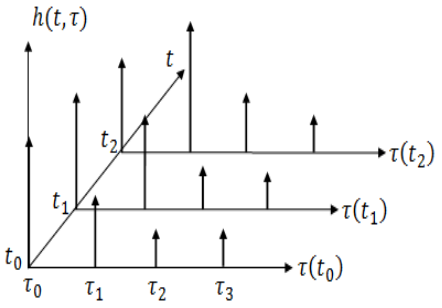


Fig 1. Tapped-delay line model of time-varying impulse response.

The measurement of impulse response is performed using the correlation method. As the probe signals, pseudorandom binary sequences (PRBSs) and linear frequency-modulated (LFM) chirps are used. A PRBS is the repetition of the maximal-length binary sequence $c_m \in \{-1,1\}$, modulated onto a binary phase-shift keyed waveform. Maximal-length binary sequences (known also as m-sequences) are widely used for communication synchronization, and transmissions of signals below the noise level [13-14]. A single "ping" of the passband PRBS of the center frequency f_C, bandwidth B, time duration T, m-sequence length M, and bit pulse shape $u(t)$ is given by:

$$p(t)=\sin(2\pi f_c t)\sum_{m=0}^{M-1} c_m u\left(t-\frac{m}{M}T\right) \quad , \qquad (1)$$

where $0\le t\le T$ and $0\le m\le M$, while a single "ping" of an LFM chirp signal is given by:

$$p(t)=\sin\left(2\pi\left[\left(f_c-\frac{B}{2}\right)t+\frac{B}{2T}t^2\right]\right) \qquad (2)$$

The channel probe signal is constructed as a concatenation of N pings:

$$s(t)=\sum_{n=0}^{N-1} p(t-nT) \qquad (3)$$

As a result of correlative measurement, time-varying impulse response $h(t,\tau)$ is obtained. The duration T and resolution T/M of a single "ping" determines range and resolution of variable τ. Moreover, the resolution of t is equal to $1/T$, while its range depends on number of "pings" and is equal to NT.

The uncertainty in time (and frequency) of probe signals is characterized by their ambiguity functions $A(\tau,\nu)$. It shows correlation filter response for a single ping as being a function of the time delay τ from the beginning of signal and frequency (Doppler) shift ν quantified for the center frequency:

$$A(\tau,\nu)=\int_{-\infty}^{\infty} p(t)\,p^*(t-\tau)\,e^{2\pi i\nu t}\,dt, \qquad (4)$$

where $0\le\tau\le T$ and $-1/(2T)<\nu<1/(2T)$. Fig. 2. shows ambiguity functions of PRBS and LFM signals of duration of $T=256\,ms$. At zero frequency shift, the ambiguity functions are the same as for PRBS and LFM, but in other areas they are very different. The PRBS offers high resolution in both delay and Doppler, but experiences clutter at frequency shift $\nu\neq 0$ [15].

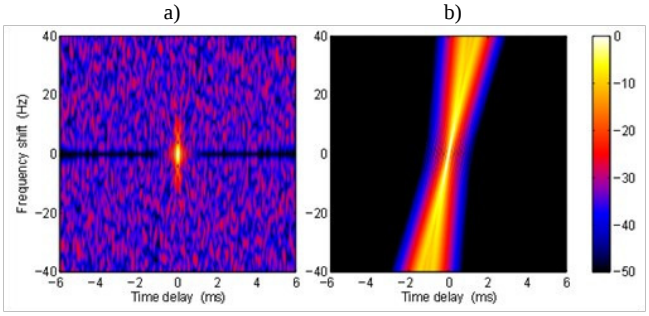


Fig 2. Ambiguity functions of PRBS (a) and LFM chirp signal (b) [15].

The chirp has a much stronger response than the PRBS for Doppler-shifted signals, but this Doppler insensitivity comes with a delay shift of $\Delta\tau=\nu\times T/B$. The LFM also shows some broadening in large frequency shifts [15].

C. The UAC impulse response

As a result of correlation measurement an estimate of TV-IR $h(t,\tau)$ is obtained. A single TV-IR estimate can be acquired but amounting to no more than the duration T of the probe sequence, which should be adapted to the geome-

try and the time variability of the channel. The channel should be sampled as often as possible to obtain a TV-IR estimate with a satisfactory time resolution. However, T is limited by the coherence time T_C of the channel. This limitation stems from the assumption that channel characteristics should remain unchanged for the duration of a single measurement in order to obtain a single estimate of the impulse response. At the same time, T should be greater than the time dispersion of the channel expressed as multipath delay spread τ_M .

Depending on the reciprocal of product of the coherence time T_C and the coherence bandwidth B_C , known as the spreading factor $SF = 1/B_C T_C$, the UAC channel can be classified as being "underspread" or "overspread".

If , the $SF < 1$ channel is assumed to be underspread. Such a channel can have a long impulse response which varies gradually, and this can be measured through the use of sufficiently long m-sequence or LMF signals. Underspread channels are also channels that change rapidly, but who possess a short TV-IR. A series of measurements through the use of short testing signals allows us to obtain knowledge on the nature of the variation in such channels.

If $SF > 1$, the channel is called overspread. In this case, the channel has a long and rapidly varying impulse response. Measurement of the TV-IR of an overspread channel is extremely difficult and gives unreliable results, if not totally impossible [12]. Measurement with the use of a short sequence allows to examine variability in the channel, but the IR will be that of visible time-aliasing [15]. On the other hand, measurement with a long testing signal will result in the collection of true information on the distribution of subsequent multipath components along the time line. However, information on channel variability will be lost.

D. WSSUS assumption

Time-varying impulse response $h(t,\tau)$ is a starting point for the construction the UAC channel model. Due to the complexity of acoustic waves' propagation mechanism, building a deterministic description is a complex task. Therefore, a statistical approach is applied, assuming (as in the radiocommunications [11]) that the channel can be described with Rayleigh distribution, making the set of the time-varying channel impulse response $h(t,\tau)$ a two-dimensional Gaussian processes with a mean of zero. With this supposition, the characteristics of the channel will be a specification of second-order statistics and the autocorrelation function of the impulse response, in addition to the IR itself, will be a subject for further analysis.

The autocorrelation of two-dimensional impulse response $h(t,\tau)$ is a four-dimensional function:

$$R_h(t,t+\Delta t,\tau,\tau+\Delta\tau) = E\left[h^*(t,\tau)h(t+\Delta t,\tau+\Delta\tau)\right] \quad (5)$$

where $-T < \Delta\tau < T$ and $-NT < \Delta t < NT$. A channel is called wide-sense stationary (WSS) if the mean value of $h(t,\tau)$ is constant and $R_h(t,t+\Delta t,\tau,\tau+\Delta\tau)$ is stationary process in t domain. Thus it can be reduce to a three-dimensional function:

$$R_h(t,t+\Delta t,\tau,\tau+\Delta\tau) \overset{WSS}{=} R_h(\Delta t,\tau,\tau+\Delta\tau) \quad (6)$$

Additionally the uncorrelated scattering (US) assumption is fulfilled, when there is no correlation between the fading coming from different signal scatters. Thus the impulse response $h(t,\tau)$ is uncorrelated in τ domain and the autocorrelation function can be denoted as:

$$R_h(t,t+\Delta t,\tau,\tau+\Delta\tau) \overset{US}{=} R_h(t,t+\Delta t,\Delta\tau) \quad (7)$$

The transfer function $H(t,f)$ of US impulse response has an autocorrelation function that is "stationary" in frequency domain f .

Under WSSUS assumption the autocorrelation function can be reduced to a two-dimensional function [16]:

$$R_h(t,t+\Delta t,\tau,\tau+\Delta\tau) \overset{WSSUS}{=} R_h(\Delta t,\Delta\tau) \quad (8)$$

The second order statistics are thus presumed to be time-invariant, and signal reflections reaching the receiver are treated as mutually uncorrelated.

In practical systems TV-IR is measured with with bandpass modulated signal, and than down-sampled into baseband. Thus, the resulting $h(t,\tau) = h_I(t,\tau) + h_Q(t,\tau)$ is complex time process, represented with real (in-phase) $h_I(t,\tau)$ and imaginary (quadrature) $h_Q(t,\tau)$ components. According to [17], for a WSS bandpass process, the in-phase and quadrature components are balanced in the sense that they have the same autocorrelation function. Also, the cross correlation of the in-phase and quadrature components must be an odd function for any pair of WSS processes. This property can be used to test whether $h(t,\tau)$ is WSS [18]. Testing US condition requires the calculation of the complex transfer function $H(t,f)$ as the Fourier transform of $h(t,\tau)$. Analysis of in-phase and quadrature components of $H(t,f)$, analogous to WSS test, is than performed.

E. Transmission Characteristics

Under the WSSUS assumption the scattering function $S(\nu,\tau)$ can be calculated as the Fourier transform in t domain. It characterizes the mean amplitude of signal reflections reaching the receiver with the delay τ and Doppler Shift ν :

$$S(\nu,\tau) = \int_{-\infty}^{\infty} h(t,\tau)e^{-j2\pi\Delta\nu\Delta t}\,d\Delta t \quad (9)$$

where $-1/2T < \nu < 1/2T$.

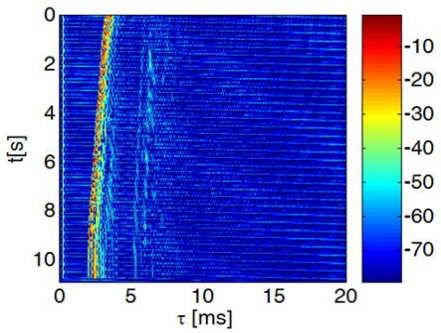


Fig 3. Time-varying Impulse response of UAC channel.

Fig. 3 and Fig. 4 show the impulse response and the scattering function of the UAC channel measured over the course of an experiment in a lake. The receiving hydrophone was placed 0.5m below the surface of the water and the speaker – 0.5m deeper. The hydrophone was moved slowly with a velocity of 15 cm/s.

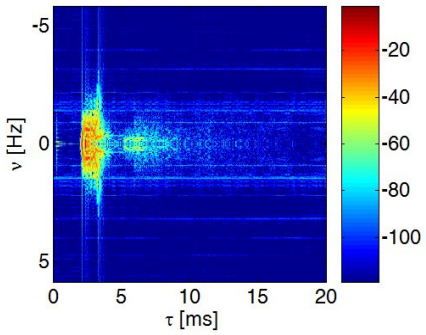


Fig 4. Scattering function of UAC channel.

As a result of integration $S(\nu,\tau)$ for the Doppler shift ν or delay τ, one of two transmission characteristics is obtained. The first is the multipath intensity profile $P(\tau)$, describing the average received power variation as a function of time delay τ (τ represents the signal's propagation delay that exceeds the delay of the first signal arrival at the receiver) [11]. For a shallow water channel, the received signal usually consists of numerous discrete multipath components (Fig. 5a). For deep water channels, due to strong refraction phenomenon, received signals are often seen as a continuum of multipath components.

For a single transmitted impulse, the time between the first and last received component represents the multipath delay spread τ_M, during which the signal power falls to some threshold level below that of the strongest component.

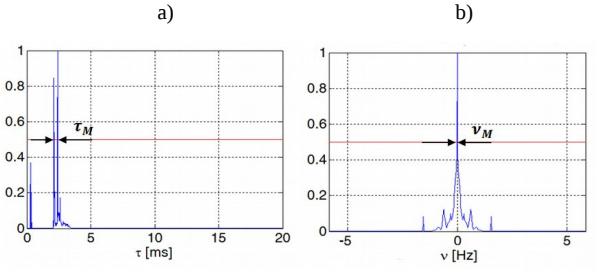


Fig 5. Multipath intensity profile (a) and Doppler power spectrum (b).

The second characteristic is the Doppler power spectral density $P(\nu)$, describing the time-variant nature of the channel. If the time-variation is random, it is seen as spectral broadening of $P(\nu)$, while a signal variation due to the transmitter or receiver movement with constant velocity reveals as a shift of $P(\nu)$ in ν domain. In case of UAC channel that fulfills the WSS assumption, $P(\nu)$ is a symmetrical function of ν. However, measurement experiments have shown, that for non-WSS channels the Doppler power spectral density can have asymmetrical shape [18]. The width of $P(\nu)$ is referred to as maximum Doppler spread, denoted by ν_M (Fig. 5b).

On the basis of autocorrelation function $R_h(\Delta t,\Delta\tau)$, the autocorrelation of the channel transfer function is calculated as:

$$R_H(\Delta t,\Delta f)=\int_{-\infty}^{\infty} R_h(\Delta t,\Delta\tau)e^{-j2\pi\Delta f\Delta\tau}d\Delta\tau \qquad (10)$$

where $-B/2<\Delta f<B/2$. It is a basis for calculating another two transmission characteristics of the communication channel. For $\Delta f=0$, the time correlation function $R_h(\Delta t)$ is obtained. Its width represents the coherence time T_C (Fig. 6a). For $\Delta t=0$ the frequency correlation function $R_h(\Delta f)$ is obtained. It represents the correlation between the channel's response to two signals as a function of the frequency difference between the two signals. It allows to determine the coherence bandwidth B_C as its width (Fig. 6b).

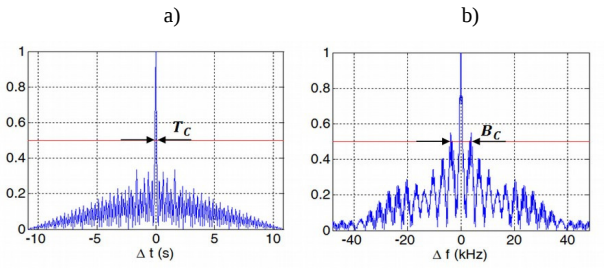


Fig 6. Time (a) and frequency (b) correlation functions.

The levels of transmission characteristics at which transmission parameters are determined depend on, amongst other factors, the technical capabilities of the particular UAC system and these are determined experimentally [11].

A set of four parameters $\{\tau_M,\nu_M,T_C,B_C\}$ is used for the designing a physical layer of the UAC system to minimize the influence of the time dispersion and time variability of the channel on the transmitted signal. The methodology is presented in the next section, in the examples of two UAC systems using different frequency diversity techniques.

F. Limitations of WSSUS assumption

In case of many underwater communication channels, especially when system terminals are in movement, the WSSUS model is of limited value. As it was shown in [18], UAC channels hardly ever fulfill the WSS and US assumptions, also in quasi-stationary conditions. But for a restricted period of time and a limited frequency range the assumptions can be satisfied. This approach is called local-sense stationary uncorrelated scattering (LSSUS) [19]. It allows to use two other transmission parameters, namely stationary time T_D and stationary bandwidth B_D, determining period of time and frequency range in which the WSSUS assumption is locally satisfied. They can extend the set of transmission parameters $\{\tau_M,\nu_M,T_C,B_C,T_D,B_D,\}$ used for the designing a physical layer of the adaptive UAC system [19].

III. Case study – OFDM System

The orthogonal frequency division multiplexing (OFDM) method is based on splitting the available transmission bandwidth into many narrow-band channels with center frequencies known as sub-carriers. The data is split into several parallel data streams, one for each sub-carrier, modulated with the use of a conventional modulation scheme (such as QAM, PSK or DPSK techniques) at a low symbol rate [11].

To adapt the OFDM signaling scheme to underwater channel conditions, the sub-carrier spacing B_{OFDM} in the frequency domain is chosen, so as to satisfy the rule:

$$B_C > B_{OFDM} \gg \nu_M \tag{11}$$

The sub-carrier spacing should be smaller than the coherence bandwidth, wherein the transfer function remains constant. It should also be much greater than the maximum Doppler spread in order to avoid interference between neighboring sub-carriers due to the Doppler shift.

In the time domain, a single OFDM symbol duration T_{OFDM} should satisfy the rule:

$$\tau_M < T_{OFDM} \ll T_C \tag{12}$$

The OFDM transmission symbol should be shorter than the coherence time T_C in which the channel's statistics can be assumed to be stationary. On the other hand, the transmission symbol should last at least as long as the corresponding signal reflections take to reach the receiver. This prevents overlapping consecutive symbols, known as inter-symbol interferences. Moreover, each OFDM symbol is preceded by a cyclic prefix, i.e. a redundant repetition of the last segment of itself. The cyclic prefix of duration T_G additionally protects the OFDM signal against inter-symbol interference [11].

Numerous sea and oceanic experimental trials of OFDM underwater communication systems have been reported [20-27]. The data transmission rate of a few kbps is achieved in shallow water channels with a depth of several hundred meters. [23]. In case of a channel with a depth of less than 100 m (and thus with much stronger multipath propagation phenomenon), only few hundred kbps are obtained [20-21]. However, there is a lack of reported results for very shallow water channels with a depth of ca. 10m.

The OFDM technique is being implemented in a laboratory model of an acoustic data transmission system, designed at the Department of Marine Electronics Systems, Faculty of Electronics, Telecommunications and Informatics, Gdansk University of Technology [28-29]. Using this model, the underwater experiments were carried out in a lake. The underwater channel was ca. 4 m deep, and experiments were performed at distances from 1 to 30 m. The speaker was placed 1 or 2 m below the surface and the hydrophone was placed 0.5, 1 or 2 m below the surface. In mobile scenarios, the hydrophone was moved slowly with a velocity of 15 cm/s. The analysis of the impulse response, measured using PRBS signal, has shown that the multipath delay spread, measured at a -10 dB threshold level of the multipath intensity profile, was ca. $\tau_M = 5\,ms$ and the maximum Doppler spread was ca. $\nu_M = 1\,Hz$. The coherence bandwidth B_C was measured as being the width of the frequency correlation function at level 0.7 of the maximum value, as well as the coherence time T_C, on the basis of the time correlation function. The results were: $B_C = 23\,Hz$ and $T_C = 1\,s$.

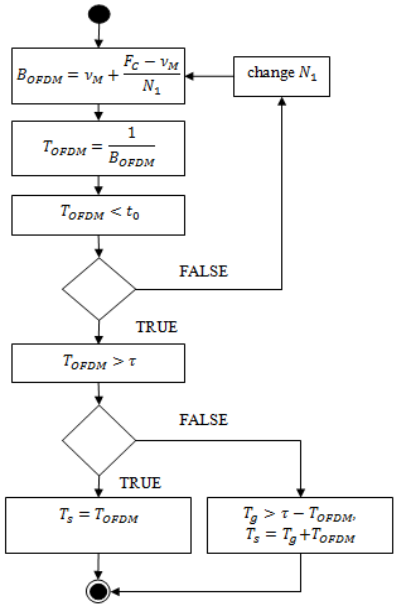


Fig 7. Procedure of calculating OFDM signaling scheme parameters.

The bandwidth of tested UAC system was 3 kHz, with a centre frequency of 5 kHz. With such fixed band and transmission parameters in the measured channel, the OFDM signaling scheme was determined as shown in Fig. 7. The sub-carrier spacing B_{OFDM} was randomly chosen from a range of (ν_M, B_C). Next, the condition was checked, that is to say, if $T_{OFDM} < T_C$, $T_{OFDM} = 1/B_{OFDM}$. If this was not the case, another B_{OFDM} value was randomly chosen. After adjusting the value of B_{OFDM} ,the following condition was checked: $T_{OFDM} > \tau_M$. If this was not the case, a cyclic prefix of the duration $T_G = T_{OFDM}/5$ was attached, lengthening the OFDM symbol duration.

With this procedure, a sub-carrier spacing of $B_{OFDM} = 5.86\,Hz$ was calculated that corresponds to symbol duration $T_{OFDM} = 0.2\,s$. Binary data, coded with a BPSK digital scheme, modulated each of the sub-carriers [18]. No channel coding was implemented. The maximum data transmission rate achieved during the underwater experiment in quasi-static scenario was 22 bps with bit error rate $BER < 10^{-3}$, and 880 bps with $BER < 10^{-1}$. In case of receiver moving in a uniform manner at a speed of about 15 cm/s, transmission rate of 615 bps was achieved with $BER < 10^{-1}$, however to perform any data transmission (<20 bps) with $BER < 10^{-3}$, it was necessary to expand a sub-carrier spacing to $B_{OFDM} = 11.72\,Hz$. This confirms the strong influence of Doppler effect on UAC system performance.

IV. Case Study – FHSS System

The second considered communication system is based on the use of spread spectrum techniques [11]. These techniques create expedient conditions for the implementation of data transmission in occurrences of harsh multipaths. Where multipath delay spread of tens of milliseconds are encountered, this excludes the use of equalization. Spread spectrum techniques were originally developed for use in military systems on account of their low probabilities of interception (LPI) and decent resistance to different types of jamming signals.

The conception of spread spectrum systems is due to the well-known Shannon equation (13), to channel capacity C, with the specified bandwidth B, and the power signal to noise ratio SNR:

$$C = B\log_2(1+SNR) \tag{13}$$

By transforming the above equations to the following form (14) we can concluded that the greater the noise that dominates the signal, the greater the possibility that it will require a wider bandwidth signal in order to receive it correctly.

$$\frac{C}{B} \simeq 1.44\,SNR \tag{14}$$

Therefore, by greatly increasing the operating bandwidth, this at the same time, allows the use of a much lower signal to noise ratio.

In general, the work of the spread spectrum system is as follows. On the transmitter's side, this is carried out by transforming an information signal into a transmission signal with a much wider bandwidth. When this is the case, this is achieved by spreading the data signal with a pseudorandom code. On the receiver's side, there must be instated a despreading operation in order to restore the transmission signal to its original bandwidth. It is important to mention the same pseudorandom code in both the receiver and in the transmitter.

There are many types of spread spectrum techniques, of which the most common are: Frequency Hopping Spread Spectrum (FHSS), Direct Sequence Spread Spectrum (DSSS), Time Hopping Spread Spectrum (THSS) and Chirp Spread Spectrum (CSS). The first of them will be briefly discussed and analyzed.

Communication systems with frequency hopping spread spectrum techniques commonly use binary frequency shift keying. In this technique, all available channel bandwidth is divided into adjacent sub-channels. The carrier among subchannels is switched by using a pseudorandom sequence. Pseudorandom sequence controls the frequency synthesizer in every signaling interval. In this technique, MFSK modulation with non-coherent demodulation is usually employed. Because of this, the use of coherent modulations is difficult to retain phase coherence during the generating process of the signal, in accordance with the hopping pattern.

In transmitters, the data signal is subjected to channel encoding and interleaving. Then goes to the input of MFSK modulator. The modulator assigns a corresponding frequency in baseband to value of sending bit with duration T_B. The produced signal is then placed in the appropriate sub-channel (frequency slot) for the time T_C - termed as "dwell time". After the amplification, the signal passes on to the channel by using the transducer.

In the receiver, the amplified signal is firstly subjected to dehopping by mixing the synthesizer output with the received signal. Then the resulting signal is demodulated into an MFSK demodulator. A synchronization signal for keeping adequate synchronism in the pseudorandom generator with frequency hopping, a received signal is extracted from a received signal. A synchronization signal between each transmitted bit is placed in the signal frame. "Coarse" synchronization is carried out based on the signal of the frame synchronization which is usually a couple of LFM signals. This signal is also used in the estimation of channel impulse response functions. Hyperbolic frequency modulation (HFM) can be used because it is more resistant to the Doppler Effect.

For the sake of the simplicity in the receiving process, and by using the fast Fourier transform (FFT), non-coherent reception is performed. Similarly, the transform is the discrete-time implementation of the corresponding matched filters. In the next step, this is subjected to an analysis of the entire operating band, in which dehopping is carried out by selecting an appropriate sub-channel.

This technique has no impact on performance in an AWGN channel. The characteristic parameter for this type of spread spectrum system is *gain processing*, which expresses the bandwidth expansion factor:

$$G_p = \frac{T_B}{T_C} \tag{15}$$

In cases in which changes of the carrier frequency are used many times during particular data bit T_B, then we are dealing with a fast frequency hopping system (FFH). The figure below shows an example of an FFH system with the frequency hopping pattern $T_B=3T_C$.

Parameters such as coherence bandwidth and coherence time are included in the following manner. The FFH technique uses frequency and time diversity to effectively counteract the effects of the multipath. Due to the fact that this system commonly works for worst-case conditions in the channel, it is necessary to determine the requirements for the resistance to the maximum multipath delay spread. The cause of this, as shown in the figure above - i.e. channel clearing time - is the length of time long enough to allow the disappearance of multipath arrivals and in correspondence with the used pseudorandom time sequence during the transmission of a single bit. Although the extension of an applied pseudorandom sequence allows for more effective counteraction in the effects of the multipath, yet this will reduce the transmission rate and requires a broadening of the operating band. However, it should be noted that an important purpose of frequency diversity is the protection of each bit against spectral nulls due to frequency-selective fading.

The employed frequencies are spaced in order to take into account the limits of the frequency change caused due to the presence of the Doppler Effect in the entirety of the available operating band. They also satisfy the condition of or-

thogonality by determining the minimum frequency separation Δf for the herein employed length of tones.

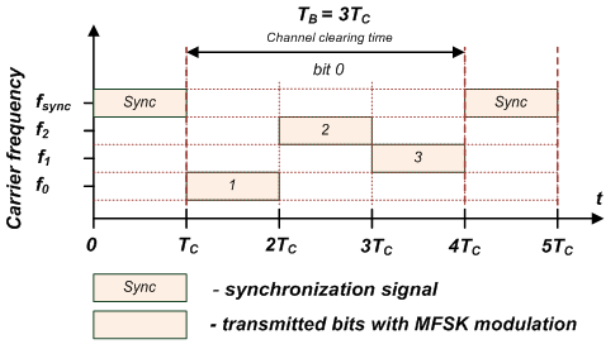


Fig 8. FFH system

The assumption concerning a communication system with this technique presented here can potentially reach maximum transmission of tens of bits per second. When the communication system takes into account the maximum value of a multipath delay spread of 30ms, then the system functioning according to the scheme shown in Fig. 8. has a value of a channel clearing time of 37.5ms, at a length Tc = 12.5ms. This yields a maximum transmission rate of 20bps for BFSK modulation, and 40bps for 4-FSK modulation, with no use of any channel coding.

Generally, a suppression mechanism in an interfering signal results from the fact that the useful signal is only transmitted at the specific moment in time, in a single narrowband channel. When the system works with the FFH technique, the interfering signal will coincide with the spectrum of the useful signal only to a small overall with its duration. The frequency hopping technique is simple to implement and is suitable in battery-powered solutions.

V. Conclusions

Due to the low propagation velocity of acoustic wave compared to electromagnetic wave, and thus greater influence of the Doppler effect, UAC systems provide transmission rates of tens or hundreds bps. The same modulation and coding techniques, applied in radiocommunication systems, allow to achieve transmission rate of Mbps and Gbps [30].

UAC channels are characterized by a large variety of propagation conditions. Designing reliable communication system requires knowledge of transmission parameters of the channel, namely multipath delay spread, Doppler spread, coherence time and coherence bandwidth. However, the possibilities of the measurement of transmission characteristics are limited, specially in case of overspread channels. New methods are needed to improve the estimation of time-varying dispersive underwater channels [12][18].

In shallow underwater channels with strong time-varying multipath propagation conditions the physical layer of data transmission is selected for the worst-case multipath delay spread and Doppler spread. The adaptation of the signaling scheme into the instantaneous transmission properties is possible only in selected well-tested channels. New modulation and coding techniques should be developed to perform reliable data transmission in time-varying propagation conditions [31].

References

[1] J. Marszal, “Digital Signal Processing Methods Implemented in Polish Navy Sonar Modernization”, *Polish Maritime Research*, Vol. 21, 2014, No 2. pp. 65-75.

[2] J. Marszal, R. Salamon, “Distance Measurement Errors in Silent FM-CW Sonar with Matched Filtering”, *Metrology and Measurement Systems*, Vol. XIX (2012), No. 2, pp. 321-332.

[3] R. Salamon, J. Marszal, W. Leśniak, “Broadband Sonar with a Cylindrical Antenna”, *ACTA ACUSTICA united with ACUSTICA*, Vol. 92, 2006, pp. 153-155.

[4] R. Salamon, J. Marszal, “Optimising the Sounding Pulse of the Rotational Directional Transmission Sonar”, *ACTA ACUSTICA united with ACUSTICA*, Vol. 88, 2002, pp. 666-669.

[5] J. Marszal, “Directivity pattern of active sonars with wideband signals”, *ACOUSTICAL IMAGING* Vol. 19, 1992, pp. 915-919.

[6] J. Marszal, R. Salamon, A. Stepnowski, “Military sonar upgrading methods developed at Gdansk University of Technology”, *Proc. IEEE Oceans'05* Europe Conference, Brest 2005, (Electronic document).

[7] I. Kochańska, “Considerations of adaptive digital communiations in underwater acoustic channel”, *Hydroacoustics*, Vol. 16, 2013, pp. 113-120.

[8] J. Schmidt, “Reliable underwater communication system for shallow coastal waters”, *Hydroacoustics*, Vol. 17, 2014, pp. 171-178.

[9] J. Schmidt, “Underwater communication system for shallow water using modified MFSK modulation”, *Hydroacoustics*, Vol. 8, 2005, pp. 179-184.

[10] R. Otnes, A. Asterjadhi, P. Casari, M. Goetz, T. Husøy, I. Nissen, K. Rimstad, P. van Walree, M. Zorzi, “Underwater Acoustic Networking Techniques”, *SpringerBriefs in Electrical and Computer Engineering*, 2012.

[11] B. Sklar, “Digital Communications: Fundamentals and Applications (2nd Edition)”, *Prentice-Hall*, 2001, pp. 944-996.

[12] I. Kochańska, “Adaptive identification of time-varying impulse response of underwater acoustic communication channel”, *Hydroacoustics*, Vol. 18, 2015, pp. 87-94.

[13] J., Marszal, R. Salamon, “Silent Sonar for Maritime Security Applications”, *Acoustical Society of America, Proceedings of Meetings on Acoustics*, Vol. 17, 070082 (2013).

[14] J. Marszal, R. Salamon, L. Kilian, “Application of Maximum Length Sequence in Silent Sonar”, *Hydroacoustics*, Vol. 15, 2012, pp. 143-152.

[15] P. van Walree, “Channel sounding for acoustic communications: techniques and shallow-water examples”, Norwegian Defence Research Establishment (FFI), FFI-rapport 2011/00007

[16] P. A. Bello, “Characterization of randomly time-variant linear channels”, *IEEE Trans*., Vol. CS-11, No 4, 1963, pp. 360–393.

[17] L. E. Franks, “Carrier and Bit Synchronization in Data Communiaction – A Tutorial Review”, *IEEE Transactions on Communications*, Vol. COM-28, No. 8, 1980 , pp. 1107 – 1121.

[18] I. Nissen, I. Kochańska, “Hydroakustik-Messung im Bornholmbecken zur lokalen Stationaritatsmodellierung beim Unterwasserschallkanal”, Fortschritte der Akustik – DAGA 2016, pp. 153-156.

[19] U. Chude-Okonkwo, R. Ngah, and T. Abd Rahman, “Time-scale domain characterization of non-WSSUS wideband channels,” *EURASIP Journal on Advances in Signal Processing*, vol. 2011, no. 1, p. 123.

[20] F. Frassati, C. Lafon, L. P.A. , and P. J.M., „Experimental assessment of OFDM and DSSS modulations for use in littoral waters underwater acoustic communications,” in *Oceans'05 Proc. MTS/IEEE*, France, 2005.

[21] S. Coatelan and A. Glavieux, „Design and test of coding OFDM system on the shallow water acoustic channel,” *in Oceans '95 Proc.. MTS/IEEE*, Brest, France, 1995.

[22] R. Bradbeer, E. Law, and E. Yeung, „Using multi-frequency modulation in a modem for the transmission of near realtime video in an underwater environment,” in *Consumer Electronics*, 2003. ICCE., Hong Kong

[23] M. Chitre, S. H. Ong, and J. Potter, „Performance of coded OFDM in vary Shallow water channels and snapping shrimp noise,” in *Ocenas '05 Proc. MTS/IEEE* , Singapore, 2005.

[24] M. Stojanovic, „Low Complexity OFDM Detector for Underwater Acoustic Channels,” in *IEEE Oceans’06 Proc. MTS/IEEE*, Boston, MA, 2006.

[25] B. Li, S. Zhou, M. Stojanovic, L. Freitag, and P. Wille, „Multicarrier Communication over Underwater Acoustic Channel with non uniform Doppler shifts," *IEEE Journal of Oceanic Engineering*, vol 33, nr 2, pp. 198-209, 2008.

[26] B. Li, S. Zhou, M. Stojanovic, and Freitag, „Pilot-tone based ZP-OFDM demodulation for an Underwater acoustic channel," in *Oceans'06 MTS/IEEE*, Boston, MA, 2006.

[27] T. Suzuki, H. M. Tran, and T. Wada, "An underwater acoustic OFDM communication system with shrimp (impulsive) noise cancelling," in *Proc. of the International Conference on Computing, Management and Telecommunications (ComManTel '14)*, pp. 152–156, IEEE, Da Nang, Vietnam, April 2014.

[28] I. Kochańska, H. Lasota, "Investigation of underwater channel time-variabiliy influence on the throughput of OFDM data transmission system", *Proceedings of Meetings on Acoustics*, POMA, 17, Acoustical Society of America, 2013.

[29] I. Kochańska, "Measurements of transmission properties of acoustic communication channels", *Hydroacoustics*, Vol. 15, 2012, pp. 91-98.

[30] A. Matoba, M. Hanada, M.W. Kim, "Throughput Improvement by Adjusting RTS Transmission Range for W-LAN Ad Hoc Network", *Proceedings of the 2014 Federated Conference on Computer Science and Information Systems*, 2014, pp. 941–946.

[31] I. Nissen, "Measurements of transmission properties of acoustic communication channels", *Hydroacoustics*, Vol. 18, 2015, pp. 113-126.

1st International Workshop on Language Technologies and Applications

DEVELOPMENT of new technologies and various intelligent systems creates new possibilities for intelligent data processing. Natural Language Processing (NLP) addresses problems of automated understanding, processing and generation of natural human languages. LTA workshop provides a venue for presenting innovative research in NLP related, but not restricted, to: computational and mathematical modeling, analysis and processing of any forms (spoken, handwritten or text) of human language and various applications in decision support systems. The LTA workshop will provide an opportunity for researchers and professionals working in the domain of NLP to discuss present and future challenges as well as potential collaboration for future progress in the field of NLP.

TOPICS

The submitted papers shall cover research and developments in all NLP aspects, such as (however this list is not exhaustive):

- computational intelligence methods applied to language & text processing
- text analysis
- language networks
- text classification
- document clustering
- various forms of text recognition
- machine translation
- intelligent text-to-speech (TTS) and speech-to-text (STT) methods
- authorship identification and verification
- author profiling
- plagiarism detection
- knowledge extraction and retrieval from text and natural language structures
- multi-modal and natural language interfaces
- sentiment analysis
- language-oriented applications and tools
- NLP applications in education
- language networks, resources and corpora

EVENT CHAIRS

- **Damaševičius, Robertas,** Kaunas University of Technology, Lithuania
- **Martinčić – Ipšić, Sanda,** University of Rijeka, Croatia
- **Napoli, Christian,** Department of Mathematics and Informatics, University of Catania, Italy
- **Wozniak, Marcin,** Institute of Mathematics, Silesian University of Technology, Poland

PROGRAM COMMITTEE

- **Boiński, Tomasz,** Faculty of ETI, Gdańsk University of Technology, Poland
- **Burdescu, Dumitru Dan,** University of Craiova, Romania
- **Cuzzocrea, Alfredo,** University of Trieste, Italy
- **Dobrišek, Simon,** University of Ljubljana, Slovenia
- **Grigonytė, Gintarė,** University of Stockholm, Sweden
- **Kapočiūtė-Dzikienė, Jurgita,** Vytautas Magnus University, Lithuania
- **Krilavičius, Tomas,** Vytautas Magnus University, Lithuania
- **Kurasova, Olga,** Vilnius University, Institute of Mathematics and Informatics, Lithuania
- **Maskeliūnas, Rytis,** Kaunas University of Technology, Lithuania
- **Meštrović, Ana,** University of Rijeka, Croatia
- **Mikelić-Preradović, Nives,** University of Zagreb, Croatia
- **Nemuraitė, Lina,** Kaunas University of Technology, Lithuania
- **Nowicki, Robert,** Institute of Computational Intelligence, Czestochowa University of Technology, Poland
- **Pappalardo, Giuseppe,** University of Catania, Italy
- **Pulvirenti, Alfredo,** University of Catania, Italy
- **Skadina, Inguna,** University of Liepaja, Latvia
- **Šnajder, Jan,** University of Zagreb, Croatia
- **Stanković, Ranka,** University of Belgrade, Serbia
- **Starczewski, Janusz,** Czestochowa University of Technology, Poland
- **Szymański, Julian,** Gdansk University of Technology, Poland
- **Tramontana, Emiliano,** University of Catania, Italy
- **Vintar, Špela,** University of Ljubljana, Slovenia

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 477–482
DOI: 10.15439/2016F153
ACSIS, Vol. 8. ISSN 2300-5963

First Automatic Fongbe Continuous Speech Recognition System: Development of Acoustic Models and Language Models

Fréjus A. A. LAleye*‡, Laurent Besacier†, Eugène C. Ezin‡ and Cina Motamed*
*Laboratoire d'Informatique Signal et Image de la Côte d'Opale
Université du Littoral Côte d'Opale, France.
50 rue F. Buisson
BP 719, 62228 Calais Cedex
Email: {laleye, motamed}@lisic.univ-littoral.fr
†Laboratoire LIG- Univ. Grenoble Alpes - BP 53,
Email: laurent.besacier@imag.fr
‡Unité de Recherche en Informatique et Sciences Appliquées
Institut de Mathématiques et de Sciences Physiques
Université d'Abomey-Calavi Bénin,
BP 613 Porto-Novo.
Email: {frejus.laleye, eugene.ezin}@imsp-uac.org

***Abstract*—This paper reports our efforts toward an ASR system for a new under-resourced language (Fongbe). The aim of this work is to build acoustic models and language models for continuous speech decoding in Fongbe. The problem encountered with Fongbe (an African language spoken especially in Benin, Togo, and Nigeria) is that it does not have any language resources for an ASR system. As part of this work, we have first collected Fongbe text and speech corpora that are described in the following sections. Acoustic modeling has been worked out at a graphemic level and language modeling has provided two language models for performance comparison purposes. We also performed a vowel simplification by removing tones diacritics in order to investigate their impact on the language models.**

I. Introduction

AUTOMATIC Speech Recognition (ASR) is a technology that allows a computer to identify the words spoken by a person in microphone. Speech recognition technology is changing the way information is accessed, tasks are accomplished and business is done. The growth of speech applications over the past years has been remarkable [1]. ASR applications have been successfully achieved for most western languages such as English, French, Italian etc., for Asian languages such as Chinese, Japanese, Indian etc, because of the large quantity and the availability of linguistic resources of these languages [2]. This technology is less prevalent in Africa despite its 2,000 languages because of lack or unavailability of these resources for most African languages (vernacular for most). Also, for the most of the time, these are not written languages (no formal grammar, limited number of dictionaries, few linguists). Despite the shortcomings, some have been investigated and now have the linguistic resources to build a speech recognition systems. For example, in the context of a project entitled ALFFA[1], the authors in [3] developed ASR systems for 4 sub-saharan african languages (Swahili, Hausa, Amharic and Wolof). Another language of West Africa (Yoruba) spoken mainly in Nigeria, in Benin and neighboring countries has been also investigated for an ASR system. [4] provides a brief review of research progress on Yorúbà Automatic Speech Recognition.

Our main objective in this paper is to introduce a first ASR system for an under-resourced language, Fongbe. Fongbe is a vernacular language spoken primarily in Benin, by more than 50% of the population, in Togo and in Nigeria. It's an under-resourced because it lacks linguistics resources (speech corpus and text data) and very few websites provide textual data. Building these resources, acoustic models and language models for Fongbe ASR becomes a challenging task. For this, we used Kaldi toolkit[2] that has allowed us to train our acoustic models on speech data that we have collected ourselves. For the language modeling, we used SRILM toolkit[3] to built trigram language models that we trained on collected text data. To enhance performance of our ASR, we subsequently transformed the vowels by normalizing different tones of Fongbe. Experiments have shown a significant improvement in the results given by the world error rates (WER).

The remainder of this paper is organized as follows. The next section describes the target under-resourced language that is Fongbe. Section 3 describes how text and speech corpora have been collected. Section 4 and 5 focus respectively on language modeling and acoustic modeling. Section 6 presents

[1]http://alffa.imag.fr
[2]kaldi.sourceforge.net/
[3]www.speech.sri.com/projects/srilm/

978-83-60810-90-3/$25.00

and comments the experimental results of WER that we obtained. Section 7 concludes this paper and presents future work.

II. Description of Fongbe language

Fongbe language is the majority language of Benin, which is spoken by more than 50% of Benin's population, including 8 million speakers and also spoken in Nigeria and Togo. The Fongbe people are the largest ethnic group in Benin. Fongbe is part of the Gbe dialect cluster and is spoken mainly in Benin [6]. It is quite widespread in the media and is used in schools, including adult literacy. The Fongbe group is one of the five Gbe dialect. J. Greenberg classifies Fongbe in the Kwa languages group in the Niger-Congo branch of the large family Niger-Kordofan [5]. It is written officially in Benin with an alphabet derived from the Latin writting since 1975. It has a complex tonal system, with two lexical tones, high and low, which may be modified by means of tonal processes to drive three further phonetic tones: rising low-high, falling high-low and mid [6]. The use of diacritical marks to transcribe the different tones of the language is essential even if they are not always marked since Fongbe is originally a spoken language. The Fongbe's vowel system is well suited to the vocalic timbre as it was designed by the first Phoneticians. It includes twelve timbres: 7 oral vowels with 4 degrees of aperture and 5 nasal vowels with 3 degrees of aperture. Its consonant system includes 22 phonemes.

Scientific studies on the Fongbe started in 1963 with the publication of Fongbe-French dictionary [7]. Since 1976, several linguists have worked on the language and many papers were published on the linguistic aspects of Fongbe. Unlike most of the western languages (English, French, Spanish, etc) and some Asian languages (Chinese, Japanese, etc) and African (Wolof, Swahili, shrugged, etc.) the Fongbe language suffers from a very significant lack of linguistic resources in digital form (text corpus and speech) despite the many linguistic works (phonology, lexicon and syntax).

III. Collection of language resources

The development of automatic continuous speech recognition system is made from a large amount of data which must contain both speech signals (for the acoustic modeling of the system) and also text data (for the language model of the system). It becomes a challenge and very difficult when it is an under-resourced language that still doesn't possess these digital resources. In this section we describe the methodology used to collect texts and audio signals of Fongbe language for building of the recognition system.

A. Speech corpus

As an audio corpus is not available for Fongbe, we proceeded to the speech signals collection to build the audio data for the system. We thus conducted the tedious task of recording the texts pronounced by native speakers (including 8 women and 20 men) of Fongbe in a noiseless environment. We have recorded at 16Khz 28 native speakers who have spoken around 1500 phrases (from daily living) grouped into 3 categories. A category is read by several speakers and contains texts that are different from contents of other categories. These recordings were made with an android application referred to as LigAikuma [8] which is developed by GETALP group of Grenoble's Computer Science Laboratory. Overall, there are around 10 hours of speech data that have been collected. First, we split the data by categories leading to a first configuration FC1: 2 categories for training (8 hours) and 1 category for testing (2 hours). Next, we split the data by speakers leading to a second configuration FC2: 20 speakers (8 hours) for training and 8 speakers (2 hours) for testing. We split the data this way firstly to make sure that category appear in test data will not appear in training and secondly, to reduce the chance of having speakers overlapping between training and testing.

TABLE I
Contents of Fongbe speech corpus.

	Speech segments	Phrases	Duration	Categories	Speakers
FC1 - config					
Train data	8,234	879	7h 35mn	C2 & C3	25
Test data	2,168	542	1h 45mn	C1	4
FC2 - config					
Train data	8,651	1,421	8h	C1, C2 & C3	21
Test data	1,751	1,410	2h	C1, C2 & C3	7

B. Text corpus

To build a language model we need to have a text corpus containing thousands of words of the given language. The standard way most commonly used to build a text corpus is the collection of texts from websites. As we have shown in previous sections, Fongbe is an under-resourced language and thus has a very limited number of websites compared to languages such as Wolof, Hausa, and above all Arabic, French and English that have a very large wide coverage on the internet and do not suffer from lack of textual data. So, based on the few websites that provide texts in Fongbe, we used RLAT [9] to crawl text from these websites covering few texts from everyday life and many texts of the Bible translated into Fongbe. RLAT enables us to crawl text from a given webpage with different link depths. For improving the quantity of texts obtained from HTML links of websites, we have added to our corpus some texts obtained from PDF files that cover many of Fongbe citations, songs and the Universal Declaration of Human Rights. After extracting all text content in web pages and pdf file, we conducted to a cleaning and normalization of the texts:

1) remove all HTML tags and codes,
2) remove empty lines and punctuations,
3) conversion of texts to Unicode,
4) remove pages and lines from other languages than Fongbe,

5) transcription of special characters and numbers,
6) delete duplicate lines.

In total, we obtained nearly 10,130 words to build our vocabulary dictionary and a corpus which contains 34,653 sentences collected from the few documents written in Fongbe that are actually available. In table II, we list the websites used to extract text for two language models (LM1 and LM2) and from which we selected 1,500 utterances (source 1) for recording speech data for the training and testing set.

TABLE II
CONTENTS OF TEXT CORPUS.

Source	Websites	Text	utterances
1	http://www.fonbe.fr	variety of texts in daily life	1,500
2	http://unicode.org/udhr/d/udhr_fon.txt	Universal Declaration of Human Rights	92
3	http://ipedef-fongbe.org/	Educational texts, songs and tales	2,200
4	http://www.vodoo-beninbrazil.org/fon.html	Educational Texts	1,055
5	https://www.bible.com/fr/bible/813/dan	The Bible	29,806

IV. LANGUAGE MODELING

Statistical language models (LM) are employed in various natural language processing applications, such as machine translation, information retrieval or automatic speech recognition. they describe relations between words (or other tokens), thus enabling to choose most probable sequences. This proves to be especially useful in speech recognition, where acoustical models usually produce a number of hypotheses, and re-ranking them according to a language model can substantially improve recognition rates [10] To compare the performance of our Fongbe recognition system, we built two language models (LM) using the same text corpus. The first language model (LM1) is built with the original texts after normalization and contain different tonal vowels. The use of tonal vowels implies that the system has to handle 26 vowels (with accented characters) considered as different tones instead of the 12 initial vowels. The second language model is built with the original texts that we modified by performing a second normalization on different tonal vowels from text corpus. The normalization was made by removing the tones from vowels and replacing accented characters by single characters. The result is that we have new entries with their transcriptions in our vocabulary dictionary. For example, the original word *axɔ́sú*, which means *king* will become in the dictionary *axɔsu*. Table III summarizes the various changes made to the vowels.

We used SRILM toolkit to train the two languages models. LM1 and LM2 were trained on 995,338 words (10,095 unigrams) by using the training data from text corpus (1,054,724 words, 33,153 sentences) without utterances used for the speech corpus (5,490 words and 1,500 sentences removed). LM1 was trained with the original texts while LM2 was trained with the modified texts by vowel normalization. To represent the uncertainty of our language models, we calculate the perplexity values of all the utterance transcriptions from speech corpus that are not contained in the various text corpus and which represents our test data to evaluate the performance of the two language models. Table V shows the perplexity values. The vowel normalization after the original text modification has positive impact on the quality of the language model by reducing in the OOV from 9.1% to 4.96%. This leads to observe a significant perplexity improvement with LM2 compared to LM1. Final system has been built using a lexicon which contains 10,130 unique grapheme words. As in [12], [11], we used grapheme as modeling unit to create our own lexicon because. An example of its content obtained after text pre-processing is shown in Table IV.

TABLE III
VOWEL NORMALIZATION.

Tonal vowels	Normalization
á	/a/
à	/a/
ā	/a/
ó	/o/
ò	/o/
ō	/o/
é	/e/
è	/e/
ē	/e/
ú	/u/
ù	/u/
ū	/u/
í	/i/
ì	/i/
ī	/i/
ɛ́	/ɛ/
ɛ̀	/ɛ/
ɛ̄	/ɛ/
ɔ́	/ɔ/
ɔ̀	/ɔ/
ɔ̄	/ɔ/

TABLE IV
EXAMPLE OF LEXICON'S CONTENT

	Word	Graphemes
Original text	axɔ́súɖuɖu hãgbɛ́	a x ɔ́ s ú ɖ u ɖ u h ã g b ɛ́
Vowel normalization	axɔsuɖuɖu haagbɛ	a x ɔ s u ɖ u ɖ u h a a g b ɛ

TABLE V
LANGUAGE MODEL COMPARISON USING THE PERPLEXITY.

LM	Vocab (words)	OOV	PPL
LM1	10,130	9.1%	591
LM2	8,244	4.96%	138

V. ACOUSTIC MODELING

In this section, we describe the methods that we used for training and testing our 2 configurations (FC1 and FC2) and

present in the next section the obtained results. The recordings and their transcriptions are used for acoustic modeling. The Acoustics models (AMs) are trained and tested on acoustic data from both FC1 and FC2 by using Kaldi acoustic modeling scripts that we have adapted to produce Kaldi scripts for Fongbe. We not only explored AM training methods but also experimented the impact of presence of tones in the utterances transcription from speech corpus by using LM1 (with tones) or LM2 (no tones). Thus, FC1 and FC2 training are performed not only with the same scripts but also by using both pronunciation dictionary. The pronunciation dictionary based grapheme that is used with LM1 contains 49 graphemes while the dictionary used with LM2 contains 28 graphemes.

The models are trained with 13 MFCC (Mel-Frequency Cesptral Coefficients) features whose coefficients are tripled with the $\Delta+\Delta\Delta$ by computing the first and second derivatives from MFCC coefficients. We also computed other feature transformation techniques such as LDA (Linear Discriminant Analysis) and MLLT (Maximum Likelihood Linear Transform) which gain substantial improvement over $\Delta+\Delta\Delta$ transformation. Subsequently, we also applied speaker Adaptation with feature-space Maximum Likelihood Linear Regression (fMLLR). Refer to the papers [13] and [14] for details on the theory of these transformation techniques implemented in Kaldi ASR. Figure 1 and Table VI show the hierarchy of the acoustics models that we trained in our experiments. In this hierarchy, we started by training monophone model using the MFCC features and we ended up training of SGMM using fMMI transformed features. The intermediate triphone models are also trained as shown in Figure 1. For decoding, we used the different trained acoustics models with the utterances from the test data. For each trained acoustic model we used the same speech parametrization and feature transformation method as was used for the given acoustic model at training time.

TABLE VI
ACOUSTICS MODELS. COMBINE* REPLACED COMBINE_TRI3B_FMMI_INDIRECT_SGMM2_5B2_MMI_B0.1

Training method	Script
Monophone	mono
Triphone	tri1
$\Delta + \Delta\Delta$	tri2a
LDA + MLLT	tri2b
LDA + MLLT + SAT + FMLLR	tri3b
LDA + MLLT + SAT + FMLLR + fMMI	tri3b_fmmi_a
LDA + MLLT + SAT + FMLLR + MMI	tri3b_mmi_b0.1
LDA + MLLT + SAT + FMLLR + fMMI + MMI	tri3b_fmmi_indirect
LDA + MLLT + SGMM	sgmm2_5b2
LDA + MLLT + SGMM + MMI	sgmm2_5b2_mmi_b0.1
LDA + MLLT + SGMM + fMMI + MMI	combine*

VI. EXPERIMENTAL RESULTS

The experiments focus on comparing the quality of ASR hypothesis measured by WER on AMs trained by different methods. To obtain the best path, we followed the standard KALDI procedures and report the best WER. The experiments were performed first on LM1 built with the original texts and using both speech data configurations. Then we conducted experiments based on the same procedures on LM2 including texts without diacritics. The interest is to measure the impact of using diacritics in language modelling from the results given by the WER. We also showed how the data speech configuration influence the quality of AMs measured by WER.

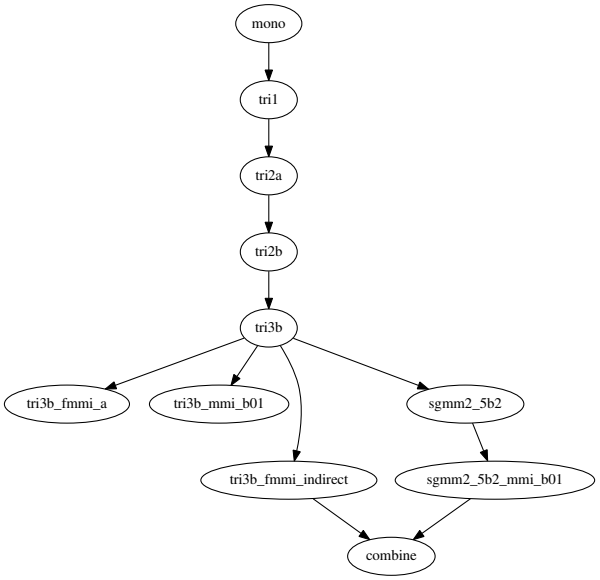


Fig. 1. Hierarchy of trained 2coustics models

A. Results before vowel normalization

In this subsection we present the results of different acoustic training methods according to data speech configuration. Table VII presents AMs results for LM1.

From the results in table VII, we can see that the monophone AM has the worst WER while the best performances are achieved with the sgmm2_5b2 AM for FC1-config and the sgmm2_5b2_mmi for FC2-config. We can thus notice that the monophone AM is typically used for the initialization of triphone models. The quality of speech recognition varies according to the used discriminative training method. The LDA+MLLT is more effective feature transformation than using $\Delta + \Delta\Delta$ features. There are subtle performance differences among the discriminatively trained acoustic model. The WER on both speech data configuration for fixed LM1 is around 44%. This can be explained by the complexity of Fongbe language for modelling the diacritics and the quality of language model used (LM1). The perplexity reported in table V justifies this assertion. Figure VI-A shows the curve performances of the acoustic training methods for both speech data configuration.

B. Results after vowel normalization

Table VIII presents two WERs of different acoustic training methods according to data speech configuration. In the second

TABLE VII
WER OF LM1-BASED ASR (WITH DIACRITICS) FOR DIFFERENT TRAINING MONOPHONE AND TRIPHONE METHODS

Speech data config/ method	WER %
FC1-config	
Monophone (a)	69.44
Triphone (b)	69.13
$\Delta + \Delta\Delta$ (c)	70.21
LDA + MLLT (d)	65.7
LDA + MLLT + SAT + FMLLR (e)	54.96
LDA + MLLT + SAT + FMLLR + fMMI (f)	55.36
LDA + MLLT + SAT + FMLLR + MMI (g)	51.11
LDA + MLLT + SAT + FMLLR + fMMI + MMI (h)	55.60
LDA + MLLT + SGMM (i)	**44.04**
LDA + MLLT + SGMM + MMI (j)	47.11
LDA + MLLT + SGMM + fMMI + MMI (k)	49.83
FC2-config	
Monophone (a)	71.97
Triphone (b)	60.37
$\Delta + \Delta\Delta$ (c)	59.74
LDA + MLLT (d)	57.52
LDA + MLLT + SAT + FMLLR (e)	51.47
LDA + MLLT + SAT + FMLLR + fMMI (f)	53.06
LDA + MLLT + SAT + FMLLR + MMI (g)	52.75
LDA + MLLT + SAT + FMLLR + fMMI + MMI (h)	52.37
LDA + MLLT + SGMM (i)	49.85
LDA + MLLT + SGMM + MMI (j)	**44.09**
LDA + MLLT + SGMM + fMMI + MMI (k)	44.17

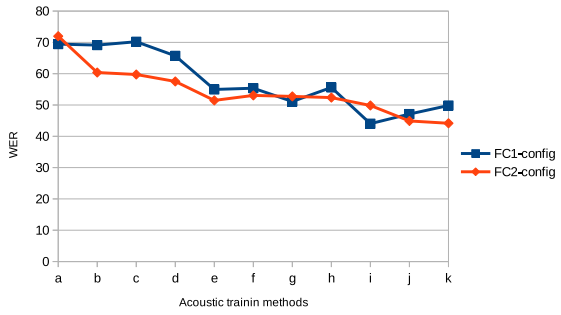


Fig. 2. Influence of speech data configuration on speech recognition quality. LM2 is fixed and only speech data and acoustic models vary. The letter in abscissa represent acoustic training methods labelled in table VI-A

column (LM2-Based ASR), we have included the WER results of ASR performed after vowel normalization (without diacritics).

Colunm of LM2-Based ASR in Table VIII also shows that triphone models significantly improve the monophone model performance. The tri2b+SAT+FMLLR acoustic model adapted to speaker from feature-space Maximum Likelihood Linear Regression reduced the WER by 6% absolute for both speech data configuration. The WER on FC1-config is lower than 20% for discriminative methods based on tri3b. For FC2-config, these acoustic training methods reduced the WER by 20%. The best results are coming from the training for Subspace Gaussian Mixture Models (SGMM), with an overall WER of 14.83% for FC1-config and 28.93% for FC2-config. The speech data divided by speakers helps us to obtain a relative gain of 14% with the best final WER of 14.83%. This leads us to choose AM training methods using SGMM for performance

TABLE VIII
WER OF LM2-BASED ASR (WITHOUT DIACRITICS) AND LM1'-BASED ASR (REMOVING OF DIACRITICS FROM HYPOTHESES AND REFERENCES OF LM1-BASED ASR).

Speech data config/ method	LM2-Based ASR	LM1'-Based ASR
FC1-config		
Monophone (a)	36.36	59.05
Triphone (b)	28.19	46.8
$\Delta + \Delta\Delta$ (c)	28.21	46.98
LDA + MLLT (d)	24.4	41.52
LDA + MLLT + SAT + FMLLR (e)	17.83	29.29
LDA + MLLT + SAT + FMLLR + fMMI (f)	19.72	31.34
LDA + MLLT + SAT + FMLLR + MMI (g)	18.93	35.59
LDA + MLLT + SAT + FMLLR + fMMI + MMI (h)	18.26	35.44
LDA + MLLT + SGMM (i)	15.23	**20.56**
LDA + MLLT + SGMM + MMI (j)	15.3	20.68
LDA + MLLT + SGMM + fMMI + MMI (k)	**14.83**	21.39
FC2-config		
Monophone (a)	52.26	57.89
Triphone (b)	38.72	47.47
$\Delta + \Delta\Delta$ (c)	38.58	46.39
LDA + MLLT (d)	35.34	42.45
LDA + MLLT + SAT + FMLLR (e)	30.74	35.63
LDA + MLLT + SAT + FMLLR + fMMI (f)	35.36	37.46
LDA + MLLT + SAT + FMLLR + MMI (g)	32.38	36.19
LDA + MLLT + SAT + FMLLR + fMMI + MMI (h)	32.94	37.52
LDA + MLLT + SGMM (i)	31.64	**31.58**
LDA + MLLT + SGMM + MMI (j)	31.36	32.75
LDA + MLLT + SGMM + fMMI + MMI (k)	**28.93**	32.02

comparison among FC1-config and FC2-config. Figure VI-B shows the evolution of WER depending on acoustic models with LM2.

It is therefore remarkable that the language model LM2 gives very satisfactory decoding results compared to LM1 standard (with diacritics). Adding diacritics in text corpus before language modelling maked the speech recognition system less efficient by increasing the WER by 44.04% compared to 15.23% (performance without diacritics). While diacritics add information, which should help the recognition system, it also increases OOV rate and perplexity of the language model (see table V).

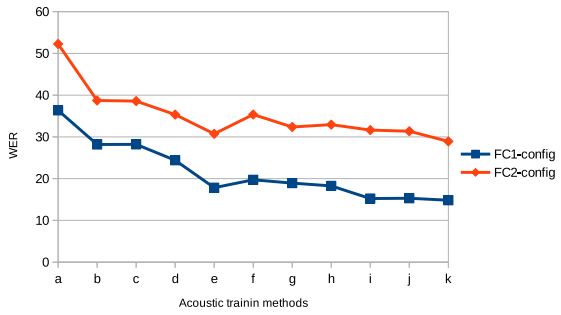


Fig. 3. Influence of speech data configuration on speech recognition quality. LM2 is fixed and only speech data and acoustic models vary. The letter in abscissa represent acoustic training methods labelled in table VI-B

For further, we performed an effective comparison of ASR performance without diacritics. To do this, we removed the

diacritics from the outputs (hypotheses) and references of ASR system built with LM1 (LM1'-Based ASR). The obtained results for this evaluation are included in the third column of Table VIII. These results can be compared to results obtained with LM2-Based ASR system (second column). This comparison leads us to assert that the removing of diacritics for different models is more effective and provides an efficient ASR system.

VII. Conclusion

In this work we introduced the first system of Fongbe continuous speech recognition by training different acoustic models using Kaldi scripts and different language models using SRILM toolkit. We also demonstrated the effect of tones on the quality of the recognition system. This leads us to conclude that with the current state of our system, the language modelling without diacritics improves significantly the recognition performances by decreasing the WER by 15.23% for speech data divided by speakers and 28.93% for speech data divided by category. Using the Kaldi recipe and the language resources we provide, researcher can build a Fongbe recognition system with the same WER obtained in this paper. For future work, firstly we will enhance the speech and text data and introduce other training techniques to further improve the performance of this first system of Fongbe recognition. Secondly, we will investigate the Fongbe re-diacritization in the context of Speech recognition.

References

[1] J. K. Tamgno and E. Barnard and C. Lishou and M. Richomme, *Wolof Speech Recognition Model of Digits and Limited-Vocabulary Based on HMM and ToolKit*, in. 14th International Conference on Computer Modelling and Simulation (UKSim), pp. 389–395, 2012 UKSim.

[2] Besacier, L., Barnard, E., Karpov, A., and Schultz, T. (2014). Automatic speech recognition for under-resourced languages: A survey. *Speech Communication*, 56:85–100.

[3] E. Gauthier and L.Besacier and S. Voisin and M. Melese and U. P. Elingui *Collecting Resources in Sub-Saharan African Languages for Automatic Speech Recognition: a Case Study of Wolof* in. 10th edition of the Language Resources and Evaluation Conference, 23-28 May 2016, Slovenia.

[4] S. A. M. Yusof and A. F. Atanda and M. Hariharan, *A review of Yorúbà Automatic Speech Recognition*, in. System Engineering and Technology (ICSET), IEEE 3rd International Conference on, pp. 242–247, Aug.2013.

[5] J. Greenberg, *Languages of Africa*, La Haye Mouton, pp. 177, 1966.

[6] C. Lefebvre and A-M. Brousseau, *A grammar of Fonge*, De Gruyter Mouton, PP. 608, December 2001.

[7] A. B. AKOHA, *Syntaxe et lexicologie du Fon-gbe: Bénin*, Ed. L'harmattan, pp. 368, January 2010.

[8] Blachon, D., Gauthier, E., Besacier, L., Kouarata, G.-N., Adda-Decker, M., and Rialland, A. (2016). Parallel speech collection for under-resourced language studies using the LIG-Aikuma mobile device app. In *Proceedings of SLTU (Spoken Language Technologies for Under-Resourced Languages)*, Yogyakarta, Indonesia.

[9] A. W. Black and T. Schultz, *Rapid Language Adaptation Tools and Technologies for Multilingual Speech Processing*, in Automatic Speech Recognition & Understanding, IEEE Workshop, pp. 51, 2009.

[10] Sebastian Dziadzio, Aleksandra Nabozny, Aleksander Smywinski-Pohl and Bartosz Ziolko, *Comparison of Language Models Trained on Written Texts and Speech Transcripts in the Context of Automatic Speech Recognition*, in Proc. Proceedings of the IEEE Federated Conference on Computer Science and Information Systems, 5, pp. 193-197, Pologne 2015.

[11] S. Seng and S. Sam and V. Bac Le and B. Bigi and L. Besacier, *Which units for acoustic and language modeling for Khmer automatic speech recognition?*, SLTU 2008.

[12] J. Billa and all, *Audio indexing of Arabic broadcast news*, in Proc. IEEE International Conference on Acoustique, Speech and Signals Processing, pp. 5-8, Orlando 2002.

[13] D. Povey and A. Ghoshal et al., *The Kaldi Speech Recognition Toolkit*, in IEEE ASRU, 2011.

[14] D. Povey and G.Saon, *Feature and model space speaker adaptation with full covariance Gaussians*, in INTERSPEECH 2006 - ICSLP, Ninth International Conference on Spoken Language Processing, Pittsburgh, PA, USA, September 17-21, 2006

[15] Lukasz Laszko, *Word detection in recorded speech using textual queries*, in Proc. Proceedings of the IEEE Federated Conference on Computer Science and Information Systems, 5, pp. 849-853, Pologne 2015.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 483–486
DOI: 10.15439/2016F316
ACSIS, Vol. 8. ISSN 2300-5963

Comparative Study of Multi-Stage Classification Scheme for Recognition of Lithuanian Speech Emotions

Tatjana Liogienė, Gintautas Tamulevičius
Vilnius University Institute of Mathematics and Informatics,
Akademijos str. 4, Vilnius, Lithuania
Email: {tatjana.liogiene, gintautas.tamulevicius}@mii.vu.lt

***Abstract*—This paper presents the experimental study of multi-stage classification based recognition of Lithuanian speech emotions. Three different criteria for feature selection were compared for this purpose: Maximal Efficiency, Minimal Cross-Correlation feature criterions, and the Sequential Feature Selection. A large database of spoken emotional Lithuanian language was used in this experiment – each of 5 emotions was represented by 1000 utterances. The results of the speaker-independent emotion recognition experiment show the superiority of multi-stage classification using the SFS technique by 0.7-8 %. This classification scheme gave the highest recognition accuracy and the smallest feature set.**

I. Introduction

SPEECH emotion recognition is a classical task of pattern classification including feature extraction, training and classification (decision making). The feature extraction step is a crucial for the successful speech emotion identification process: appropriate and relevant feature set is a key component of any valid and efficient recognition system.

Various feature sets have been proposed for speech emotion recognition [1]-[5]. In straightforward manner composed feature sets often contain a few hundred or even thousand features and this can become problematic in case of limited datasets. Thus various feature selection or transformation techniques are applied for reduction purposes [2], [3], [6], [7]. Various parallel, serial, and hierarchical classification schemes have been proposed and proved to be more effective for speech emotion recognition [2], [4], [8], [9] also.

In this paper we present multi-stage classification based recognition of Lithuanian speech emotions. Section II contains the review of multi-stage classification of speech emotions. The multi-stage classification scheme using three different feature selection criteria is presented in next section. The results of the experimental study are given in Section IV and concluded in Section V.

II. Multi-stage Classification of Speech Emotions

The classification of speech emotion can be implemented in two ways. The simplest is to classify emotions in one step using one general feature set for all emotions. Usually this means a very large but not optimal feature set.

The interest in sophisticated classification schemes has been noticeable in last few years. Variations of classification scheme include multi-stage classification (when the whole recognition process is implemented in a few steps), multiple classifier schemes (different classifiers are dedicated to separate emotions or emotion groups), pair-wise classification and others. All these classification schemes can be arranged into three groups: serial, parallel, and hierarchical (Table I).

The serial combination of classifiers considers the speech emotion classification process as the consecutive identification of one or more separate emotions during one classification step. *N-1* separate classifiers will therefore be needed to identify *N* emotions [2]. The parallel scheme is based on the concurrent identification of separate emotions – all the emotions are analyzed by a set of classifiers during one step [2], [9]. The third and the biggest group of multiple classifier systems is based on the hierarchical organization of the classification process according to some criterion.

The speaker's gender, three dimensional emotion model based groups and other criteria are used for hierarchical organization of the classification process. In general, the hierarchical group contains attributes of both the serial and parallel schemes [1]-[9].

As we can see, multi-stage organization of the speech emotion classification process results in a complicated process and the accuracy obtained varies from 50 % up to 88 %. Nevertheless, the above-mentioned multi-stage classification schemes outperform single-step schemes and provide an opportunity to modify the feature set for a particular emotion or emotion group without affecting another. This should be considered as the main advantage of multi-stage classification of speech emotion.

III. Feature Selection Based Multi-stage Classification

Considering the above-mentioned advantages, we proposed a multi-stage classification scheme for speech emotion recognition [10]. The main idea of the proposed scheme is the grouping of emotions for different classification stages. All groups of emotional speech

978-83-60810-90-3/$25.00

TABLE I
EMOTIONAL SPEECH CLASSIFICATION SCHEMES

No	Authors	Classification Schemes	Number of Emotions	Language	Accuracy
1.	W.-J. Yoon, K.-S. Park [8]	Two-step hierarchical classification	2	Chinese	80.7%
2.	J. Liu, et al. [1]	Enhanced co-training algorithm	6	Chinese	75.9% male, 80.9% female
3.	Z. Xiao, et al. [3]	Hierarchical classification	6	German	76.4%.
4.	M. Lugger, etal. [2]	Hierarchical combination of classifiers	6	German	88.8%.
5.	M. Kotti, F. Paterno [7]	Psychologically-inspired binary cascade classification scheme	6	German	87.7%
6.	C.-C. Lee, et al. [5]	Hierarchical binary decision tree approach	5	German	48.27%
7.	L. Chen, X. Mao, Y. Xue, and L. L. Cheng [6]	Three-level classification model	6	Chinese (Mandarin)	86.5%, 68.5%, and 50% (for each level)
8.	E. M. Albornoz, D. H. Milone, and H. L. Rufiner [4]	Two-stage hierarchical classification	7	German	71.75%
9.	M. Lugger, M.-E. Janoir, and B. Yang (2009) [2]	Serial combination of classifiers	6	German	96.5%
10.	M. Lugger, M.-E. Janoir, and B. Yang [2]	Parallel combination of classifiers	6	German	92.6%
11.	A. Milton and S. Tamil Selvi [9]	Class-specific multiple classifiers scheme	7	German	80.6%

utterances are labeled in several stages. During the first stage all utterances are classified into predefined groups. During successive stages, these groups are divided into subgroups or separate emotions. This classification scheme enables us to use different (more effective, we suppose) feature sets per classification node, thereby improving the overall recognition rate. The feature set for every node (we will call this set as subset) is formed individually according to performance on the emotional group analyzed.

Three feature selection techniques were applied for the multi-stage classification scheme: Maximal Efficiency criterion (ME), the criterion of the Minimal Cross-Correlation of features (MC), and the Sequential Forward Selection (SFS) based technique.

A. Maximal Efficiency Feature Selection Criterion

This criterion is applied by making an assumption about the aggregate efficiency of features with maximal individual efficiency i.e. of features giving the lowest classification error. The formation of a feature subset using the ME selection criterion is carried out

$$f_m^{(l)} = \arg\min_j E(f_j^{(l)}), j = 1,...,J. \qquad (1)$$

Here $E(f_j^{(l)})$ is a classification error of the j-th feature in the l-th level $f_j^{(l)}$. J is a total number of features in the l-th classification level.

The feature subset is initialized once and repeatedly extended with the most effective features $f_j^{(l)}$. The evaluation of every subset case is carried out and the extension process is stopped when the overall efficiency of the subset is not improved. Thus the selection procedure of J features from M feature set will require analysis of J+M-1 feature subsets.

B. Minimal Cross-Correlation Criterion

In this case an assumption is made as to the efficiency of linearly independent features. Independent features make the set more effective than strongly correlated ones. Thus by selecting linearly independent features we seek for a more effective subset.

MC criterion based feature selection is initiated with the most efficient feature thus ensuring the discriminative power of the subset. The analyzed feature subset is expanded by adding features with the minimal cross-correlation value

$$f_m^{(l)} = \arg\min_j \left| R\left(f_0^{(l)}, f_j^{(l)}\right)\right|, j = 1,...,J. \qquad (2)$$

Here $f_0^{(l)}$ is the feature with highest classification accuracy for analyzed emotion group. $R(f_0^{(l)}, f_j^{(l)})$ is the cross-correlation of the $f_0^{(l)}$ and the new feature $f_j^{(l)}$.

Again, the expansion of the feature subset is stopped when the efficiency of the feature subset begins to diminish. Similar to ME criterion the selection procedure of J features from M feature set using MC criterion will result in analysis of J+M-1 subsets.

Considering the unknown distribution of emotion feature values, the Spearman coefficient was selected to evaluate the correlation of the features. Moreover, the Spearman correlation is hypothesized as being more robust to data outliers, an aspect we find important in the case of speech emotion features.

C. Sequential Forward Selection

The SFS technique is one of the acquisitive search algorithms aiming to find the most significant subset of the features, and the aggregate efficiency of the feature subset is considered rather than individual properties of the features.

The selection of features starts from initialization of the empty feature subset F_0. The subset is extended with a feature $f_j^{(l)}$ making the new subset F_{i+1} more effective

$$f_m^{(l)} = \arg\max_j \left[E\left(F_i + f_j^{(l)}\right) - E\left(F_i\right) \right], j = 1,...,J. \quad (3)$$

The feature set extension step is repeated until the efficiency of newly obtained feature set F_{i+1} increases or while $j \le J$. $J \times M$ different feature subsets should be analyzed to select J-th order feature subset using this procedure. Therefore, the SFS will require number analyzed feature subsets grows significantly in comparison with aforementioned criteria.

The applied Maximal Efficiency and Minimal Cross-Correlation feature selection criteria, and the Sequential Forward Selection Technique make locally optimal choices and will thus give suboptimal feature subsets.

IV. Experimental Study

In this study we decided to perform a thorough comparison of the aforementioned feature selection criteria for the recognition of Lithuanian speech emotions. Three versions of the multi-stage classification scheme were implemented using different feature selection criteria and applied to the Lithuanian speech emotion identification task.

We have chosen recognition tasks for 3 emotions (anger, joy, neutral), 4 emotions (anger, joy, neutral, sadness), and 5 emotions (anger, joy, neutral, sadness, and fear). 1000 examples of each emotion (recorded by 5 females and 5 males) were analyzed during the experiment [11].

The initial full set consisted of 6552 different speech emotion features including time and frequency domain features, mel scale features, probabilities of voicing in speech, and their various derivatives (first and second order differentials, statistics, distribution data) [12].

A non-parametric K-Nearest Neighbor classifier was chosen for experimental testing. The value $K = 5$ was selected considering the large size of the data sets.

A two-stage classification scheme was designed assuming low-pitch and high-pitch emotion classes in the first classification stage. These two groups are classified into separate emotions during the second stage using a group-specific feature subset.

Considering the number of examples for each emotion, a 10-fold cross-validation scheme was selected to obtain more robust results. As every speaker pronounced 100 emotional sentences, the speech emotion recognition experiment was performed in speaker-independent mode.

The average recognition results are given in Fig. 1, detailed results for every emotion are given in Table II (the Maximal Efficiency criterion is denoted as *ME*, Minimal Cross-Correlation criterion denoted as *MC*, and the SFS technique denoted as *SFS*).

As we can see in Fig. 1, the SFS based multi-stage recognition of speech emotions showed the highest accuracy in all cases. Its superiority over other selection criteria based schemes was 0.7-8 %. Two things should be pointed about these results. To begin with, the average results are much lower than our previously obtained results. In the case of 5 emotions, the recognition rate was 50 % approximately (not impressive in comparison with results from other studies). Besides, the superiority of the SFS based multi-stage scheme is much smaller in comparison with previous results. There are two possible reasons for the lower results:

- Non-professional actors: the valence of the emotions is much lower in comparison with emotions expressed by professional actors. Consequently, the classification of these patterns is a more challenging task.
- The size of the database: the much larger set of emotional utterances contains more different verbal expressions, thus the variability of the utterances is much wider and more confusing.

The MC criterion produced the lowest recognition rates for emotional cases. In case of 5 emotions, the recognition rate was 41.6 %. Nevertheless, the single-stage recognition of 5 emotions (using the entire set of features) has shown an accuracy of 28.4 % only. Thus multi-stage classification has obvious superiority over single-stage classification.

Analysis of the recognition results for particular emotions reveals that anger and the neutral state are the most difficult emotions to recognize among the others.

Figure 2 shows the dependence of the obtained feature set size (order) on the number of recognized emotions. Again, the SFS based multi-stage scheme demonstrated the highest efficiency. The size of the feature sets was 2.5-4 times smaller in comparison with the cases of ME and MC selection criteria. The highest order was obtained for the MC criterion; the total size of the feature set was 90-110. In general, the order of the feature set increases with the number of analyzed emotions. This could be caused by suboptimal feature selection techniques.

Analyzing individual results from every speaker, we have noticed a fluctuation in recognition rate amongst speakers. For example, the average recognition results of speaker #9

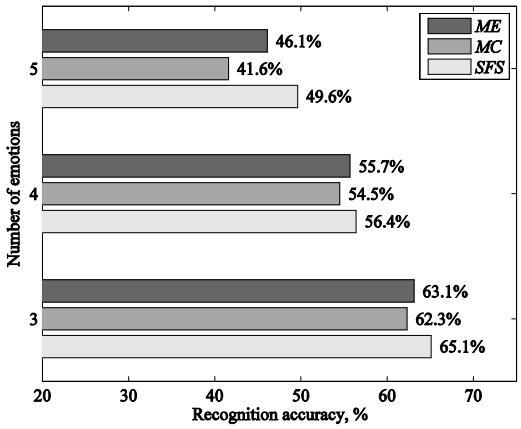


Fig 1. Average emotion recognition results

TABLE II
RECOGNITION RATES FOR PARTICULAR EMOTIONS

Criterion	Number of emotions	Recognition accuracy, %				
		Anger	Joy	Neutral	Sadness	Fear
ME	3	56.6	67.3	65.5	–	–
	4	53.5	62.2	50.1	56.8	–
	5	42.7	48.1	45.1	49.3	45.2
MC	3	54	62.6	70.3	–	–
	4	49.6	58.3	46.4	63.8	–
	5	42.4	38.1	40.3	52.8	34.6
SFS	3	57.6	71.7	66	–	–
	4	52	67.6	48	57.8	–
	5	49.4	60.6	41.2	50.3	46.7

were 1.5-2.5 times lower than average. The results of speaker #10 were 1.2-1.3 times higher than the average results. The reason is the suboptimal feature selection aiming for the highest average recognition rate not for the individual one.

V. CONCLUSIONS

In this paper the results of a comparative study of a multi-stage classification scheme for Lithuanian speech emotion recognition are presented. Three different feature selection criteria were applied for recognition purposes: Maximal Efficiency, Minimal Cross-Correlation of features and Sequential Forward Selection. The following conclusions can be drawn from the results:

- The average recognition rate was 62-65 % for the 3 emotion set (anger, joy, and neutral), 55-57 % for the 4 emotion set (anger, joy, sadness, and neutral), and 42-50 % for the 5 emotion set (anger, joy, sadness, fear, and neutral). The results are not impressive in comparison with results from other studies, but the large set of emotional utterances should be considered as the main factor for the accuracy obtained.
- Sequential Forward Selection based scheme shows higher performance in comparison with individual feature properties based selection criteria (Maximal Efficiency and Minimal Cross-Correlation in our case). The superiority was 0.7-8 %.
- The recognition of a large number of emotional utterances requires large feature sets. Increasing the number of recognized emotions also expands the size of the required feature set. Consequently, speaker and text-independent speech emotion recognition would require for huge feature sets.

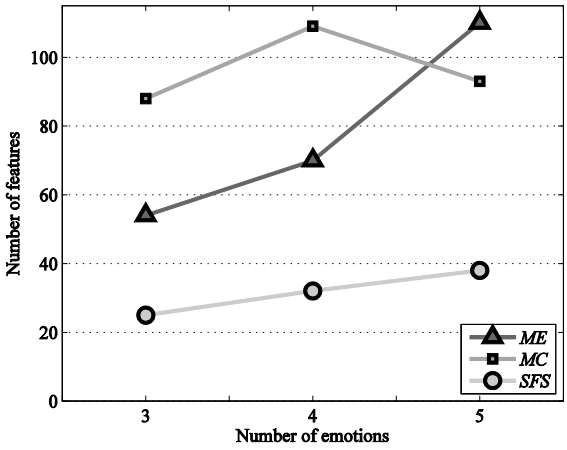


Fig 2. Feature order dependence on number of emotion

REFERENCES

[1] J. Liu, C. Chen, J. Bu, M. You, and J. Tao, "Speech Emotion Recognition using an Enhanced Co-Training Algorithm," *2007 IEEE International Conference on Multimedia and Expo,* pp. 999–1002, July 2007, http://dx.doi.org/10.1109/ICME.2007.4284821.

[2] M. Lugger, M.-E. Janoir, and B. Yang, "Combining classifiers with diverse feature sets for robust speaker independent emotion recognition," *17th European Signal Processing Conference,* pp. 1225–1229, 2009, http://dx.doi.org/10.5281/zenodo.41415.

[3] Z. Xiao, E. Centrale, L. Chen, and W. Dou, "Recognition of emotions in speech by a hierarchical approach," *3rd International Conference on Affective Computing and Intelligent Interaction and Workshops,* pp. 1–8, September 2009, http://dx.doi.org/10.1109/ACII.2009.5349587.

[4] E. M. Albornoz, D. H. Milone, and H. L. Rufiner, "Spoken emotion recognition using hierarchical classifiers," *Computer Speech & Language,* pp. 556–570, 2011, http://dx.doi.org/10.1016/j.csl.2010.10.001.

[5] C.-C. Lee, E. Mower, C. Busso, S. Lee, and S. Narayanan, "Emotion Recognition Using a Hierarchical Binary Decision Tree Approach," *Speech Communication,* pp. 1162–1171, 2011, http://dx.doi.org/10.1016/j.specom.2011.06.004.

[6] L. Chen, X. Mao, Y. Xue, and L. L. Cheng, "Speech emotion recognition: Features and classification models," *Digital Signal Processing,* pp. 1154–1160, 2012, http://dx.doi.org/doi:10.1016/j.dsp.2012.05.007.

[7] M. Kotti and F. Paterno, "Speaker-independent emotion recognition exploiting a psychologically-inspired binary cascade classification schema," *Iternational Journal of Speech Technology,* pp. 131–150, 2012, http://dx.doi.org/10.1007/s10772-012-9127-7.

[8] W.-J. Yoon and K.-S. Park, "Building robust emotion recognition system on heterogeneous speech databases," *2011 IEEE International Conference on Consumer Electronics,* pp. 825–826, 2011, http://dx.doi.org/10.1109/TCE.2011.5955217.

[9] A. Milton and S. Tamil Selvi, "Class-specific multiple classifiers scheme to recognize emotions from speech signals," *Computer Speech and Language,* pp. 727–742, 2014, http://dx.doi.org/10.1016/j.csl.2013.08.004.

[10] G. Tamulevicius and T. Liogiene, "Low-order multi-level features for speech emotion recognition," *Baltic Journal of Modern Computing,* pp. 234–247, 2015.

[11] J. Matuzas, T. Tišina, G. Drabavičius, and L. Markevičiūtė, "Lithuanian Spoken Language Emotions Database," Baltic Institute of Advanced Language, 2015. [Online]. Available: http://datasets.bpti.lt/lithuanian-spoken-language-emotions-database/.

[12] F. Eyben, M. Wollmer, and B. Schuller, "OpenEAR - Introducing the Munich open-source emotion and affect recognition toolkit," *3rd International Conference on Affective Computing and Intelligent Interaction and Workshops,* pp. 1–6, September 2009, http://dx.doi.org/10.1109/ACII.2009.5349350.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 487–490
DOI: 10.15439/2016F128
ACSIS, Vol. 8. ISSN 2300-5963

Neuro-heuristic voice recognition

Dawid Połap
Institute of Mathematics
Silesian University of Technology
Kaszubska 23, 44-100 Gliwice, Poland
Email: Dawid.Polap@gmail.com

***Abstract*—Protection of private data, signing electronic documents are selective use of identity verification. In this work, the problem of voice verification has been discussed. The method of verifying the voice based on the methods of artificial intelligence was presented. Numerous tests were performed to demonstrate the effectiveness of the presented solution - research results are shown and discussed in terms of advantages and disadvantages.**

I. Introduction

THE digitization of today's world has started some time ago. At almost every step we associate with the technology in one form or another. Each one of us carries a cell phone or laptop and therefore has continuous access to the Internet. On the Internet, people handle all the daily issues - buy goods, pay bills or store images and a variety of data. These are the times in which it is hard to remain anonymous what requires continuous improvements in data transmission and data protection.

One of the most advanced trends in computing today is artificial intelligence, which finds its use in almost any application. One of the first approaches of this trend were neural networks, which are mainly used as classifiers, i.e. classification of signatures [1], surface structures [2] and multimedia applications [3]. Modern methods of mathematics and artificial intelligence are increasingly finding application in medicine and other aspects. In [4] and [5], the authors presented an interesting approach to the processing of EEG signals. A very important achievement is also fuzzy logic, which is often combined with neural networks in order to not only increase precision, but allow the classification of other measures such as linguistics [6] and [7]. Similarly we can use voice recognition in AAL environments to help users [8] and improve data acquisition [9]. These aspects of artificial intelligence methods are also widely used in verification of people. On a daily basis, we meet with the problem of verification in banks, stores (e.g. when signing checks), companies or in different institutions for which a signature, voice, or even features of the iris are a confirmation of our identity. Voice verification is one of the least expensive in implementation of solutions of this type – just a microphone and expert application are needed. Different approaches are analyzed in order to maximize the precision of the voice classification. In [10], the authors presented the statistical approach, again in [11] used hidden Markov models for the same purpose. In this work, I would like to introduce an innovative way of extracting features of the voice sample using heuristic and neural classifier.

II. Audio signal processing

For the purposes of analysis of digital sound samples, sound file must be represented by a numerical value, or a function that enables analysis. One of the best known methods is the Discrete Fourier Transform (DFT) presented in [12]. Transform is called transformation of vector that represents the numerical values of the signal $\vec{x} = [x_0, x_1, \ldots, x_{N-1}]$ into $\vec{z} = [z_0, z_1, \ldots, z_{N-1}]$, where the value of $z_i \in \mathbb{C}$. DFT is performed according to the following formula

$$z_k = \frac{1}{N} \sum_{n=0}^{N-1} x_n \exp(\frac{-2kni\pi}{N}), \tag{1}$$

where k is harmonic number, n is number of signal samples and N is total number of samples.

One way to display the audio signal is to use a spectrogram [13] which is a graph of amplitude spectrum signal over time. The construction of the spectrogram operates on the principle of dividing the signal (obtained by the use of short-time fast Fourier transform with Hamming function) on parts for which the amplitude of harmonic components are calculated. During the analysis of this graph, we assume that

- the power at a given frequency is based on the color - the value is higher if the color is warmer;
- the frequency increases with the height of the point on the chart.

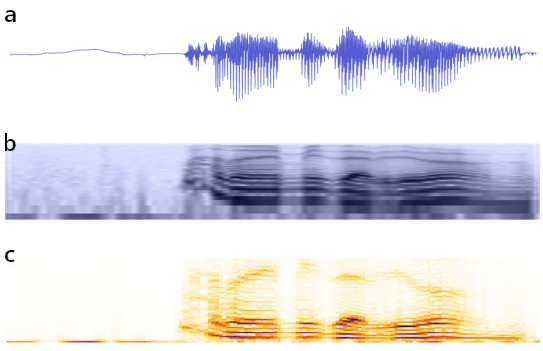


Fig. 1: Graphical representation of sound samples with the sentence "My name is Han Solo": (a) the DFT of the signal (b) spectrogram of the recorded sound with noise (c) spectrogram of the clear signal.

978-83-60810-90-3/$25.00

Recorded sound is exposed to register all types of noise - in the case of voice verification in a large company, people standing in a queue already make noise as well as any other voices. Noise is clearly visible on spectrograms interiors as a plurality of light color. In order to correct analysis of sound, noise should be removed to leave only the desired recording. Noise reduction can be achieved by a multi-band spectral subtraction described in [14]. The algorithm assumes that speech spectrum will be divided into N bands and the estimate of the clean speech spectrum can be obtained by performing the following formula for each band i for k value

$$|\hat{S}_i(k)|^2 = |Y_i(k)|^2 - \alpha_i \delta_i |\hat{D}_i(k)|^2 \qquad l_i \leq k \leq h_i, \quad (2)$$

where $\hat{S}$ means the magnitude spectra of the clean speech and $\hat{D}$ of the noise, Y is the magnitude spectra of the incoming signal, l_i is the beginning frequency of the ith band and similarly h_i is the ending frequency, δ_i is a tweaking parameter which is set by empirical way. The parameter α_i can be calculated by

$$\alpha_i = \begin{cases} 5 & \psi_i < -5 \\ 4 - 0.15 * \psi_i & -5 \leq \psi_i \leq 20 \\ 1 & \psi_i > 20 \end{cases}, \quad (3)$$

where

$$\psi_i = 10 \log_{10} \left[\left(\sum_{k=l_i}^{h_i} |Y_i(k)|^2 \right) / \left(\sum_{k=l_i}^{h_i} |\hat{D}_i(k)|^2 \right) \right]. \quad (4)$$

The samples of this methods to process audio file are shown in Fig. 1. Another way to present the audio signal is a periodogram, which for the first time was shown by Arthur Schuster [15]. It is obtained by the modulus squared of the DFT and presents the Power Spectral Density (PSD) estimate, as shown in Fig. 3.

III. Preprocessing for Neural Network

Creating samples is an important part for the pattern recognition problem. Each sample must not only represent input data but save as much information about the sample using the fewest number of values. For that purpose, a model of the general pattern based method on a large number of input samples is proposed. All samples representing the specific person will be created using general pattern.

A. Preparation of the aggregate sample

In the first step of processing sound samples, spectrograms are created. Each spectrogram is subjected to noise removal. On the basis of all samples, the aggregated sample is prepared. This process is about creating one sample spectrogram, which retain repeating features in all processed ones. At the beginning of the method, the $w \times h$ array is created, where w is the width and h the height of all samples - each point (x, y) on spectrogram is corresponding to one cell in the array. Initially, each cell is set to 0. For each point (x, y) color pixel on each bitmap is verified. In the case where the pixel is not white, the value on this position (x, y) is increased by 1 in the array. Color $[R_{old}, G_{old}, B_{old}]$ is selected and the new sample is created. For each pixel, the color is calculated using a value of m from an array and the following formula

$$C_{new} = C_{old} * (1 - m\varsigma), \quad (5)$$

where C is one of the color components (R, G or B) and ς means the color shade and it is calculated as

$$\varsigma = \frac{2}{\max\limits_{0 \leq i < w \wedge 0 \leq j < h} m_{ij}}. \quad (6)$$

In the next section of this paper, created spectrogram will be called a *general spectrogram*. The process of creating a *general spectrogram* is shown in Fig. 2.

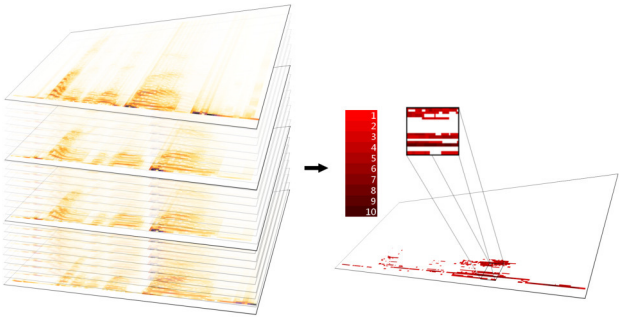


Fig. 2: Visualization of creating a *general spectrogram* based on input spectrograms with attached legend of colors - the darker shade of red, the point often occurs on the input samples.

B. Heuristic detection of key-points

Using the so-created *general spectrogram*, heuristic algorithm will be used to find the key points of the image. The coordinates of key points will allow to create the mask of the features by which we will be able in a quick and effective way to retrieve specific values from the spectrogram to perform training using verification vectors.

As a heuristic algorithm, Flower Pollination Algorithm (FPA) [16] was selected, which is a mathematical model of a natural phenomenon pollination of flowers in the spring. The original algorithm assumes some basic rules aimed at simplifying some dependence

- Global pollination (e.g.: biotic phenomenon) is represented as Levy flights,
- Local pollination is interpreted as abiotic and self-pollination,
- Pollen is carried by the wind what is modeled by a random factor $p \in [0, 1]$.

FPA is used to find the important points in the image. Population of flowers is placed in random places on *general spectrogram*. Then, the simulations of global and local pollination are performed in order to find the best points. Point

selection is done according to the fitness function, which is defined as

$$F(x_i) = \begin{cases} \min\limits_{0 \le j < N} 0.2 * B(x_j) & \text{if } x_{best} = 0 \\ \sqrt{B^2(x_i) - B^2(x_{best})} & \text{for others} \end{cases}, \quad (7)$$

where x_i is a point representing a pollen, x_{best} is a point with the best fitness value in actual iteration, and the function $B(\cdot)$ is the brightness of a pixel that takes values of $[0, 1]$, N is the size of population.

In each iteration, two operations are performed – global and local pollination. Global pollination moves pollen x_i over the spectrogram image according to

$$x_i^{t+1} = x_i^t + L(x_i^t - x_{neighboring}), \quad (8)$$

where t is the number of iteration, $x_{neighboring}$ is the nearest point to x_i and $L(\cdot)$ is a function of Levy flight understood as

$$L(x, \kappa, \mu) = \sqrt{\frac{\kappa}{2\pi}} \frac{e^{-\kappa/(2(x-\mu))}}{(x-\mu)^{3/2}}, \quad (9)$$

where κ and μ are specified parameters.

The second operation is a local pollination which takes place over the neighborhood pixels for a given point x_i and it is defined as

$$x_i^{t+1} = x_i^t + \epsilon(x_j^t - x_k^t), \quad (10)$$

where x_j and x_k are neighboring pollens.

The algorithm returns the best adapted solution with its motion path. From a mathematical point of view, the trajectory is a closed consisting of two-dimensional coordinates of key points found on the spectrogram. Created set will allow to get the appropriate features based on the spectrogram in a very short time. Moreover, for every person, FPA will find a different trajectory, and therefore the possibility of fraud during the verification respectively decreases.

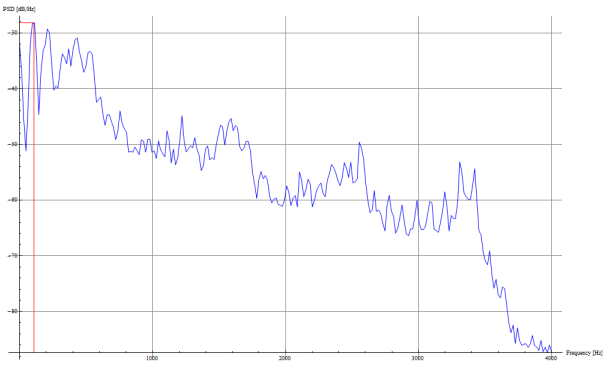


Fig. 3: Periodogram of the sentence "Han Solo" with the selected maximum value of PSD.

C. A general vector for identification

In order to analyze every sound, a file should be saved as a vector of numbers representing as much as possible features. In addition, the vector should contain the minimum number of values, because then a learning process occurs much more efficient. The proposed model in the general formula for a vector in this work is the combination of the features extracted from the spectrogram and the periodogram using the method described in Sec. III.

The proposed model of the sample can be defined as follows

$$[B(x_0), B(x_1), \ldots, B(x_n), f_{max}, id], \quad (11)$$

where $B(x_i)$ is brightness value of pixel at the key-point x_i obtained from FPA, f_{max} is the maximum value of the PSD determined from periodogram (see Fig. 3) and id is the designation of the owner of the sample.

IV. Neural Network

The beginning of the history of neural network models is dated to the first half of the twentieth century [17]. Until today, new and modified models of the network, or even learning algorithms of this type of networks are created. A neural network is understood as a complex object composed of many layers. The simplest models contain only two - input and output layer. In this type of layers, learning vector is inserted into input layer and the calculation result is forwarded to the next layer. In more complex structures the hidden layers are added between input and output. Their number depends on the size of the problem, for which a structure has been created. Each layer is made of neurons through which the layers are connected – each neuron of one layer is connected to each neuron in the next one. Neurons use activation function, whereby the value is calculated. In [18], different activation functions were proposed, but the most common is a unipolar sigmoid function. Moreover, each connection between two neurons is burdened with a certain weight in the range of $[0, 1]$ and this contributed to a number of learning methods for improving received results.

One method of learning neural network is the back-propagation algorithm for the first time mentioned in [19]. To this day, it is one of the most commonly used training algorithms for neural networks. It is a type of supervised learning which is based on minimizing the error function of the output layer up to the achievements of the first hidden layer. The error calculated in this way is used to modify the weights on the connections between neurons. The formula for calculating an error for the neuron k is defined as follows

$$\delta_k = \begin{cases} out_k(1 - out_k)(ex_k - out_k) & \text{the output layer} \\ out_k(1 - out_k) \sum\limits_{o \in out} w_{ok}\delta_k & \text{the hidden layer} \end{cases}, \quad (12)$$

where out means the output values from neuron k, ex is the expected value at the output of the neuron. The calculated error values are used in updating the weights by

$$w_i = w_i + \delta w_i. \quad (13)$$

V. Experiments

Using the method of pre-processing sound samples described in Sec. II and III are processed. In experimental research 100 different sound files belonging to four people

were used. During the processing, the following parameters were used

- FPA – 100 flowers, 15 iterations, $\kappa = 0.4$, $\mu = 0.35$,

After creating samples, obtained vectors have been placed in a database and used for training the network until the error value was smaller than 0.1 with the proportion of mixed samples – 80% to train and 20% to verify. Subsequently, using all samples for examined network an accurate measurements were calculated in Tab. I.

TABLE I: Average accuracy for the identification

Id	Correctly classified	Incorrectly classified	Accuracy
1	14	11	56%
2	20	5	80%
3	23	2	92%
4	18	7	72%

The results allow to calculate the effectiveness of the verification network, which reaches 75% with 25 samples of voice for one person. The results in the table indicate that for a person no. 1 efficacy haughty only 56% accuracy, again for a person no. 3 it was 92%. Such a large discrepancy between the results can be an effect of a bad sound recording or different voice tones. Another problem that might be a big obstacle in good verification is possible jitters or hoarseness. This can be accomplished by recording the samples for a few days—not only by one day.

VI. Final Remarks

In times where almost everything revolves in the digital world, any method of data security become an important area of IT. Newer and more complex algorithms can increase the security not only of information in computer networks, but many companies, where the entrance is authorized.

The innovative idea of the features extraction from the processed audio samples is presented in this work. Experiments based on artificial intelligence methods have been performed and discussed. The proposed method of voice verification due to the 75% effectiveness is a good alternative to existing methods. Moreover, the use of heuristics to find a sequence of points can efficiently assist in verification of possibility of fraud - each sequence of points is different because of the basic assumption of randomness for heuristics. Therefore proposed solution seems to be right development to increase security in man-machine interactions.

VII. Acknowledgment

Author acknowledge contribution to this project of Operational Programme: "Knowledge, Education, Development" financed by the European Social Fund under grant application POWR.03.03.00-00-P001/15.

References

[1] D. Połap and M. Woźniak, "Flexible neural network architecture for handwritten signatures recognition," *International Journal of Electronics and Telecommunications*, vol. 62, no. 2, pp. 197–202, 2016, DOI: 10.1515/eletel-2016-0027.

[2] G. Capizzi, G. Lo Sciuto, C. Napoli, E. Tramontana, and M. Woźniak, "Automatic classification of the fruit defects based on co-occurrence matrix and neural networks," in *Proceedings of the Federated Conference on Computer Science and Information Systems - FedCSIS'2015*. 13-16 September, Lodz, Poland: IEEE, 2015, pp. 861–867, DOI: 10.15439/2015F258.

[3] M. Knop, T. Kapuscinski, W. K. Mleczko, and R. A. Angryk, "Neural video compression based on RBM scene change detection algorithm," *Lecture Notes in Artificial Intelligence - ICAISC'2016*, vol. 9693, pp. 660–669, 2016, DOI: 10.1007/978-3-319-39384-1_58.

[4] R. Damaševičius, M. Vasiljevas, I. Martišius, V. Jusas, D. Birvinskas, A. Venčkauskas, and M. Woźniak, "BoostEMD: an extension of EMD method and its application for denoising of EMG signals," *Elektronika IR Elektrotechnika*, vol. 21, no. 6, pp. 57–61, 2015, DOI: 10.5755/j01.eee.21.6.13763.

[5] I. Martisius, D. Birvinskas, R. Damasevicius, and V. Jusas, "EEG dataset reduction and classification using wave atom transform," in *Proceedings of the 23rd International Conference on Artificial Neural Networks and Machine Learning; ICANN 2013 - Volume 8131*. New York, NY, USA: Springer-Verlag New York, Inc., 2013, pp. 208–215, DOI: 10.1007/978-3-642-40728-4_26.

[6] K. Cpałka, O. Rebrova, R. Nowicki, and L. Rutkowski, "On design of flexible neuro-fuzzy systems for nonlinear modelling," *International Journal of General Systems*, vol. 42, no. 6, pp. 706–720, 2013.

[7] C. Napoli, G. Pappalardo, E. Tramontana, R. K. Nowicki, J. T. Starczewski, and M. Woźniak, "Toward automatic work groups classification based on probabilistic neural network approach," *Lecture Notes in Artificial Intelligence - ICAISC'2015*, vol. 9119, pp. 79–89, 2015, DOI: 10.1007/978-3-319-19324-3_8.

[8] R. Damaševičius, M. Vasiljevas, J. Salkevicius, and M. Woźniak, "Human activity recognition in aal environments using random projections," *Computational and Mathematical Methods in Medicine*, vol. 2016, pp. 4 073 584:1–4 073 584:17, 2016, DOI: 10.1155/2016/4073584.

[9] M. Woźniak, D. Połap, R. K. Nowicki, C. Napoli, G. Pappalardo, and E. Tramontana, "Novel approach toward medical signals classifier," in *IEEE IJCNN 2015 - 2015 IEEE International Joint Conference on Neural Networks, Proceedings*. 12-17 July, Killarney, Ireland: IEEE, 2015, pp. 1924–1930, DOI: 10.1109/IJCNN.2015.7280556.

[10] Y. Zhang, S. Sankaranarayanan, and F. Somenzi, "Statistically sound verification and optimization for complex systems," in *Automated Technology for Verification and Analysis*. Springer, 2014, pp. 411–427.

[11] G. Hinton, L. Deng, D. Yu, G. E. Dahl, A.-r. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. N. Sainath *et al.*, "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," *Signal Processing Magazine, IEEE*, vol. 29, no. 6, pp. 82–97, 2012.

[12] S. Winograd, "On computing the discrete fourier transform," *Mathematics of computation*, vol. 32, no. 141, pp. 175–199, 1978.

[13] J. L. Flanagan, *Speech analysis synthesis and perception*. Springer Science & Business Media, 2013, vol. 3.

[14] S. Kamath and P. Loizou, "A multi-band spectral subtraction method for enhancing speech corrupted by colored noise," in *IEEE international conference on acoustics speech and signal processing*, vol. 4. Citeseer, 2002, pp. 4164–4164.

[15] A. Schuster, "On the investigation of hidden periodicities with application to a supposed 26 day period of meteorological phenomena," *Terrestrial Magnetism*, vol. 3, no. 1, pp. 13–41, 1898.

[16] X.-S. Yang, "Flower pollination algorithm for global optimization," in *Unconventional computation and natural computation*. Springer, 2012, pp. 240–249.

[17] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *The bulletin of mathematical biophysics*, vol. 5, no. 4, pp. 115–133, 1943.

[18] P. Sibi, S. A. Jones, and P. Siddarth, "Analysis of different activation functions using back propagation neural networks," *Journal of Theoretical and Applied Information Technology*, vol. 47, no. 3, pp. 1264–1268, 2013.

[19] P. Werbos, "Beyond regression: New tools for prediction and analysis in the behavioral sciences," 1974.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 491–495
DOI: 10.15439/2016F332
ACSIS, Vol. 8. ISSN 2300-5963

Web Services Ontology Population through Text Classification

José A. Reyes-Ortiz
Autonomous Metropolitan University
Azcapotzalco, Mexico City, Mexico
Email: jaro@correo.azc.uam.mx

Maricela Bravo
Autonomous Metropolitan University
Azcapotzalco, Mexico City, Mexico
Email: mcbc@correo.azc.uam.mx

Hugo Pablo
Autonomous Metropolitan University
Azcapotzalco, Mexico City, Mexico
Email: hpl@correo.azc.uam.mx

***Abstract*—In this paper, we describe the process by which web services ontologies are populated from a web services collection. The general approach relies on a global ontology model that is used to represent automatically web services. The model is enriched with web service instances classified into a taxonomy. The main idea is to extract taxonomic relations (*isTypeOf*) from web services using a supervised classifier of textual descriptions attached to web services. The entire process for ontology population involves the following tasks: text extraction from web service descriptions, classification of text descriptions and extraction of taxonomic relations (instances of classified web services). An experimentation was carried out with a collection of web service, which shows promising results and the feasibility of our approach.**

***Index Terms*—web services classification; ontology population; web services ontology population; text classification.**

I. Introduction

WEB services are reusable software components through which it is possible to build and integrate new applications without having to implement all elements of a system. Nowadays, Web services have become more popular due to their proliferation for offering storage services and resource management in the cloud. Web services are available in both public and private repositories using descriptions in XML and natural language, such as: English, Spanish or German. There are several public repositories of Web services, for example: SOAP Web service directory supported by Membrane[1]; Repository of Visual Web Service[2]; ProgrammableWeb[3]; OWLS-TC[4].

Web services are described using the standard WSDL and OWL-S. Both consist of an XML file, in which necessary elements to achieve a detailed description of web services is defined.

Programmers and application developers can use web services like software components, but they need search them into a large volume of web service published in repositories. This task is commonly known as web services discovery. However, web services discovery remains a difficult and error-prone task, since web services repositories offer keywords-based search mechanisms. In addition, web service repositories are organized in static structures that do not allow a flexible and dynamic organization of services. As a solution to this problem, ontologies can organize web services repositories with semantic relations and taxonomic relations in order to offer a semantic organization of them. In addition, this structure can help to discover and test web services in the work presented in [1].

In this paper, we present an approach for web services classification based on the frequency of *1-grams* (words), in order to populate a web services ontology. The main aim of this paper is to improve the structure repositories in order to facilitate web services discovery by providing an ontology-based semantic structure.

The rest of the paper is organized as follows. In Section II, we report the state of the art related with classification and ontology population in web services. Section III describes Web Services Descriptions Language (WSDL) and an ontology-base service description language (OWL-S). Section IV presents the global ontology model used for populate web services. Section V presents our approach for web services ontology population. Experiments and results are presented in Section VI. Finally, conclusions and future work are shown in Section VII.

II. Related Works

Web services are described using parameters (input and output) names, data type names and operation names. These elements have been exploited for several purposes. As in [2] and [3] that have used text processing with parameters and operations names in order to obtain the similarity between web services; content-based approaches for web services classification have been proposed in [4], [5], [6], [7], [8] and [9], which are described below. Web services clustering from WSDL documents in order to facilitate web services discovery proposed in [10] and [11], also in [12] a clustering-based approach to web service categorization in order to form a hierarchy of service taxonomy is presented.

In this paper, we rely on text classification to enable web services ontology population. Text classification is a task widely used, as in [13]. Moreover, web services classification is a task that has been addressed in different ways. Regarding with web services ontology population using text descriptions

[1]http://www.service-repository.com/
[2]http://www.visualwebservice.com
[3]http://www.programmableweb.com
[4]http://projects.semwebcentral.org/projects/owls-tc/

classification of web services, poor works have been propounded. However, web services classification is close to web service ontology population since it needs a class name to be instantiated. Thus, we have conducted our research of related works in two aspects: approaches for supervised classification of web services; and works for ontology population.

Using OWLS-TC collection, we have the following works: [4] uses web services textual descriptions, such as: operations, inputs/outputs textual descriptions. They classify web services with a *Naive-Bayes* classifier. In [5], the classification of web services is based on support vector machine algorithm and it is achieved by calculating a similarity between words using *WordNet* and a domain taxonomy in order to reach an efficient classification of web services in the collection. A similar web services classification has been proposed in [6], in this case it is based on sets, they propose a representation of web service descriptions with vector space model and an entropy-based weighting of all terms.

There are works that not use the public collection mentioned above. They use private collections of web services. [7] showed that using quality attributes (reliability, documentation, performance, and response time), it is possible to classify and predict the quality of a web service; they have used a private collection of 364 web services. Unsupervised classification is applied in [8], where an automatic classifier is presented based on tags embedded in WSLD documents for each web service, its method was tested with 951 web services distributed in 19 categories. In addition, other relevant work is presented in [9], they expose a text mining approach to web services classification by identifying key concepts in textual documentation services, but only in a specific domain.

Another field for this paper is web services ontology population. Some works that have presented ontology learning methods includes ontology population for web services domain. As in [14] that enhances an existing ontology with similarity relations between operation of web services. An ontology learning mechanism is proposed in [15] in order to enable *RESTful* semantic web services using syntactic and semantic descriptions. And [16] propose an automatic extraction method that learns domain ontologies from textual documentations attached to Web services.

III. Web Services Description

The recommended SDL for the Web Service implementation is named Web Service Description Language (WSDL), which is currently a well-established W3C standard. WSDL defines an XML grammar for describing networked services as collections of communication endpoints capable of exchanging messages. In this work, we consider WSDL 2.0, that is the latest version, which incorporated important changes in the description of a service. WSDL 2.0 changes the *definitions* tag with the *description* tag. The main difference between WSDL 2.0 and previous versions are: the *targetNamesSpace* is a required attribute of the definitions element in WSDL 2.0; message constructs are removed in WSDL 2.0; operator overloading is not supported in WSDL 2.0; *PortType* is renamed as Interface; Interface inheritance is supported by using the extends attribute; and Port is renamed as Endpoint.

Also, OLW-S is an ontology-based service description language, which provides a semantic description of web services. OWL-S is based on *Service* class that presents a *Profile* class in order to describe the service functionality, which is described by *Service Model* class. OWL-S contains descriptions in natural language focused on the understanding by humans. A challenge is faced when this user-focused description need to be processed.

IV. Global Ontology Model

This section presents a global ontology model exposed in [17], which is used to populate web services from text descriptions in our approach. It is also widely explained here.

We use Manchester Syntax for OWL 1.1 [18] in order to present the global ontology due to it is a user-friendly syntax. Thus, the ontology model created in [17] represents the following relevant classes for our approach.

```
Class: GeneralService
Class: GeneralOperation
Class: GeneralEffect
Class: GeneralPrecondition
Class: GeneralParameter
Class: InputParameter
     SubClassOf: GeneralParameter
Class: OutputParameter
     SubClassOf: GeneralParameter
```

The ontology model also includes object properties, relation between classes, as follows:

```
ObjectProperty: hasOperation
     Domain: GeneralService
     Range: GeneralOperation
ObjectProperty: hasEffect
     Domain: GeneralOperation
     Range: GeneralEffect
ObjectProperty: hasPrecondition
     Domain: GeneralOperation
     Range: GeneralPrecondition
ObjectProperty: hasInput
     Domain: GeneralOperation
     Range: InputParameter
ObjectProperty: hasOutput
     Domain: GeneralOperation
     Range: OutputParameter
```

The idea of this paper is that the web service instances are created and classified into subclasses of *GeneralService* class and thus, we have an extended ontology model from presented in [17]. Under this assumption, Figure 1 shows nine subclasses that we have proposed to be populated.

V. Web Service Ontology Population

Web Service Ontology Population is carried out by text classification of each Web service description in WSDL tags and natural language. Text classification associates predefined

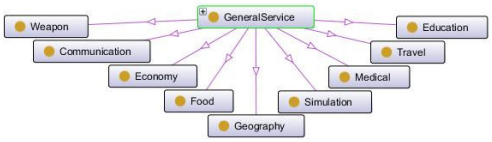


Fig. 1. Our taxonomy proposed from GeneralService class

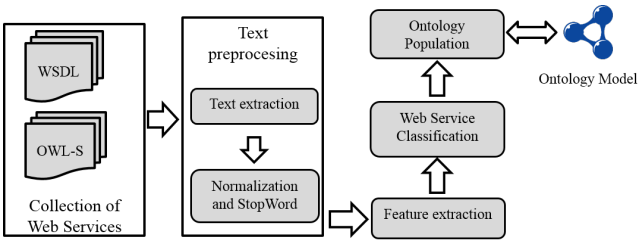


Fig. 2. Architecture of Web service ontology population

categories to a Web service from text analysis. Figure 2 shows our architecture for web service ontology population from web service descriptions through text classification. From a general ontology model, our process of ontology population is able to discover taxonomic relations to enhance our initial model and thereby achieve the classification of web services.

Web service ontology population uses a collection of public web services that are classified in any of nine proposed categories. Each web service in our collection has associated a WSDL description and an ontology-based extension of this description in OWL-S. We have proposed four phases in order to populate a web service ontology, which are described below.

A. Text preprocessing

WSDL files and OWL-S files that describes web service functionality are analyzed in order to identify and extract the text descriptions, which is useful for achieve content-based classification. Thus, the content of *serviceName* and *textDescription* labels from *Profile* class is extracted, such labels contain natural language text. From WSDL file, the service name (*wsdl: service name*), names of operations (*wsdl: operation name*) and the data type names of messages, both simple or complex (*xsd: simpleType name* and *xsd: complexType name*). This natural language text and full names of elements are used to characterize, classify and represent the web service instances into the corresponding class in our extended ontology model.

Text extracted is segmented into canonical words. In web services domain is common to find complex names of services, operations and data type names, which are formed by two or three complex words. They are segmented into lexical words by considering the switching from lowercase to uppercase and the use of subscripts as separator of lexical words. As an example: *getAddressLocation* or *get_address_location* is separated into the following lexical words: [*get*] [*address*] [*location*].

In addition, the texts are normalized by applying a conversion to lowercase, removing punctuation and stop words, which they not add meaning to the services and therefore are considered non-functional for classification based on service content.

B. Features extraction and selection

Lexical words of each web service were normalized and selected in order to represent a space model to have a formal representation of each web service. This model uses the lexical words (*1-gram*) like features. From our collection of web services 1801 words are extracted and represented in vectors.

Features are weighted using the *Bag-of-Word(BoW)* model, which consists of a collection of texts and their vocabulary that is considered features in our research. Each web service is expressed like a vector $S_j = (w_{1j}, w_{2j} w_{nj})$, where w_{ij} expresses the relevance that produces a feature i in a text j. In our case, a lexical word (*1-gram*) of vocabulary represents a feature i and web service descriptions represent a text j.

The approach used to obtain the relevance of a feature (*1-gram*) in the text of a web service is based on applying Term Frequency Inverse Document Frequency (*TF-IDF*) in order to determine what lexical word in the collection of texts might be more important for a text description of a Web service. This weighting uses the Term Frequency (*TF*) of our vocabulary in a text (1) and Inverse Document Frequency (*IDF*) that determines whether the term is common in the text collection (2), then the final equation to calculate a *TF-IDF* of a *1-gram* is shown in (3).

$$TF(t_i, S_j) \tag{1}$$

In order to determine the value of *Term Frequency (TF)*, we use the number of times that the term t occurs in a web services description *S*.

$$IDF(t_i, S_j) = \log \frac{|S|}{1 + |sS : t_i s|} \tag{2}$$

Inverse Document Frequency (*IDF*)is obtained by dividing the total number of web service descriptions by the number of web service descriptions that contain the term, and then, logarithm of such quotient is taken.

$$w_{ij} = TF(t_i, S_j) \times IDF(t_i, S_j) \tag{3}$$

A feature selection process was carried out in order to reduce our vector space. The aim of features selection is to reject irrelevant features to obtain the best subset to improve the accuracy in web services classification. For this process, we apply an analysis of attribute correlation and wrapping scheme based on decision trees, widely used as the features selection algorithm, in order to select relevant features. Such process has obtained a subset of 69 lexical words (*1-gram*) like relevant features.

C. Web service classification

Web services classification is based on features extracted from web service descriptions in WSDL and OWL-S and weighted with *TF-IDF*. The features are represented in vectors, which are used by a supervised classifier, widely used in machine learning to estimate the predictive function of each class of a collection.

As a supervised classifier is used to obtain the category of a web service, we have divided our collection in two sets: a training and test set. The main aim of this phase is to build a classifier of web services using their text description. Nine categories are considered by the classifier to be built: *Communication, Economy, Education, Food, Geography, Medical, Simulation, Travel and Weapon*. They are automatically assigned to each web service.

A rule-based classifier was built under machine learning schema using the *1-gram* words like features. C4.5 [19] was used like classification model for web services. This classifier builds decision trees from a set of training data by a selection of the best attributes that obtain effectively partitions from training set, then, the rules of the decision tree are applied to the test set in order to evaluate the results.

C4.5 classifier was trained with 899 web service descriptions, adjustments in the trees are applied for obtaining the corresponding rules, which are used with the test data for obtaining the similar output. The difference in the output determines the performance of our web services classifier. We have implemented the C4.5 for web services classification using WEKA tool [20].

D. Ontology population

This section presents the process of automatic population of Web services ontology. This task uses the assigned categories in the previous phase, they are used to create class instances in order to populate the global ontology model used for the representation of Web services.

Ontology population is the process of adding class instances and relation instances between individuals of the ontology into an existing ontology [21]. Thus, ontology population takes the category assigned to each Web service and it creates an ontological instance into the corresponding class from global ontology model. A *isTypeOf* relation is created between the instance of the Web service and the corresponding class.

Our ontology population process is illustrated with an example. Given a pre-processed text of the web service identified like *ActivityDestination*, as shown below, we determine his class in order to be populated.

Web service description: *name: activity destination service; operation name: get destination; input name: activity name; output name: destination name; description: this web service provides the destinations where an activity is available.*

The following code represents the population of class *Travel* service with the *ActivityDestination* service.

```
Individual: ActivityDestination
    Types: Travel
    Facts: hasOperation GetDestination
    Facts: hasName "Activity destination
service"
```

And the *GetDestination* operation has an input and an output, which are described as follows:

```
Individual: GetDestination
    Types: GeneralOperations
    Facts: hasInput ActivityName
           hasOutput DestinationName
```

VI. Experimentation and Results

The main idea of our experiments is to evaluate the classification task, which was carried out with *C4.5* classifier. Thus, with the evaluation of our classification task, we are evaluating the correct population of global ontology model used to represent Web Services. It is because that the population of web services ontology corresponds to create instances in the class discovered by *C4.5* classifier.

According to the assessment presented above, we carried out our experiments. We use selected word of the vocabulary like features.

Our evaluation was carried out with version 3.0 of OLWS-TC collection, which consist of 1129 web services described using WSDL and OWL-S. As this collection is considered to test, their web services are pre-classified into the following classes:*Communication, Economy, Education, Food, Geography, Medical, Simulation, Travel and Weapon*. It was divided into two groups: 899 web services for training web services classifier and 230 web services for testing our approach.

The experimentation was performed with 899 web services in order to obtain the classification model and then, it was applied to 230 web services to be tested. Also, they used the vector of *1-grams* like features weighted with *TF-IDF* for each web service and the *C4.5* classifier with the following parameters: *confidence factor used for pruning the tree = 0.25* and *the minimum number of instances per leaf = 2*.

We evaluate the results in terms of *Precision(P)*, *Recall(R)* and *F-measure(F1)*. Metrics widely used in classification tasks. In our case, these metrics compare the results to be evaluated with external values of trust (web services previously classified).

Table I shows the results of our evaluation, which consists of testing *C4.5* classifier with features extracted (*1-grams*) and weigh them using *TF-IDF*.

The results showed in Table I emphasize that the use of selected features (words) present promising results reaching 92.1 % of correctly classified web services.

We have used OLWS-TC collection for our experiments. This collection has been used as a benchmark for works that have proposed web services classification. Therefore, it is possible to compare our results with previously proposed approaches, such as Zhang and Pan [4], Wang et al. [5], Chen et al. [6] and Yuan and Jian [8]. In Table II a comparison of proposed approaches with our approach is presented in terms

TABLE I
RESULTS OF THE WEB SERVICES CLASSIFICATION

Class	Precision	Recall	F1
Communication	0.927	0.879	0.903
Economy	0.955	0.953	0.954
Education	0.843	0.944	0.891
Food	0.857	0.706	0.774
Geography	0.982	0.900	0.939
Medical	0.955	0.863	0.906
Simulation	0.933	0.875	0.903
Travel	0.961	0.909	0.934
Weapon	0.974	0.925	0.949
Average	**0.924**	**0.921**	**0.921**

of accuracy, representing the number of correctly identified true or false classifications of web services.

TABLE II
RESULTS WITH OWL-TC COLLECTION

Approach	Accuracy
Zhang and Pan	0.41
Wang et al.	0.89
Chen et al.	0.85
Yuan and Jian	0.87
Our approach	**0.921**

The results show the effectiveness of our approach compared with other approaches under the same Web services collection and the same evaluation criteria. Although the results are not so encouraging for *Food* class as for other classes, our solution can help developers to recovery Web services and reuse existing software components in a disorganized repository of web services.

VII. CONCLUSIONS AND FUTURE WORK

This paper has presented an approach for web service ontology population through a text classification technique. We have employed *C4.5* classifier with *1-grams* using the weighting *TF-IDF* like features to represent a web service using the vector space model. A selection of features was carried out in order to reduce the space of representation, obtaining 69 words like relevant features.

The main contributions of this paper are as follows: (a) we present an approach to classify web services using their descriptions and a text classification technique; (b) we populated a web service ontology; (c) and we have used a benchmark collection for testing called OWLS-TC, in which we have demonstrated that our solution outperforms other approaches in terms of average resulting classification.

As future work, we plan to discover non-taxonomic relations or concepts [22] between web services and functionalities, input/output descriptions to enhance web service ontology in order to facilitate discovery and composition of web services.

In addition, semantic similarity between operations and between classified web services is relevant to extract for helping to software developers in web service discovery.

ACKNOWLEDGMENT

This work has been supported by *PRODEP* as part of project no. UAM-PTC-478. Authors thank the *SNI-CONACyT*.

REFERENCES

[1] I. Bluemke, M. Kurek and M. Purwin, "Tool for Automatic Testing of Web Services", *Proc. of the 2014 Federated Conference on Computer Science and Information Systems*, Warsaw, 2014, pp. 1553-1558. doi: 10.15439/2014F93

[2] F. Liu, Y. Shi, J. Yu, T. Wang and J. Wu, "Measuring similarity of web services based on WSDL," in *Proc. of the 2010 IEEE International Conference on Web Services*, Florida, USA, 2010, pp. 155-162.

[3] M. Bravo and M. Alvarado, "Similarity measures for substituting Web services," Web Service Composition and New Frameworks in Designing Semantics: Innovations, pp. 143-170, 2012.

[4] J. Zhang and D. Pan, "Web Service Classification," Dan Pan, Jing Zhang [EB/OL], 2008.

[5] H. Wang, Y. Shi, X. Zhou, Q. Zhou, S. Shao and A. Bouguettaya, "Web service classification using support Vector Machine," in *Proc. of the 22nd IEEE International Conference on Tools with Artificial Intelligence*, Arras, France, 2010, pp. 3-6.

[6] L. Chen, Y. Zhang, Z. L. Song and Z. Miao, "Automatic web services classification based on rough set theory," Journal of Central South University, vol. 20, pp. 2708-2714, 2013.

[7] R. Mohanty, V. Ravi and M. R. Patra, "Web-services classification using intelligent techniques," Expert Systems with Applications, vol. 37(7), pp. 5484-5490, 2010.

[8] L. Yuan-jie and C. Jian, "Web service classification based on automatic semantic annotation and ensemble learning," in *Proc. of the 26th International on Parallel and Distributed Processing Symposium Workshops & PhD Forum*, Shanghai, China, 2012, pp. 2274-2279.

[9] R. Nisa and U. Qamar, "A text mining based approach for web service classification," Information Systems and e-Business Management, pp. 1-18, 2014.

[10] J. Wu, L. Chen, Z. Zheng, M. R. Lyu and Z. Wu, "Clustering web services to facilitate service discovery," Knowledge and information systems, vol. 38(1), pp. 207-229, 2014.

[11] K. Elgazzar, A. E. Hassan and P. Martin, "Clustering WSDL documents to bootstrap the discovery of web services," in *IEEE International Conference on Web Services (ICWS)*, Florida, USA, 2010, pp. 147-154.

[12] Q. Liang, P. Li, P. C. Hung and X. Wu, "Clustering web services for automatic categorization," in *IEEE International Conference on Services Computing*, Bangalore, India, 2009, pp. 380-387.

[13] H. S. Nguyen, S. H. Nguyen and W. Świeboda, "Semantic explorative evaluation of document clustering algorithms," *Proc. of the 2013 Federated Conference on Computer Science and Information Systems*, Krakow, 2013, pp. 115-122.

[14] M. Bravo, J. Rodríguez and A. Reyes, "Enriching Semantically Web Service Descriptions," in *On the Move to Meaningful Internet Systems: OTM Conferences*, Amantea, Italy, 2014, pp. 776-783.

[15] Y. J. Lee and C. S. Kim, "A learning ontology method for restful semantic web services," in *IEEE International Conference on Web Services (ICWS)*, Washington DC, USA, 2011, pp. 251-258.

[16] M. Sabou, "Learning web service ontologies: an automatic extraction method and its evaluation," Ontology learning from text: methods, evaluation and applications, vol. 123, pp. 125-139, 2005.

[17] M. Bravo, J. Rodríguez and J. Pascual, "SDWS: Semantic Description of Web Services," International Journal of Web Services Research, vol. 11(2), pp. 1-23, 2014.

[18] M. Horridge and P. F. Patel-Schneider, "Manchester syntax for OWL 1.1," OWL: Experiences and Directions, Washington, USA, 2008.

[19] R. Quinlan, *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers, 1993.

[20] S. R. Garner, "Weka: The Waikato environment for knowledge analysis," in *Proc. of the New Zealand Computer Science Research Students Conference*, 1995, pp. 57-64.

[21] P. Buitelaar and P. Cimiano, "Ontology learning and population: bridging the gap between text and knowledge," vol. 167, IOS Press, 2008.

[22] P. Szwed, "Concepts extraction from unstructured Polish texts: A rule based approach," *Proc. of the 2015 Federated Conference on Computer Science and Information Systems*, Lodz, 2015, pp. 355-364. doi: 10.15439/2015F280

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 497–502
DOI: 10.15439/2016F296
ACSIS, Vol. 8. ISSN 2300-5963

A Real-Time Audio Compression Technique Based on Fast Wavelet Filtering and Encoding

Nella Romano*, Antony Scivoletto*, Dawid Połap†
*Department of Electrical and Informatics Engineering, University of Catania, Viale A. Doria 6, 95125 Catania, Italy
Email: nromano919@gmail.com, antonyscivoletto@gmail.com
†Institute of Mathematics, Silesian University of Technology, Kaszubska 23, 44-100 Gliwice, Poland
Email: dawid.polap@gmail.com

***Abstract*—With the development of telecommunication technology over the last decades, the request for digital information compression has increased dramatically. In many applications, such as high quality audio transmission and storage, the target is to achieve audio and speech signal codings at the lowest possible data rates, in order to offer cheaper costs in terms of transmission and storage. Recently, compression techniques using wavelet transform have received great attention because of their promising compression ratio, signal to noise ratio, and flexibility in representing speech signals. In this paper we examine a new technique for analysing and compressing speech signals using biorthogonal wavelet filters. In particular, we compare this innovative compression method with a typical VoIP encoding of human voice, underlining how using wavelet filters may be convenient, mainly in terms of compression rate, without introducing a significant impairment in signal quality for listeners.**

***Index Terms*—Wavelet Analysis; Audio Compression; Digital Filters; VoIP; SIP, Quality of Services.**

I. Introduction

SPEECH is a very basic way for people to convey information to each other by means of human voice, within a bandwidth of around 4 KHz. The growth of the computer industry has invariably led to the demand for quality audio data. Analogue audio signals, such as voice speeches, or music, are often represented digitally by repeatedly sampling the waveform and representing it by the resulting quantized samples. This technique is known as *Pulse Code Modulation (PCM)*. PCM is typically used without compression in high-bandwidth audio devices (e.g., in CD players), but compression is essential where the digital audio signal has to be transmitted by means of a communication medium, such as a computer or telephone network [1]. In order to send real-time audio data over a communication link, data compression has be used due to the mismatch with the available link bandwidth [2].

Compression of signals is based on redundancy removal between neighbouring samples or between the adjacent cycles [3]. In data compression, it is desired to represent data by as small as possible number of coefficients within an acceptable loss of quality. Therefore, compression methods rely on the fact that information, by its very nature, is not random but exhibits an intrinsic order and pattern, so that the essence of the information can often be represented and transmitted using less data than would be required for the original signal [2].

Compression techniques can be classified into one of two main categories: *lossless* and *lossy*. Lossless compression works by removing the redundant information present in an audio signal, but preserving its quality and the complete integrity of the data. However, it offers small compression ratios, hence it can be used if we have no stringent requirements; furthermore, it does not guarantee a constant output data rate, since the compression ratio is highly dependent on the input data. On the other hand, one advantage of lossless compression is that it can be applied to any data stream. In lossy coding, the compressed data does not preserve bit-wise equivalence with the original data. The goal of this kind of compression is to maximize the compression ratio or the bit rate reduction, with reduced cost in terms of loss in quality [2].

Compression methods can be classified into three functional categories: *direct methods*, when the samples of the signal are directly handled to provide compression; *parameter extraction methods*, if a preprocessor is employed to extract some features that are later used to reconstruct the signal; *transformation methods*, such as Fourier transform, wavelet transform, and discrete cosine transform. In this latter, the wavelet transform is computed separately for different segments of the time-domain signal at different frequencies. This makes wavelet filtering good for signals having high frequency components for short duration and low frequency components for long duration, such as images, video frames and speech signals [3].

In this paper we show an innovative technique to process and compress an audio signal using a 3.7 biorthogonal wavelet filter, by using thresholding techniques in order to eliminate some insignificant details of the signal, then obtaining a lossy compression that allows us to significantly reduce audio bit-stream length, without compromising the sound quality. While such a technique has been selected because of its remarkable performances, due to the intrinsic nature of the implemented wavelet transforms and filters, it is also possible to implement the coding and decoding pipeline directly on hardware. Therefore, the proposed system not only constitutes an outperforming compression software, but also a possible hardware interface. The latter can be thought both as a consumer-side phone-box or as a provider-side switchboard integrated system.

978-83-60810-90-3/$25.00

II. VoIP Encodings

Nowadays, modern telecommunication is mainly based on the following steps: voice information is digitalised (by sampling), then encoded, and then transmitted as a packets stream [4].

A. Encoding

The encoding process starts by producing digital "rough" results during the sampling phase, then normally reducing the bit rate (so, the bandwidth) of the sampled data through a suitable compression. There may be many coding techniques and among these we can mention:

- *Differences encodes*: if the next sample differs not much from the previous one, then we transmit the difference (which requires a lower number of bits) with respect to the original sample; a typical example is video encoding used in MPEG, which adopts a differential coding both regarding the previous frame and the next one [5].
- *Weighted encodes*: if certain samples are often present within the voice stream, we adopt a convention which codifies them through a smaller number of bits in order to save bandwidth (used e.g. in compression techniques such as ZIP) [5].
- *Loss encodes*: it is based on the principle that, for human ear, certain audio signals are practically ignored. This type of encoding causes such parts of signals to be erased, and the resulting encoding becomes leaner because there are less data to send (this technique is used in MP3 audio compression) [6].

With these techniques we can obtain significant signal compressions. Due to the various application fields, different types of codecs, characterised by different complexity, have been developed. In order to determine which of these should be used in a VoIP service, it is important to take into account the features in terms of bandwidth, encoding delay, and the voice quality reproduced on the receiving part. The main encoding used for the telephony transport on digital lines is the PCM, described in ITU-T G.711 recommendation, which produces a flow of 64 kbps [5]. This encoding is very simple and widespread (for the reasons mentioned above) particularly among telcos. The standard which offers the highest compression, maintaining good voice quality, is the DR SCS (Dual Rate Speech Coding Standard) or G.723. It can go up to speeds of 5.3 kbps, and is commonly adopted on videoconferencing systems over analog lines [5], [7].

Other popular voice coding standards for telephony and packet voice include [5]:

- G.711 - Describes the 64 Kbps PCM voice coding technique outlined earlier; G.711-encoded voice is already in the correct format for digital voice delivery in the public phone network or through Private Branch eXchanges (PBXs);
- G.728 - Describes a 16 Kbps low-delay variation of CELP voice compression;
- G.729 - Describes CELP compression that enables voice to be coded into 8 Kbps streams; two variations of this standard (G.729 and G.729 Annex A) differ largely in computational complexity, and both generally provide speech quality as good as that of 32 Kbps ADPCM.

B. Packets encoding

While the previous steps (sampling and coding) can be partially used even on a digital telephone network, the packaging operation is peculiar to packet networks. A package, by its nature, is composed by a series of headers so that the package can reach properly the destination. These headers can not be eliminated; this fact implies they must be present independently from the number of packets sent and their size. In voice over IP, the typical header has a size of 58 bytes (18 bytes Ethernet, 20 bytes IP, UDP 8 bytes, 12 bytes RTP): if each package carried only one data byte, the efficiency would become equal to approximately 1.7%, as a voice stream at 64kbps would generate a traffic of 3.7Mbps. Unfortunately, it is not possible to use packages of arbitrary size, because by decreasing the package encoding time, the delay would increase. A reasonable packaging delay values are in the order of 20-40 ms [5].

Compression efficiency is possible, in this scenario, to work around the problems as well as with respect to the used bandwidth, if the compression is achieved by other coding techniques, such as wavelet compression.

III. Wavelets

In recent years, wavelet theory has been developed as a unifying framework for a large number of techniques for wave signal processing applications, such as multiresolution analysis, sub-band coding and wavelet series expansions [8]. The idea of analysing a signal at various time frequency scales with different resolutions has emerged independently in many mathematics, physics and engineering fields. In fact wavelet analysis is capable of revealing aspects of data that other signal analysis techniques cannot take into account, especially when breakdown points, discontinuities in higher derivatives, and other self-similarity occur. Furthermore, because it let us obtain a different representation of data than those offered by traditional techniques, it can help us to efficiently compress or de-noise a signal without any appreciable degradation [9], [10], [11]. A significant advantage of using wavelets for speech coding is that the compression ratio can easily be optimised, while most other techniques have fixed compression ratios keeping all the other parameters constant.

Wavelet analysis is the breaking up of a signal into a set of scaled and translated versions of an original wavelet. The wavelet transform of a signal decomposes the original signal into wavelets coefficients at different scales and positions.

A. Wavelets in continuos domain

Wavelets are continuos "basis" functions constructed in order to satisfy certain mathematical properties. A wavelet

is defined as a function $\psi(x)$ which must be subject to the following constrains:

1) $\int_{-\infty}^{+\infty} \psi(x)\, dx = 0$

2) $\| \psi(x) \|^2 = \int_{-\infty}^{+\infty} \psi(x)\psi^*(x)\, dx = 1$

Moreover, a whole family of wavelet functions can be obtained by just shifting and scaling as:

$$\psi_{j,k} = \sqrt{2^j}\ \psi(2^j t - k), \quad i, j \in \mathbb{N} \tag{1}$$

The idea behind wavelets is that by stretching and translating one such wavelet function $\psi(t)$ (also called mother wavelet), we can represent a signal $f(t) \in L^2(\mathbb{R})$ as:

$$f(t) = \sum_{j,k} b_{j,k}\ \psi_{j,k}(t) \tag{2}$$

where $b_{j,k}$ are called *wavelet coefficients* of the signal f in the wavelet basis given by the inner product of $\psi_{j,k}$ [12]. The wavelet coefficients represent the original signal in the wavelet domain [10]. An advantage of wavelet transforms is that the windows vary.

B. Discrete Time Case

In the discrete time case, two methods have been developed independently, namely *sub-band coding* (widely used in voice compression) and *multiresolution signal analysis*.

1) Multiresolution analysis: With this method we can derive a lower resolution signal by lowpass filtering with a half-band low-pass filtering having impulse response $g(n)$. This results in a signal $y(n)$ where

$$y(n) = \sum_{k=-\infty}^{k=+\infty} h(k) * f(2n - k) \tag{3}$$

Now based on subsampled version of $f(n)$ we want to find an approximation, $a(n)$, of the original signal $f(n)$: this is done by inserting a zero between every sample, because we need a signal at the original scale for comparison. Therefore, there is some redundancy in the number of samples, which will be proven useful for compression.

2) Sub-band coding schemes: The idea behind this technique is based on the following methodology: a pair of filters are used, i.e. a low pass and a high pass filter; we decompose a sequence $X(n)$ into two subsequences at half rate, or half resolution, and this by means of an orthogonal filter. This process can be iterated on either or both subsequences. In particular to obtain finer frequency resolution at lower frequencies, we iterate the scheme on the lower band only (see Figure 1). Thus, applying wavelet transform on a speech signal helps in compressing it to meet the stringent demands of bandwidth consumption while maintaining the quality and integrity of a signal [10].

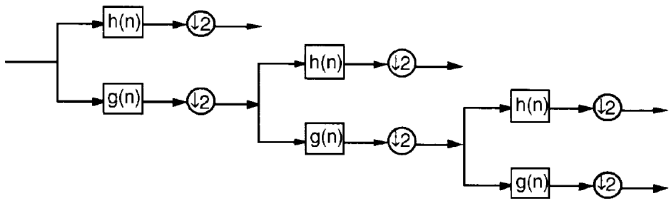


Fig. 1: Filter bank tree of discrete wavelet transform implemented with two discrete time filters (low-pass and high-pass filter, respectively)

C. Biorthogonality

The wavelet filter used in this work for audio signal compression is the 3.7 biorthogonal wavelet transform. A biorthogonal wavelet is a wavelet where the associated wavelet transform is invertible but not necessarily orthogonal; the first number indicates the order of the synthesis filter, while the second number refers to the order of the analysis filter. Even if frequency responses from biorthogonal filters may now not show any symmetry and the energy of the decomposed signal may also not equal the energy of the original signal, there are some reasons behind using biorthogonal wavelets, such as their compact support: this means we do not require IIR filters to approximate the signal. This fact is good because IIR filters are more difficult to treat and they also require extra computing power [12].

IV. GPU Parallel Computing with CUDA

The advent of multicore CPUs and many core GPUs have allowed the development of applications that transparently scale to take advantage of the increasing number of processor cores. Recently, the parallel programming field has seen an increasing success with the development of a novel technology, called Common Unified Device Architecture (CUDA).

A. GPU architecture

A CUDA-enabled GPU is composed of several MIMD (multiple instruction multiple data) multiprocessors that contain a set of SIMD (single instruction single data) processors. Each multiprocessor has a shared memory that can be accessed from each of its processors, and also shares a bigger global memory. Shared memory buffers reside physically on the GPU as opposed to residing in off-chip DRAM, so because of this, the latency to access shared memory tends to be far lower than typical buffers [13].

In CUDA programming model, an application consists of a *host* program that executes on the CPU and other parallel *kernel* programs executing on the GPU. A kernel program is executed by a set of parallel threads, where a thread is a subdivision of a process in two or more sub-processes, which are executed concurrently by a single processor. The host program can dynamically allocate device global memory to the GPU and copy data to/from such a memory from/to the memory on the CPU. Moreover, the host program determines

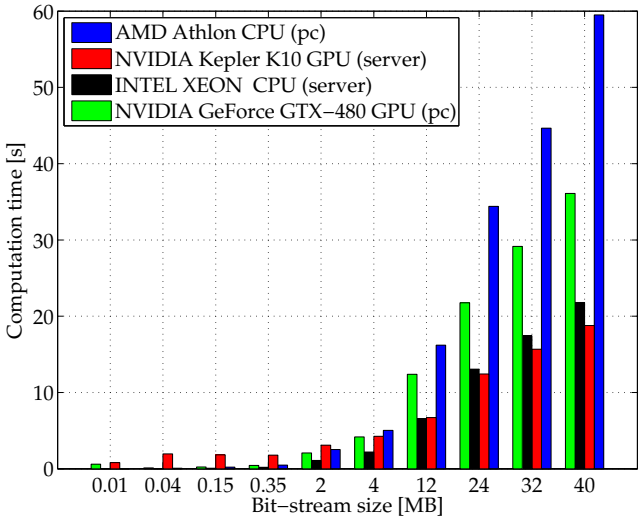


Fig. 2: Comparison among parallel and sequential programs on PC and Server

the number of threads used to execute kernel. Set of threads are grouped in blocks, and each block shares its own memory. It follows also that interactions between CPU and GPU should be minimised in order to avoid communication bottlenecks and delays due to data transfers [14], [15], [16].

B. Performing wavelet transforms on GPU

Classical wavelet transform implementations can be too computationally costly to cater for real time systems, hence we have investigated a parallel and fast approach to the problem at hand. As we have shown in [12], it is possible to adopt a fast GPU-oriented processing system to obtain the wavelet decomposition within the speed requirements of a VoIP service.

Figure 2 shows CPUs and GPUs timing performances for a conventional host, named PC, and for a Server. PC has been equipped with an AMD Athlon 64X2 having CPU clock frequency up to 2.9 GHz and a NVIDIA GeForce GTX 480 GPU; in Server, instead, has been equipped with an Intel Xeon CPU having a clock frequency up to 3.4 GHz, and a NVIDIA Tesla Kepler K10 GPU with 1536 cores.

V. Wavelet Compression

The goal of this work is to show the potential in terms of audio signals compression processed by wavelet filters. In particular, here we used a 3.7 biorthogonal wavelet filter. We will evaluate the wavelet compression efficiency and we will compare it with encoded audio trace through the typical PCM-16-bit VoIP standard at frequency of 8KHz. This analysis looks at both the quantitative aspects of encoding (the *compression factor*), and the audio quality perceived by listeners.

A. Experimental Setup

Initially an audio signal has been imported as a simple numerical array, whose size and the sampling frequency are fixed. In our setup, the audio signal has a duration of about 8 seconds and is sampled at a frequency of 22050 Hz. As explained in Section III, we know that, when we consider details with increasingly sample rate, the wavelet filter selects audio signal components which are at higher frequencies and which are characterised by a short time decay. These characteristics are typical of noise and background, therefore removable from the audio signal. It is then convenient to intervene more strongly on this portion of the information when coming to lossy compression. This result is achieved by appropriately setting thresholds. The detailed coefficients are removed from the last two sub-bands.

The result of superimposition between original and compressed signal is shown in Figure 3b. Once the two last sub-bands are removed, the coefficients set is compressed, taking also advantage of the zero-coefficients redundancy. As explained in Section III-B2, the sub-band coding process performs a dyadic signal processing. In other words, by the first subsampling cycle, we get a details vector, called *d1*, whose size is equal to half of original signal's size; for the second scale, a details vector *d2* is achieved, whose size will be equal to half of *d1* vector's size, i.e. one quarter (25 %) of the original signal size. Moreover, since we have suppressed the last two bands of the original signal, we take into account only the first quarter of the coefficient vector, because all other elements are consecutive zeros that were generated during the compression phase. Furthermore, the existence of other non-consecutive zeros inside the coefficient vector helps the compression further. Finally, we have compressed the coefficients using a gzip compressor. This final result is then transmitted by the VoIP client.

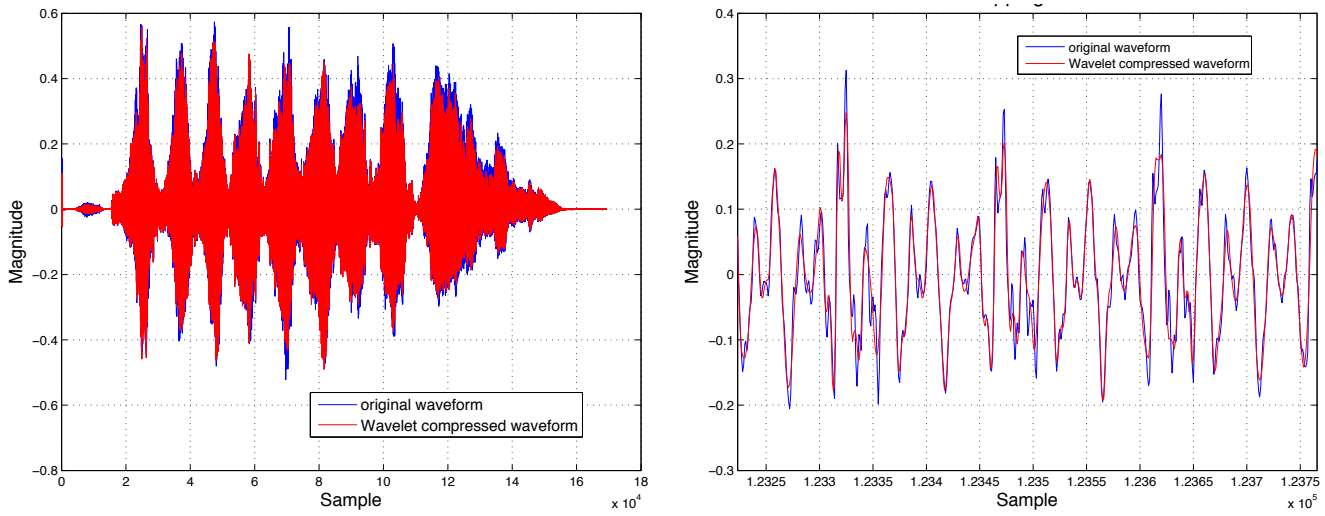
The reconstruction phase proceeds reversely to the compression stage: the starting point is the compressed data; the data stream is then decompressed with gzip. The decompressed data are initially incomplete, because we have to add zeros in order to reach the original size of the coefficients vector. In this way we can reconstruct the original digital signal.

B. Streams

The presented procedure is recursively applied for audio results. In this paper, we have analysed audio tracks that have a minimum duration of 8 seconds and a maximum of 10 minutes. For each of them, we have taken into account:

- the size of starting file, saved in .wav format and sampled at 22050 Hz;
- the size of file encoded using the 16-bit PCM to 8 KHz VoIP standard;
- the size of gzip-compressed file after wavelet filtering.

From the comparison among these three quantities, for different audio traces durations, the final result is shown in Figure 4. By observing Figure 4, it is possible to note a marked difference in the occupied bandwidth between the first two audio signals, namely the original and the VoIP one, compared to the proposed results obtained by wavelet compression. We have determined that the VoIP standard can give a compression factor of about 4, while the wavelet filtering allows us to obtain



(a) Global view of superimposition between original and compressed signal

(b) Superimposition between original and compressed signal in detail

Fig. 3: Superimposition between original and wavelet compressed signal

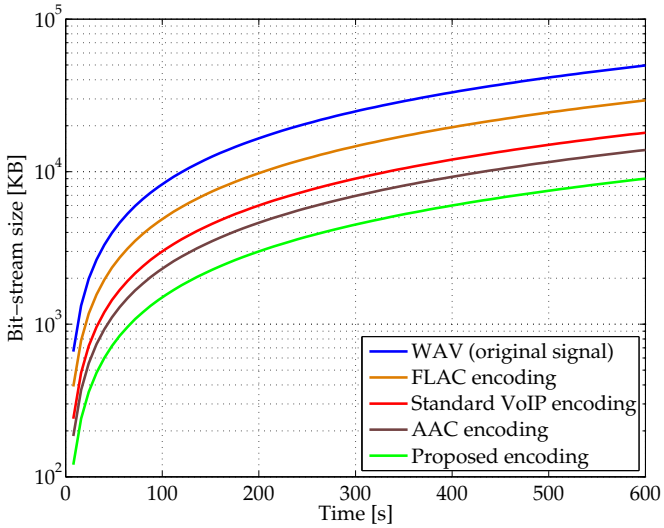


Fig. 4: Comparison among different standard audio data encodings with respect to our proposed wavelet compressed format

a compressed file approximately 10 times smaller than the original one. Therefore, our approach can improve the standard VoIP protocol compression.

VI. Related Works

Some works in the literature have analysed similar problems regarding audio compression techniques, and the relative loss of quality as well as regarding the delivery of content with limited network resources[17]. In [18] the authors use optimal adaptive wavelet selection and wavelet coefficients quantization procedures together with a dynamic dictionary approach. It takes advantages of the masking effect in human hearing. They minimise the number of hits required to represent each frame of audio material at a fixed distortion level. In [19], the authors show that the WP decomposition provides sufficient resolution to extract the time-frequency characteristics of the input signal and the WP based audio compressor provides transparent sound quality at compression rates comparable to the MPEG compressor with less than one third of the computational effort.

In [20], the authors have proposed a novel filter bank scheme which switches between a MDCT and a wavelet filter bank based on signal characteristics. A tree structured wavelet filter bank with properly designed filters offers natural advantages for the representation of non-stationary segments, such as attacks.

Interesting results are usually achieved when using a neural network approach to retain the relevant information on huge amount of sparse data [21], [22], [23], and some experiments have been performed on the compression of images [24]. Moreover, several solutions have been proposed for separating the components representing neural networks and the logic using them [25], [26], [27], [28]. In [29], the authors have proposed to apply wavelet analysis and audio compress technology in audio watermarking. Their approach performs a wavelet transform to determine the local audio properties. In [30] the authors have provided a novel scheme to join the wavelet packets and perceptual coding to construct an algorithm that is well suited to high-quality audio transfer for internet and storage applications. Finally, in [31] the authors examined how a compression method employs an approximation of a psychoacoustic model for wavelet packet decomposition. It has a bit rate control feedback loop particularly well suited to matching the output bit rate of the data compressor to the bandwidth capacity of a communication channel.

VII. Conclusion

In this paper we have investigated the compression potential of a 3.7 biorthogonal wavelet filter based technique for VoIP telecommunication. We have seen how, by using this compression technique, it is possible to obtain a greater compression factor than that obtainable with other traditional encodings. This fact is advantageous because the obtained audio stream is more efficiently compressed and requires less bandwidth for transmission.

The wavelet filtering is also interesting because makes it possible to implement it in a physical device. In fact, the proposed approach can be realised in hardware, basing on digital signals processing and compressing through wavelet filters. Furthermore we have to consider that nowadays modern 3G technology, widely used in mobile telephony, provides users with a quite limited data transmission bandwidth. For this reason wavelet compression approach, featured by a compression factor greater than ones used in other codings, meets practical needs of people who want to exchange data in a wireless medium.

Acknowledgements

This work has been supported by Project Knowledge Education Development financed by the European Social Fund POWR.03.03.00-00-P001/15.

References

[1] D. Monro, "Audio compression," August 2004, US Patent App. 10/473,649.

[2] O. O. Khalifa, S. H. Harding, and A.-H. Abdalla Hashim, "Compression using wavelet transform," *International Journal of Signal Processing*, vol. 2, no. 5, pp. 17–26, 2008.

[3] H. Kaur and R. Kaur, "Speech compression and decompression using DWT and DCT," *Int. J. Computer Technology and Applications*, vol. 3, no. 4, pp. 1501–1503, August 2012.

[4] F. D. Rango, M. Tropea, P. Fazio, and S. Marano, "Overview on VoIP: Subjective and Objective Measurement Methods," *International Journal of Computer Science and Network Security*, vol. 6, no. 1, pp. 140–153, 2006.

[5] J. Davidson, *Voice over IP fundamentals*. Cisco press, 2006.

[6] [Online]. Available: http://www.ucci.it/docs/ICTSecurity-2003-18b

[7] M. Mehić, M. Mikulec, M. Voznak, and L. Kapicak, "Creating covert channel using sip," in *Federated Conference on Computer Science and Information Systems (FedCSIS)*. IEEE, 2014, pp. 182–192.

[8] S. Mallat, *A wavelet tour of signal processing: the sparse way*. Academic press, 2008.

[9] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Transactions on Information Theory*, vol. 41, no. 3, pp. 613–627, 1995.

[10] V. Malik, P. Singh, A. kumar Singh, and M. Singh, "Comparative analysis of Speech Compression on 8-bit and 16-bit data using different wavelets," *International Journal of Computer Trends and Technology (IJCTT)*, vol. 4, no. 5, May 2013.

[11] M. B. Abdallah, J. Malek, A. T. Azar, H. Belmabrouk, J. E. Monreal, and K. Krissian, "Adaptive noise-reducing anisotropic diffusion filter," in *Federated Conference on Computer Science and Information Systems (FedCSIS)*. IEEE, 2015, pp. 1–28.

[12] A. Scivoletto and N. Romano, "Performances of a parallel cuda program for a biorthogonal wavelet filter," in *Proceedings of the International Symposium for Young Scientists in Technology, Engineering and Mathematics (System)*, 2016.

[13] J. Sanders and E. Kandrot, *CUDA BY EXAMPLE, An Introduction to General-Purpose GPU Programming, NVIDIA*.

[14] F. Bonanno, G. Capizzi, S. Coco, C. Napoli, A. Laudani, and G. Lo Sciuto, "Optimal thicknesses determination in a multilayer structure to improve the spp efficiency for photovoltaic devices by an hybrid femcascade neural network based approach," in *International Symposium on Power Electronics, Electrical Drives, Automation and Motion (SPEEDAM)*. IEEE, 2014, pp. 355–362.

[15] M. Wozniak, D. Polap, G. Borowik, and C. Napoli, "A first attempt to cloud-based user verification in distributed system," in *Asia-Pacific Conference on Computer Aided System Engineering (APCASE)*. IEEE, 2015, pp. 226–231.

[16] Z. Marszalek, M. Wozniak, G. Borowik, R. Wazirali, C. Napoli, G. Pappalardo, and E. Tramontana, "Benchmark tests on improved merge for big data processing," in *Computer Aided System Engineering (APCASE), 2015 Asia-Pacific Conference on*. IEEE, 2015, pp. 96–101.

[17] K. Kaczmarski, M. Pilarski, B. Banasiak, and C. Kabut, "Content delivery network monitoring with limited resources," in *Federated Conference on Computer Science and Information Systems (FedCSIS), 2013*. IEEE, 2013, pp. 801–805.

[18] D. Sinha and A. H. Tewfik, "Low Bit Rate Transparent Audio Compression using Adapted Wavelets," *IEEE Transactions On Signal Processing*, vol. 41, no. 12, December 1993.

[19] M. Black and M. Zeytinoglu, "Computationally efficient wavelet packet coding of wide-band stereo audio signals," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 5. IEEE, 1995, pp. 3075–3078.

[20] D. Sinha and J. D. Johnston, "Audio compression at low bit rates using a signal adaptive switched filterbank," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 2. IEEE, 1996, pp. 1053–1056.

[21] C. Napoli, G. Pappalardo, M. Tina, and E. Tramontana, "Cooperative strategy for optimal management of smart grids by wavelet rnns and cloud computing," *IEEE Transactions on Neural Networks and Learning Systems*, (in press) 2015.

[22] C. Napoli, G. Pappalardo, E. Tramontana, R. Nowicki, J. Starczewski, and M. Woźniak, "Toward work groups classification based on probabilistic neural network approach," in *Proceedings of Artificial Intelligence and Soft Computing*, ser. Lecture Notes in Computer Science, vol. 9119. Springer, 2015, pp. 79–89.

[23] C. Napoli, G. Pappalardo, and E. Tramontana, "A mathematical model for file fragment diffusion and a neural predictor to manage priority queues over bittorrent," *International Journal of Applied Mathematics and Computer Science*, vol. 26, no. 1, pp. 147–160, 2016.

[24] M. Wozniak, C. Napoli, E. Tramontana, and G. Capizzi, "A multiscale image compressor with rbfnn and discrete wavelet decomposition," in *International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2015, pp. 1219–1225.

[25] C. Napoli and E. Tramontana, "An object-oriented neural network toolbox based on design patterns," in *Proceedings of the International Conference on Information and Software Technologies (ICIST)*, ser. CCIS, vol. 538. Springer, 2015, pp. 388–399.

[26] A. Calvagna and E. Tramontana, "Delivering dependable reusable components by expressing and enforcing design decisions," in *Proceedings of IEEE Computer Software and Applications Conference (COMPSAC) Workshop QUORS*, Kyoto, Japan, July 2013, pp. 493–498.

[27] G. Capizzi, G. L. Sciuto, C. Napoli, E. Tramontana, and M. Woźniak, "A novel neural networks-based texture image processing algorithm for orange defects classification," *International Journal of Computer Science & Applications*, vol. 13, no. 2, pp. 45–60, 2016.

[28] R. Giunta, G. Pappalardo, and E. Tramontana, "Superimposing roles for design patterns into application classes by means of aspects," in *Proceedings of ACM Symposium on Applied Computing (SAC)*. Riva del Garda, Italy: ACM, March 2012, pp. 1866–1868.

[29] L. Cui, S.-x. Wang, and T. Sun, "The application of wavelet analysis and audio compression technology in digital audio watermarking," in *Proceedings of International Conference on Neural Networks and Signal Processing*, vol. 2. IEEE, 2003, pp. 1533–1537.

[30] P. Srinivasan and L. H. Jamieson, "High-Quality Audio Compression Using an Adaptive Wavelet Packet Decomposition and Psychoacoustic Modeling," *IEEE Transactions On Signal Processing*, vol. 46, no. 4, April 1998.

[31] K. Dobson, N. Whitney, K. Smart, P. Rigstad, J. Yang *et al.*, "Method and apparatus for wavelet based data compression having adaptive bit rate control for compression of digital audio or other sensory data," oct 1998, US Patent 5,819,215.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 503–512
DOI: 10.15439/2016F334
ACSIS, Vol. 8. ISSN 2300-5963

Supercombinator Set Construction from a Context-Free Representation of Text

Michal Sičák, Ján Kollár
Techical University of Košice, Department of Computers and Informatics
Letná 9, 042 01 Košice, Slovakia
Email: {michal.sicak, jan.kollar}@tuke.sk

***Abstract*—Grammars might be used for various other aspects, than just to represent a language. Grammar inference is a large field which main goal is the construction of grammars from various sources. Written text might be analysed indirectly with the use of such inferred grammars. Grammars acquired from processed text might grow into large structures as the inference process could be continuous. We present a method to decompose and store grammars into a non-redundant set of lambda calculus supercombinators. Grammars decomposition is based on their structure and each distinct element is stored only once in such a structure. We present a method that can create such a set from any context-free grammar. To prove this and to show the possible applications in the field of natural language processing we present a case study performed on samples from two books. Those samples are the entire Book of Genesis from The King James Bible and the first 24 chapters of War and peace by Tolstoy. We obtain context-free grammars with the Sequitur algorithm and then we process them with our method. The results show significant decline in the number of grammar elements in all cases.**

I. Introduction

REPRESENTATION of extracted information from text is a question that correlates with cognitive science as researchers try to emulate the processes that runs in our head [1]. Natural languages might differ from their formal counterparts, but they can be described with the same formal theory as Chomsky [2] pointed out almost 60 years ago.

From a more pragmatic stand point of view, extracted information may be represented in a grammar form. This is usually the goal of the grammar inference (or grammar induction) field. As Gold in [3] stated, only superfinite grammars can be inferred from a text. By text Gold means a set of positive samples, i.e. samples that belong to the language that inferred grammar is representing. Yet recent advances in this field has shown, that although in a strict formal way Gold theorem still holds, by using heuristic, statistical or evolutionary methods, we are able to infer more complex grammars, like context-free grammars (CFG). De la Higuera presents a review of possible inference methods in [4]. And not only formal languages can be inferred. We are able to perform inference on natural languages as well, as Onnis, Waterfall and Edelman point out [5].

As we infer grammars of large quantities of text, we may indeed obtain large grammars. If those texts are similar, for example we are processing books written in the same language by the same author, then the resulting grammars are in a risk of having a large number of rules and lots of similar information stored separately. This phenomena is called structural explosion. Where structurally similar, but symbolically different rules of grammar are inferred and then stored each separately. In this paper we present a way how to represent any CFG grammar in a non-redundant form. This form is a single structure composed of lambda calculus supercombinators. In the section II we reason why such a form is useful and how it relates to the field of natural language processing.

The main contributions of this paper are:

- We present the entire process of supercombinator set construction from any CFG grammar. The theoretical background is presented in the section III and the detailed description itself is in the section IV. It is a multi step process, each step is explained in a separate section accompanied with appropriate examples.
- Our approach has been tested on larger scale experiments that we present in the section V. We used Sequitur algorithm [6] to create context-free grammars from book samples and then we run our process on them. We present the results, that show significant reduction of grammar elements. This indicates the prevention of structural explosion.
- In the section VI we discuss the possibilities of text analysis that results from the usage of Sequitur algorithm and our supercombinator set acquisition process. We point out that even from the simple grammars, that are the result of Sequitur algorithm, we can extract useful structures contained in a natural language text.

II. Motivation

The basic idea behind this work stems from the engineering discipline of grammarware [7]. Simply put, we can use grammars for other purposes, than just for the representation of a language. With our approach we can decompose and store a CFG into a non-redundant form. In this section we explain, what good that process brings and how it relates to the field of natural language processing.

Why do we need to store a grammar inside a non-redundant structure? If we were given a small, predefined, static CFG, that we use only as a base for a parser for example, than our process could indeed be redundant itself. However, we can

This work was supported by project KEGA 031TUKE-4/2016 "Integrating software processes into the teaching of programming".

obtain a CFG from previously unknown or rather unprocessed text. And we might want to process lots of such texts for purposes like information extraction, language learning, text analysis, grammar inference or others. And consider that we would want to obtain one structure from all of those texts. If we were to keep those grammars stored in a basic CFG form, we could find ourselves buried under structurally same, yet semantically different rules. This is so especially if we are to process grammars that are large yet their rules are rather simple, as we show in the section V. Such rules might differ only in symbols but all grammar operations are the same. So instead of having a sequence of two different symbols stored for each pair of symbols separately, we can store one structural representation of two symbol sequence and link it with its respective symbols. Those symbols are also stored non-redundantly, so in case we have large terminal symbols (like words), having links to them is more memory efficient.

Our approach is based on the lambda calculus principle. The idea of the application and abstraction that is inherent to the lambda calculus offers a way to represent grammar rules that are separated from their terminal symbols. Therefore the entire structure is represented as one big set of supercombinators. The applications of those supercombinators on other supercombinators might be perceived as links. In that case we may view such a set as some network like structure. We explain this principle in further detail in [8] and [9], but also here, in the section III.

Now, how our work relates to the natural language processing? Well, the supercombinator form has few advantages. The non-redundancy has already been mentioned. The second advantage is that it describes and decomposes the structure of grammar rules. This for example opens the possibility to a form of text structure analysis. By searching for similar structures of text, combined with appropriate grammar inference mechanism, we might obtain information that may be used for various forms of analysis, like author or style identification. However, this possibility is out of scope of this paper.

Our process strongly relates to the field of grammar inference. Should we apply our process on a finite string of symbols, we would obtain just two supercombinators, out of which one would be long and would represent entire sequence of words. And that is not what we want, since we want to capture structure. So we need to start from a grammar form.

In this paper we are processing simple (in the structure, not in the size) CFGs obtained by Sequitur algorithm application (see section V) on a text written in a natural language, in our case the English language. We do not need to use Sequitur algorithm only. There are many ways to infer (or induce) a grammar from any text, see section VII. Therefore we can use any CFG grammar in our process.

III. Background

First of all, we need to explain how the entire process of supercombinator form acquisition works. The first version of this process has already been explained in [9]. There we have used only regular languages to create our supercombinator set. And thus the process was left in the theoretical space, i.e. we have not used real world case study. Our other work [8] presented an introduction to context-free languages application via higher order principle. In this section we build up on this principle towards building complete set of supercombinators from any CFG.

A. Enriched Lambda Calculus

Ordinary lambda expression e is defined by the rule (1).

$$e \to a \mid x \mid e\ e \mid \lambda x.e \tag{1}$$

Where a represent any constant, x a lambda variable, $e\ e$ is lambda application and $\lambda x.e$ is lambda abstraction. We can enrich this simple definition with grammar operations, thus the result of lambda expression reduction would not be a single value but the grammar expression, either regular or, as we later show, even context-free. As an example for regular-language-enriched lambda calculus have an expression (2).

$$L = \lambda x_1.\ \lambda x_2.\ x_1 \mid x_2 \tag{2}$$

This expression takes two variables as arguments and results into a regular expression, that consists of the alternative operation applied on both arguments. Therefore lambda application $L\ a\ b$, where a and b are terminal symbols, yields a regular expression $a \mid b$. We may notice, that we have used an infix notation for the alternative operation. This is only a syntactic sugar however.

To enrich our basic lambda calculus definition (2) we just add a regular expression option as another alternative. The formal definition of that regular expression is shown in (3).

$$r \to a \mid r_1 \mathbf{+} \ldots \mathbf{+} r_n \mid r_1 \mid \ldots \mid r_n \mid (r)^{\mathbf{*}} \mid (r) \tag{3}$$

The first element represents terminal symbol. Then the structures of concatenation, alternative and Kleene closure meta-operations are defined. The final element represents ordinary bracketing. Note, that the operators of regular expression themselves are depicted in a bold font, so we can distinguish them from the meta-operators. Concatenation is usually depicted without its operator or with $'.'$ operator. We have used the $'+'$ operator as the dot is already used in the lambda abstraction. Also, the expression $r_1 + r_2$ is equal to the expression $r_1\, r_2$. Should we use operator-less notation in our extended lambda calculus notation however, it could cause a confusion between a lambda application and the concatenation itself. Therefore we are going to stick with the plus operator should we use the concatenation inside of a lambda expression.

The fact, that we may define lots of different and specialized operations for regular expressions is accounted for in this paper. We are not restricting our algorithm for supercombinator form construction with predefined static regular expression operations (i.e. only those three defined in (3)). We use an abstract function that acts as a placeholder for any operation. This idea is further explained in the section IV-A.

We show in Tab I a supercombinator set obtained from the expression $a\,b\,|\,(c)^*$. This expression serves as a good example, since it contains the alternative, concatenation and

TABLE I
SUPERCOMBINATOR SET FOR GRAMMAR $ab|(c)^*$.

Supercombinators			Arguments
L^0	=	$\lambda x_1.\ x_1$	$\{\ a,\ b,\ c\ \}$
L^1	=	$\lambda x_1.\ L^0\ x_1$	$\{\ a\ \}$
L^2	=	$\lambda x_1.\ \lambda x_2.\ L^1\ x_1 + L^0\ x_2$	$\{\ a\ b\ \}$
L^3	=	$\lambda x_1.\ (L^0\ x_1)^*$	$\{\ c\ \}$
L^4	=	$\lambda x_1.\ \lambda x_2.\ \lambda x_3.\ L^2\ x_1\ x_2 \mid L^3\ x_3$	$\{\ a\ b\ c\ \}$

closure operations. The set itself was obtained with the use of process defined in [9]. The process in this paper is slightly different, but any actual difference is pointed out.

The important accompanying part of any supercombinator is its permissible argument string set. There may be more than one permissible argument string for each supercombinator as we can see in the case of L^0. Al three terminal symbols are possible arguments in this case. Only one argument string is allowed for any other supercombinator. We obtain the original expression $ab|(c)^*$ by β-reduction of the supercombinator L^4 with its argument string abc. The original expression would not be the result, if we allow any other argument strings to accompany that supercombinator. Note, that the permissible arguments are represented in the memory only once. They are connected with their supercombinators with the use of links. Argument strings are necessary for the reconstruction of original expression, any other argument string would lead to different expressions being created.

Supercombinators in Table I represent a structurally decomposed form of regular expression $ab|(c)^*$. Some of those supercombinators are applied more than once, like L^0. This is the solution to the structural explosion, since no multiple occurrences of equal supercombinators exist. There exists one top supercombinator for every expression It's the one by which β-reduction we obtain the original expression. In this case it's the L^4. The supercombinator form is reusable. So if we were to process more expressions, all already existing supercombinators would be reused. Only new arguments links would be added in that case. A simple example: if we were to add expression consisting of one symbol, say d, the supercombinators L^0 argument set would be enriched by the d symbol and at the same time it would become top supercombinator for that expression (but only if used with the d symbol).

Each supercombinator represents a part of the entire expression structure. As we see in Table I, the identity function is represented by L^0, a sequence of two variables by L^2 or a closure over one variable by L^3.

B. Higher Order Regular Expressions

We have published a paper under this name [10] where we noted that the difference between regular and context-free expressions is not really that significant. And that we may view context-free expressions as higher order regular expressions, i.e. expressions that may take another expression as an argument.

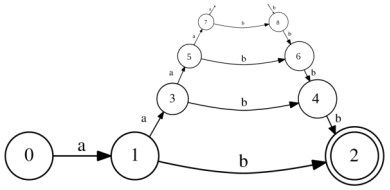


Fig. 1. The infinite state automaton obtained from higher order expression $A \rightarrow aAb \mid ab$.

The simple example is shown in the expression (4).

$$A \rightarrow aAb \mid ab \tag{4}$$

This expression represents the language $a^n\, b^n$, standard example of a nonregular language. The expression (4) is in BNF form, but we may view it as an expression, where we have an alternative of two subexpressions. The first one is a sequence of three symbols. The middle symbol represents a jump inside another iteration of expression evaluation. Expression 4 is represented by the automaton depicted in Fig. 1. It's an infinite state automaton. The important fact here is the idea that there is not that big of a gap between regular and context-free grammars.

This idea of higher order expression translates into our enriched lambda calculus easily. The higher order jump is nothing else but an ordinary lambda application. The first element in the alternative of expression (4) can be translated into the following lambda expression:

$$L^{aAb} = \lambda x_1.\lambda x_2.\ x_1\ +\ (L^A\ x_1\ x_2)\ +\ x_2 \tag{5}$$

Where L^A is the top supercombinator for the expression (4), from which the supercombinator L^{aAb} is being applied. In this scenario, both supercombinators have the same arguments, so there is no need to extend the possible argument set for supercombinators in the lower levels of hierarchy.

It's important to note, that supercombinator (5) is not in the final, complete form, since it applies directly its arguments to the result. In the final form, argument applications are always translated into the identity supercombinator $L^0 = \lambda x.\ x$.

C. Basic Idea

We have presented the idea of regular-language-enriched lambda calculus and the way to view CFGs as context-free expressions or higher order regular expressions. Now we are going to show the outline of our process, where we translate those expressions into a supercombinator set.

Every context-free expression is based on a set of context-free grammar rules. Under the BNF definition (or EBNF if were tu use closure and option operations), these rules have form of $A \rightarrow r$, where A designates the nonterminal and r an expression into which the nonterminal is transformed. The expression has a form as defined in (3) with the addition of nonterminal symbol as a possible element.

Any nonterminal symbol may have more rules that define its expansion. But those rules can be merged together with

the alternative operation without any consequences. , process in (6). Our approach requires all such expressions to be merged together.

$$\left.\begin{array}{l} A \ \rightarrow ab \\ A \ \rightarrow (c)^* \end{array}\right\} \Rightarrow \ A \rightarrow ab \mid (c)^* \qquad (6)$$

There exist one top supercombinator for every expression. This is the supercombinator from which we are able to reconstruct the original expression by applying it to its permissible argument string. We may extract that string even before we start to transform the expressions. This is an important property for implementing higher order jumps.

Simple outline of our transformation process is:

1) First step is to transform the input into the internal expression tree form. Each rule is transformed into separate tree, thus we obtain a set of trees, precisely one tree per nonterminal. Each tree is named after the nonterminal, that it represents.
2) Then we obtain the list of permissible arguments for every tree in the set. These lists do not only include arguments obtained by searching the tree for terminal symbols, but also include all terminals inside each nonterminal accessible from the current expression tree. Each terminal symbol occurs in every final argument string precisely once. No duplicates are allowed.
3) Each expression tree is transformed into separate, independent set of supercombinators. Each supercombinator of that set has exactly one argument string, since they haven't been merged yet. Every occurrence of nonterminal in expression is replaced by temporary supercombinator that points to that nonterminal. We have already acquired their permissible argument strings in the previous step.
4) Equal supercombinators are merged for each expression separately. Supercombinators now may contain more than one argument string in their permissible argument set. This process is done for each expression independently, thus may be performed in parallel.
5) Temporary supercombinators representing other nonterminals are now replaced by top supercombinators of expression represented by those nonterminals.
6) Equal supercombinators from all expressions are new merged together. The result of this is a single supercombinator set that represents a structurally decomposed original grammar.

IV. Supercombinator Set Acquisition from Context-free Grammar

We present the details of the transformation process in this section. All the steps of transformation are illustrated with appropriate examples.

A. Tree Creation

We start our process in the similar fashion as in our previous work, by expression transformation into a tree form. However, the trees used in the current process are different.

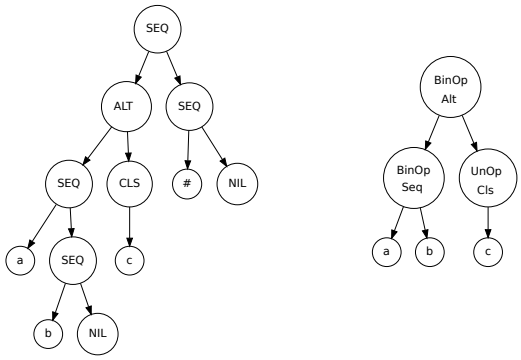


Fig. 2. The old and the new type of expression tree for $a\,b \mid (c)^*$.

Listing 1. Type of a context-free expression tree.

```
data Tree = Leaf Term | Jump NTerm
   | UnOp OpName Tree | BinOp OpName Tree Tree
   | MultiOp OpName [Tree]
```

As mentioned earlier, we tried to separate the operations from the basic structure. While in our previous work, the operations themselves were part of a tree as nodes, now they have been abstracted away.

We have selected $A \rightarrow a\,b \mid (c)^*$ as an example expression. Its depiction on Fig. 2 shows the old and the new type of tree. This figure immediately shows the difference in node amounts. The current tree form, the one on the right, is much more space efficient.

The formal definition of this new tree form is depicted in Listing 1. Constructors `Term` and `NTerm` represent terminals and nonterminals respectively. We can see that instead of a direct use of an operator as a node, we use abstract nodes. They are `UnOp` for unary operations, `BinOp` for binary operations and `MultiOp` for n-ary operations. `OpName` holds the name of that operation, which semantic is not defined here, since for now we only care about the structure. The semantics needs to be supplied for each intended operation for the lambda calculus to work, but it's presented independently from a tree.

We have implemented two different ways of processing operations. One is by constructing the tree only with the use of `MultiOp` nodes, that hold the entire subtree in a list. And the second is by using a combination of `BinOp` and `UnOp` nodes. This difference is crucial, since it produces different supercombinators, as we show in the section IV-C and further expose by the experiment in the section V-A. The list structure can hold any operation arity, therefore may be used exclusively. The other two operations should be used in tandem to achieve currying in application of supercombinators.

Currying is a well known phenomenon, where we may perceive any n-ary function as a sequence of unary high order functions that are applied to the arguments. We may view the sequence $abcd$ as either being transformed to the node $MultiOp(a, b, c, d)$ or as currying like tree $BinOp(a,\ BinOp(b,\ BinOp(c,\ d)))$.

Both tree approaches can be combined, as we did in our

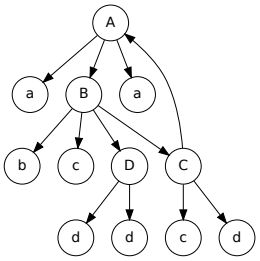


Fig. 3. Graph of elements from the context-free expression (7).

previous work. We have used `MultOp` like nodes for an alternative operation, `BinOp` type for a sequence and `UnOp` for a closure. We have also used special `Nil` node to sign the termination of a whole expression or a sequence. This special node is the reason, why there is L^1 supercombinator in Table I. However, it's equivalent to the L^0 supercombinator and therefore redundant.

The `Jump` node presents an actual jump to the nonterminal designated by the name in this node. This basically turns the set of trees into a directed, possibly cyclic, graph.

B. Argument Lists Extraction from Nonterminals

This step is essential for our process, since we are using jumps to other expressions. And each jump brings a new possibility of permissible terminal symbols at the input of their top lambda expression. Take expressions (7) as an example.

$$\begin{aligned} A &\rightarrow a\ B\ a \\ B &\rightarrow b\ c\ D \mid C \\ C &\rightarrow c\ d \mid A \\ D &\rightarrow d\ d \end{aligned} \tag{7}$$

The only terminal symbol actually inside of the expression A is the symbol a. But others, namely b, c and d are accessible. As they are accessible from the B and the C expressions. The expression D has only the symbol d as accessible.

From a set of terminals and nonrerminals of all expressions, a single directed cyclic graph is constructed. Each nonterminal forms a node which links to its body, and terminals represent leaves. The graph of expression (7) is depicted in Fig. 3.

Afterwards, a depth first search is performed for every nonterminal node in order to find all possible symbols from it. If we omit nonterminals from such a path, only a string of terminals remains. The final step is the removal of all duplicates, so the symbol strings contain each symbol at most once. This process is shown in (8) for nonterminal A. Thus symbol string for expression A is $abcd$, $bcda$ for B and $cdab$ for C. D has only one symbol, d, as a permissible symbol.

$$\begin{aligned} A \Rightarrow A\ a\ B\ b\ c\ D\ d\ d\ C\ c\ d\ a \Rightarrow \\ \Rightarrow a\ b\ c\ d\ d\ c\ d\ a \Rightarrow a\ b\ c\ d \end{aligned} \tag{8}$$

C. Construction of Supercombinator Sets

The process described in this section is performed on all expressions separately, therefore can be executed in parallel.

Listing 2. Definition of a supercombinator data type.

```
data Lambda = LeafLambda LambdaId
 | JumpLambda LambdaId Nterm
 | UnLambda LambdaId OpName SubLambda
 | BinLambda LambdaId OpName SubLambda SubLambda
 | MultiLambda LambdaId OpName [SubLambda]
```

Every single expression yields independent supercombinator set. We used to perform this process with the use of a direct tree transformation in our previous work. We have chosen to do it differently this time.

The basic idea behind obtaining supercombinators is that every operation node of a tree yields exactly one supercombinator. Therefore we have defined a data type for them, shown in Listing 2. All constructors are named after the nodes of a tree they represent. Each of them has its own *id*. All *id*s are unique, since we haven't performed merge yet. The *id* references allow us to use linear form instead of interconnected tree form, hence we may perceive them as pointers.

Each supercombinator has its own permissible argument string. We have separated that string from the underlying data type in order to keep its core aspects tidy and in order. It's functionally irrelevant whether the argument string is represented directly within the data type or as an accompanying list of arguments within a tuple.

The `LeafLambda` constructor represents already mentioned supercombinator $L^0 = \lambda x\ .\ x$, specified only by its *id*. Its argument string consists only from the original symbol of a `Leaf` node. `JumpLambda` is temporary supercombinator that serves as a place-holder for expression name and hold permissible argument strings of that expression. We have already obtained them during the previous step. Other three forms of supercombinators are functionally similar since they represent operations. `SubLambda` represents a supercombinator that is being applied to that operation. It holds only a pointer to the real supercombinator and references to arguments inside the argument string.

The direct mapping between supercombinators and our internal form is shown on the example (9). The ids are upper indices of the L symbol. The arity of a lambda expression is obtained from the length of its argument string. The lists in `SubLambda` tuple refer to the index of an argument inside the argument string. Such an argument string has the length of 2 in this case.

$$\begin{aligned} &BinLambda\ 1\ Sum\ (2,\ [0,1])\ (0,\ [0]) \iff \\ &\iff L^1 = \lambda x_0.\lambda x_1.L^2\ x_0\ x_1 \mid L^0\ x_0 \end{aligned} \tag{9}$$

The permissible argument string for any supercombinator is obtained as a merge of their children arguments. Therefore it is really important to perform step 2, as the jump supercombinators now have an entire argument string of an expression they are referring to.

Using `MultOp` nodes exclusively in a tree yields different supercombinators as an equivalent tree composed of `BinOp` and `UnOp` nodes. The former results into n-ary operation

lambda expressions, where one operation is applied over multiple sub expressions. The latter method usually yields more supercombinators, but each has at most two sub expressions applications in its body, i.e. its underlying operation is at most binary.

D. Merge within Expression Supercombinator Set

The previous step may yield supercombinators that are equal, only their argument strings may be different. All such supercominators are merged, so no duplicates are present inside each set. This step, as the previous one, can be performed in parallel over all expression sets.

The equality of supercombinators, formally described in [9], means, that in order for supercombinators to be equal, they need to have the same arity and need to contain same `SubLambda` elements. In terms of definition in Listing 2 that means equality of all elements excluding the *id*. All equal supercombinators need to be merged. Even some supercombinators which contain references that initially lead to different supercombinators might be equal, since the sub-supercombinators may have been merged in the previous iteration of the merge step. Example of a merger is presented in Table II, where we see the initial and merged supercombinator set of the expression $ab \mid cd$.

We start the merge process with the identification of all equal supercombinators within a set. Then we group them together. We now have some groups of equal supercombinators and the rest of the set. Those groups can now be merged, even in parallel. Well, only their argument strings are merged into a single set and only one supercombinator comes out of this process. Its *id* now needs to be updated in the entire set, replacing the old, now unused *id*s. After this update, new equal supercombinators may be present, so we need to repeat this process until no new equal supercombinators are found. The process depicted in Table II had two merge iterations. Argument sets always consist of strings with the same length, since the arity of equal supercombinators needs to be the same. The resulting set of supercombinators is now duplicate free.

E. Removal of Jump Supercombinators and the Merge of All Sets

At this moment, we possess numerous sets of supercombinators that we need to connect together and then merge them to a single set. Each temporary jump supercombinator is now replaced by a top supercombinator of an expression it points to. It is important to keep the *id*s in between the sets from clashing, i.e. each set needs to have different *id*s. After this step, another merge is applied that joins all the sets to a single, duplicate free set.

For a simple grammar (10) we obtain the final supercombinator set in lambda calculus form, depicted in table III. If we consider the nonterminal A to be the starting symbol, then the top supercombinator is in our case L^2.

$$\begin{aligned} A &\rightarrow a\,B \mid a \\ B &\rightarrow b \mid A\,b \end{aligned} \tag{10}$$

TABLE III
SUPERCOMBINATOR SET OF GRAMMAR (10).

Supercombinators	Arguments
$L^0 = \lambda x_0.\ x_0$	$\{\{a\},\{b\}\}$
$L^1 = \lambda x_0.\ \lambda x_1.\ L^0\ x_0 + L^4\ x_1\ x_0$	$\{\{a\ b\}\}$
$L^2 = \lambda x_0.\ \lambda x_1.\ L^1\ x_0\ x_1 \mid L^0\ x_0$	$\{\{a\ b\}\}$
$L^3 = \lambda x_0.\ \lambda x_1.\ L^2\ x_0\ x_1 + L^0\ x_1$	$\{\{a\ b\}\}$
$L^4 = \lambda x_0.\ \lambda x_1.\ L^0\ x_0 \mid L^3\ x_1\ x_0$	$\{\{b\ a\}\}$

If we use a special form of β-reduction, we obtain the grammar (10) back. By special form we mean replacing all top supercombinators in the body of expression by their nonterminal representation. This is an important step, since it prevents the infinite loop. Already mentioned L^2 supercombinator is the top for A and the L^4 supercombinator is the top for B. But should we use in the β-reduction only on the top supercombinator of an entire grammar, the A, we get a transformed version of a grammar, as shown bellow in (11). We can conclude, that we may use the supercombinator form to transform a form of grammar.

$$A \rightarrow\ L^2\ a\ b \rightarrow^*\ a\ (b \mid A\ b\ ) \mid a \tag{11}$$

V. EXPERIMENTAL RESULTS

We have been using only abstract grammar symbols so far. Those abstract symbols, like a for terminals and A for nonterminals are useful for the explaining purposes, but they do not represent anything we can find in the real world directly. Thus we have performed experiments on the book samples. As mentioned before, we obtain a supercombinator set that represents the structure of elements. But we cannot just use our process over a fixed sequence of words, since that would result into one big supercombinator (and one L^0 supercombinator of course). We need a grammar first. And for that reason we have chosen Sequitur algorithm.

Sequitur algorithm, created by Nevill-Manning and Witten [6], creates context-free grammar from a linear sequence of discrete symbols. The words of English language in our case. The resulting grammar generates only the input text, therefore it's possible to use our supercombinators for text analysis of that text, since no other text is possible to generate either from the Sequitur grammar or our supercombinator set.

We use *the Book of Genesis* from the King James Bible and chapters from Leo Tolstoy novel *War and Peace*[1]. We have chosen these books because they are written in different styles and the former has many unknown authors while the latter was written by a single well known author, therefore may be less schematic and fragmented. We use various samples from these two books, all of which are listed in Table IV. We also list abbreviations, by which we are going to reference them, and we list the total word count of each section as well.

[1]The books were obtained from Project Gutenberg, located at http://www.gutenberg.org, where they are distributed under GNU Free Documentation license 1.2.

TABLE II
EXPRESSIONS $ab \mid cd$ MERGE OF DUPLICATE SUPERCOMBINATORS.

Supercombinators	Arguments	Merged Supercombinators
$L^0 = \lambda x_1.\, x_1$	$\{\, a \,\}$	
$L^1 = \lambda x_1.\, x_1$	$\{\, b \,\}$	
$L^2 = \lambda x_1.\, x_1$	$\{\, c \,\}$	
$L^3 = \lambda x_1.\, x_1$	$\{\, d \,\}$	$L^0 = \lambda x_1.\, x_1$
$L^4 = \lambda x_1\, x_2.\, L^0\, x_1 + L^1\, x_2$	$\{\, ab \,\}$	
$L^5 = \lambda x_1\, x_2.\, L^2\, x_1 + L^3\, x_2$	$\{\, cd \,\}$	$L^1 = \lambda x_1\, x_2.\, L^0\, x_1 + L^0\, x_2$
$L^6 = \lambda x_1\, x_2\, x_3\, x_4.\, L^4\, x_1\, x_2 \mid L^5\, x_3\, x_4$	$\{\, abcd \,\}$	$L^2 = \lambda x_1\, x_2\, x_3\, x_4.\, L^1\, x_1\, x_2 \mid L^1\, x_3\, x_4$

TABLE IV
LIST OF BOOKS AND THEIR SECTIONS USED IN THE EXPERIMENTS.

Book Name	Used Sections	Abbreviation	Words
Book of Genesis	Sections 1 - 3	*Gen3*	1429
Book of Genesis	Sections 1 - 6	*Gen6*	3840
Book of Genesis	Sections 1 - 12	*Gen12*	7306
Book of Genesis	All	*Gen*	39797
War and Peace	Chapter 1	*WP1*	2015
War and Peace	Chapter 2	*WP2*	1378
War and Peace	Chapter 3	*WP3*	1469
War and Peace	Chapter 4	*WP4*	1417
War and Peace	Chapters 1 - 4	*WP1-4*	6279
War and Peace	Chapters 1 - 12	*WP12*	17755
War and Peace	Chapters 1 - 24	*WP24*	37538

TABLE V
SEQUITUR GRAMMAR SAMPLE OBTAINED FROM THE BOOK OF GENESIS.

Rule		Rule body
78	→	`have dominion over the fish` **29** `sea` **79 96**
79	→	**56** `the`
80	→	`all` **61**
81	→	**56** `every`

The resulting Sequitur grammar uses only concatenation operation. Using Sequitur to decompose text of literature, i.e. text without rigorous structure, inevitably leads to the state, where the first rule consists of a long sequence of terminals and nonterminals, while the rest of the rules have rather low arity. Despite this fact, we still might obtain interesting results by performing our process, as the portion of lower arity rules might be represented by a single supercombinator. To imagine how Sequitur grammar looks, see Tab V, where we show a sample of the grammar obtained from *Gen*. We see that nonterminals are represented by numbers and terminals by actual words.

What do we expect from our experiments? Since our process captures the structure of rules, we expect the processed Sequitur grammar to have less elements, represented by supercombinators, than is the count of the rules. We show in Table VI the count of Sequitur rules, each column represents different arity. The last column represent the actual arity of the first rule, it does not represent the amount of rules. The Book of Genesis contains more, and larger, repetitive patterns than War and Peace as we can see in Table VI.

Note, that in Table VI we have 9-ary rule for *Gen3* and *Gen6*, but no such a rule is present in *Gen12*. This is not an error, but a result of the text amount *Gen12* contains. The rule in question here is: `have dominion over the fish` **29** `sea` **79 96**. It generates the phrase `have dominion over the fish of the sea and over the fowl of the air`. In case of *Gen12*, the rule for this phrase is different: `have dominion` **531** `fish` **574 86 111**. More sub-phrase rules have been created, since they are reused elsewhere in *Gen12* text. Two 9-ary rules in *Gen* sample are unrelated, since they generate different phrases.

A. *Difference between Binary and N-ary Trees.*

The first experiment is going to show us, whether we should use binary[2] or n-ary trees. As we mentioned in Section IV-C, n-ary trees might result into a smaller set of supercombinators. On the other hand, a binary tree always results in supercombinators with the underlying operation having arity of maximum two. Without the merge operation in mind, the decomposition of an n-ary operation with the use of an n-ary trees would yield one supercombinator with its underlying operation having arity of n. But with the binary trees, we would get maximum of $n-1$ supercombinators with at most binary operations. The actual arity of those $n-1$ supercombinators would gradually drop from maximum of n to 2 in case that every terminal symbol is different for that particular operation. We expect larger set of supercombinators obtained from binary trees than from n-ary based on those mentioned facts.

The difference between those two approaches is illustrated in Table VII for unary and Table VIII for n-ary tree approach. The fields represent the total number of supercombinators described by their operation arity shown in the second row. As mentioned earlier, binary trees yield always supercombinators with at most binary operations. Since only the concatenation operation was used, no supercombinator with unary operation was created in both cases. Supercombinator with nullary operation is the $L^0 = \lambda x.x$. We see, that in all cases it's present only once in each set, so our merge works.

The total amount of supercombinators is dramatically different for binary and n-ary strategies. That's a direct result of the fact mentioned at the beginning of this section. Although

[2]Those trees aren't actually binary, since unary nodes are possible via $UnOp$ constructor. They are called binary for brevity.

TABLE VI
THE AMOUNT OF SEQUITUR GRAMMAR RULES DIVIDED BY THEIR ARITY.

Sample	2-ary	3-ary	4-ary	5-ary	6-ary	7-ary	8-ary	9-ary	11-ary	1st rule
Gen3	122	18	2	5	-	-	-	1	-	762
Gen6	339	35	7	5	-	-	-	1	-	2116
Gen12	660	61	13	5	3	1	-	-	-	3906
Gen	3363	206	56	22	6	1	2	2	1	19512
WP1	120	6	2	-	-	-	-	-	-	1678
WP2	68	6	1	1	-	-	-	-	-	1173
WP3	73	7	-	-	-	-	-	-	-	1244
WP4	84	7	-	-	-	-	-	-	-	1189
sum	345	26	3	1	-	-	-	-	-	-
WP1-4	421	23	6	1	-	-	-	-	-	4812
WP12	1369	50	6	-	-	-	-	-	-	12582
WP24	2963	106	14	2	-	-	-	-	-	24901

TABLE VII
SUPERCOMBINATOR SETS OBTAINED FROM TWO BINARY TREE APPROACH.

Sample	0-ary	2-ary	Total
WP1	1	1682	1683
WP2	1	1173	1174
WP3	1	1246	1247
WP4	1	1191	1192
Sum	4	5292	5296
Merged	1	5272	5273
WP1-4	1	4825	4826

TABLE VIII
SUPERCOMBINATOR SETS OBTAINED FROM n-ARY TREE APPROACH.

Sample	0-ary	2-ary	3-ary	4-ary	5-ary	1st rule	Total
WP1	1	4	3	1	-	1 (1678)	10
WP2	1	3	2	1	1	1 (1173)	9
WP3	1	4	2	-	-	1 (1244)	8
WP4	1	3	2	-	-	1 (1189)	7
Sum	4	14	9	2	1	4	34
Merged	1	5	5	1	1	4	17
WP1-4	1	8	6	2	1	1 (4812)	19

all supercombinators in the binary section have at most binary operations, the real arity of them can vary. In case of the *WP1* sample, the maximum arity was 728. That is a smaller number than the actual arity of the first rule of Sequitur grammar, which is 1678. This is because the supercombinator form is non-redundant, i.e. the words (the actual arguments of supercombinators from which the total arity is calculated) do not repeat themselves.

If we compare the sum of supercombinators obtained from four different War and Peace chapters and the number of supercombinators obtained after the merge (on supercombinator level) we see no significant difference in the case of Binary trees (5296 compared to 5273). In the n-ary tree case however, the difference is an exact half (34 to 17).

In comparison with the processed grammar of four chapters together, the sample *WP1-4*, we see the difference in case of binary trees (5273 for merged against 4826 for *WP1-4* sample). In the other case however, the difference is insignificant (17 compared to 19). In case of other arities in n-ary tree case, the merged has unified mostly supercombinators with lower arity operations. The difference between merge and *WP1-4* strategy is rather insignificant.

We can thus conclude, that in the case of Sequitur generated grammars, the n-ary tree approach seems to yield significantly smaller set of supercombinators, therefore we are going to use this approach further on in this paper.

B. Constructing Supercombinator Set from a Sequitur Grammar

Straightening out the issue of what tree forms to use we can now proceed to the evaluation of larger book parts. The results are shown in Table IX.

Supercombinators are again divided by their operation arity. This is an especially useful feature for comparison with the arity of Sequitur grammars. We can see that for every arity of the original Sequitur rules (see Table VI) there exists at least one supercombinator with the same operation arity. This result was expected, since we have used n-ary trees. And the actual number of supercombinators is always lower or equal to the number of rules. We can say, that our process does not create any unexpected supercombinators that might introduce different structure into the original text. The arity of the first rule is the same as the arity of the operation within the top supercombinator, which is again the result of the n-ary tree usage.

Another interesting result is the comparison of the difference between binary rules and their respective supercombinators. Since the Book of Genesis seems to have more equal parts, as concluded in Section V, it has more supercombinators with larger arity operation. This also means that the sets of lower arities are larger than in the case of War and Peace. We see that difference in *Gen* and *WP24*, where the former has dropped from 3363 to 126 and the latter from 2963 to 33 for binary operations. This is due to the fact, that supercombinators with lower arity operations may contain sub-supercombinators with

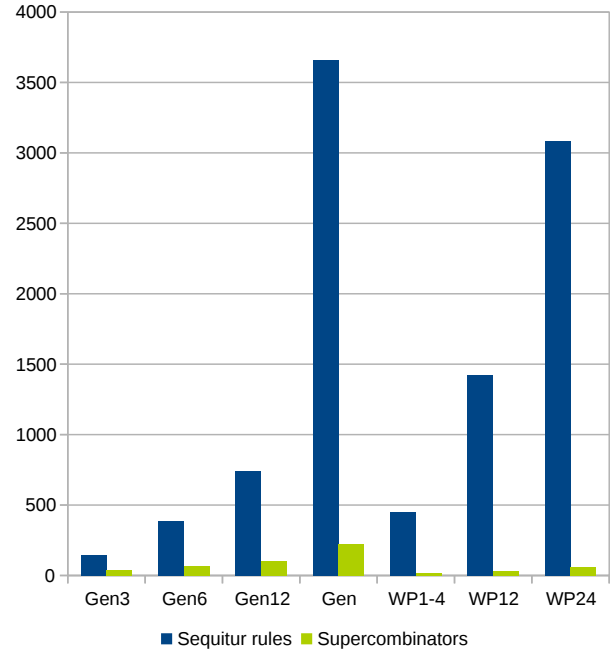


Fig. 4. Total number of Sequitur rules compared with their resulting supercombinator count.

any other operation arities. We see that *Gen* has more larger arity operations than *WP24*.

Figure 4 represents the total difference between the number of Sequitur rules and supercombinators obtained from them. The difference is rather significant. This shows, that our process is capable of data reduction. Different grammar rules with the same structure are merged into a single supercombinator, and as we can see, this has a significant impact in the case of Sequitur generated grammars.

VI. Discussion

Presented results show that our approach can prevent the structural explosion. Although this has already been pointed out in our previous work, we have performed small experiments on a simple regular grammar examples. This paper presents larger scale experiments performed on context-free grammars that support these claims.

We may conclude from the results in Table IX that if the final supercombinator set contains supercombinators with higher operation arities, the total number of supercombinators with the lower ones rises with them as is the case of *Gen* sample. We do not observe this for the *WP24* sample in such a scale, since it has less supercombinators with higher arities.

The different tree approaches that we have defined result into different supercombinator sets. Should we use only binary like trees, the resulting set of supercombinators in the case of Sequitur generated grammars would be larger and thus the promised reduction of size would be impossible. The other, list oriented approach does not suffer from that drawback and delivers the promised results.

Another interesting discovery lies with the analysis of the text itself. Although Sequitur algorithm may be useful for analysis, for example if we want to find out most occurring phrase or longest occurring repeating phrase (like the rule 78 one in Table V), our approach takes this notion one step further. By one step further we mean the analysis of the abstract structures, that supercombinators are. One of those is the argument string length of L^0 supercombinators, that presents absolute count of all words occurring in the text. This number obviously differs from the word count in Table IV, since we do not store duplicates in our structure.

The comparison of arities signifies the difference between The Book of Genesis and War and Piece, but that discovery can be inferred from pure Sequitur results, that we show in Table VI. But should we want to find out, what structure is most used, our supercombinator form is probably better suited for that question. But this is not so visible from the current results, since we have used Sequitur, and thus produced only supercombinators with the concatenation operation. Thus further research with better grammar inference method is necessary.

VII. Related Work

Our work relates with the field of grammar inference. As already mentioned in the Introduction, not only formal inference but the induction of natural languages grammar can be incorporated with our supercombinator set construction mechanism. The induction of grammar can be achieved with the use of various different methods. Onnis, Waterfall and Edelman use cognitive graphs in their model ADIOS to infer CFG in [5]. Adriaans and Van Zaanen created the model EMILE [11], where they use probabilistic methods. Klein and Manning developed model based on constituency [12] that is also capable to induce CFG from text. As our background is in the computer languages field, Stevenson and Cordy presented concise review of the state of the art in [13], where they present various methods of grammar inference.

Our supercombinator form might be practically used in tandem with ontology extraction methods as our supercombinator form of a grammar might be used to identify concepts of a certain kind. In the work of Carvalho, Almeida, Pereira and Henriques [14] we see the use of ontologies that help the concept identification. Other uses of ontologies might be for information retrieval [15], detection of concept similarity across different information media [16] or even for the detection of mental illness from written text [17].

Related methods for concept extractions include rule based approach, as Szwed used in [18]. There we can see extraction of concepts from written text. The rules used are based on Petri nets. Other related method for text analysis is summarization. Example of this method is presented by Jassem and Pawluczuk in [19]. Those methods focus on semantic side of a text, where our supercombinator approach focuses primarily on the structure. The actual meaning is treated separately, therefore we can concentrate more clearly on those separate aspects.

TABLE IX
RESULT COUNT OF SUPERCOMBINATORS DIVIDED BY THEIR OPERATION ARITY.

	0	2	3	4	5	6	7	8	9	11	Last	Total
Gen3	1	25	7	2	4	-	-	-	1	-	1 (762)	41
Gen6	1	40	14	5	5	-	-	-	1	-	1 (2116)	67
Gen12	1	53	25	12	5	3	1	-	-	-	1 (3906)	101
Gen	1	107	71	43	20	6	1	2	2	1	1 (19512)	255
WP1-4	1	8	6	2	1	-	-	-	-	-	1 (4812)	19
WP12	1	15	15	3	-	-	-	-	-	-	1 (12582)	35
WP24	1	29	20	9	2	-	-	-	-	-	1 (24901)	62

VIII. CONCLUSION

We have presented a way to represent any CFG nor-redundantly in a single set of supercombinators. The process has been described in detail, where we show it in separate steps. Many of those steps might be performed in parallel, so better computation time is achievable.

Applications of our supercombinator structure have been presented on the samples taken from literature. The Book of Genesis and War and Piece by Tolstoy were used. Since our process works only on grammars, we needed to process those samples first with Sequitur algorithm. This algorithm constructs CFG from a finite string of symbols, in our case words. We show , that our representation significantly reduces the size of entire structure, since it is non-redundant. Further on, we have discussed the possibilities of a structure analysis, that is possible to perform on our supercombinator structure.

In our future work, we would like to extract more information with the use of better inference mechanism that Sequitur algorithm brings. In the section VII we show various possible ways to induce a grammar from text samples so we would like to actually use them in our experiments to further prove the abilities of our supercombinator form construction method.

REFERENCES

[1] S. Edelman, "On the nature of minds, or: truth and consequences," *Journal of Experimental & Theoretical Artificial Intelligence*, vol. 20, no. 3, pp. 181–196, 2008. doi: 10.1080/09528130802319086. [Online]. Available: http://dx.doi.org/10.1080/09528130802319086

[2] N. Chomsky, *Syntactic Structures*. Mouton and Co., 1957.

[3] E. M. Gold, "Language identification in the limit," *Information and control*, vol. 10, no. 5, pp. 447–474, 1967. doi: 10.1016/S0019-9958(67)91165-5. [Online]. Available: http://dx.doi.org/10.1016/S0019-9958(67)91165-5

[4] C. De La Higuera, "A bibliographical study of grammatical inference," *Pattern recognition*, vol. 38, no. 9, pp. 1332–1348, 2005. doi: 10.1016/j.patcog.2005.01.003. [Online]. Available: http://dx.doi.org/10.1016/j.patcog.2005.01.003

[5] L. Onnis, H. R. Waterfall, and S. Edelman, "Learn locally, act globally: Learning language from variation set cues," *Cognition*, vol. 109, no. 3, pp. 423–430, 2008. doi: 10.1016/j.cognition.2008.10.004. [Online]. Available: http://dx.doi.org/10.1016/j.cognition.2008.10.004

[6] C. G. Nevill-Manning and I. H. Witten, "Identifying hierarchical structure in sequences: A linear-time algorithm," *J. Artif. Intell. Res.(JAIR)*, vol. 7, pp. 67–82, 1997. doi: doi:10.1613/jair.374. [Online]. Available: http://dx.doi.org/doi:10.1613/jair.374

[7] P. Klint, R. Lämmel, and C. Verhoef, "Toward an engineering discipline for grammarware," *ACM Trans. Softw. Eng. Methodol.*, vol. 14, no. 3, pp. 331–380, Jul. 2005. doi: 10.1145/1072997.1073000. [Online]. Available: http://doi.acm.org/10.1145/1072997.1073000

[8] J. Kollár, M. Sičák, and M. Spišiak, "Towards machine mind evolution," in *Computer Science and Information Systems (FedCSIS), 2015 Federated Conference on*. IEEE, 2015. doi: 10.15439/2015F210 pp. 985–990. [Online]. Available: http://dx.doi.org/10.15439/2015F210

[9] J. Kollár, M. Spišiak, and M. Sičák, "Abstract language of the machine mind," *Acta Electrotechnica et Informatica*, vol. 15, no. 3, pp. 24–31, 2015. doi: 10.15546/aeei-2015-0025. [Online]. Available: http://dx.doi.org/10.15546/aeei-2015-0025

[10] M. Sičák, "Higher order regular expressions," in *Engineering of Modern Electric Systems (EMES), 2015 13th International Conference on*. IEEE, 2015. doi: 10.1109/EMES.2015.7158427 pp. 1–4. [Online]. Available: http://dx.doi.org/10.1109/EMES.2015.7158427

[11] P. W. Adriaans and M. Van Zaanen, "Computational grammar induction for linguists," *Grammars*, vol. 7, pp. 57–68, 2004.

[12] D. Klein and C. D. Manning, "Corpus-based induction of syntactic structure: Models of dependency and constituency," in *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*. Association for Computational Linguistics, 2004. doi: 10.3115/1218955.1219016 pp. 478–485. [Online]. Available: http://dx.doi.org/10.3115/1218955.1219016

[13] A. Stevenson and J. R. Cordy, "Grammatical inference in software engineering: an overview of the state of the art," in *Software Language Engineering*. Springer, 2013, pp. 204–223. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-36089-3_12

[14] N. Carvalho, J. J. Almeida, M. J. Pereira, and P. Henriques, "Probabilistic synset based concept location," in *SLATe'12—Symposium on Languages, Applications and Technologies*. Alberto Simões and Ricardo Queirós and Daniela da Cruz, 2012. doi: 10.4230/OASIcs.SLATE.2012.239 pp. 239–253. [Online]. Available: http://dx.doi.org/10.4230/OASIcs.SLATE.2012.239

[15] J.-A. Asensio, N. Padilla, and L. Iribarne, "Information retrieval using an ontological web-trading model," in *Proceedings of the 2013 Federated Conference on Computer Science and Information Systems*. IEEE, 2013, pp. pages 243–249.

[16] M. Wang, X. Liu, L. Huang, B. Lang, and H. Yu, "Ontology-based concept similarity integrating image semantic and visual information," in *Proceedings of the 2014 Federated Conference on Computer Science and Information Systems*, ser. Annals of Computer Science and Information Systems, vol. 2. IEEE, 2014. doi: 10.15439/2014F273 pp. pages 289–296. [Online]. Available: http://dx.doi.org/10.15439/2014F273

[17] P. Šaloun, "From lightweight ontology to mental illness indication," in *Scientific Conference on Informatics, 2015 IEEE 13th International*. IEEE, 2015. doi: 10.1109/Informatics.2015.7377799 pp. 9–12. [Online]. Available: http://dx.doi.org/10.1109/Informatics.2015.7377799

[18] P. Szwed, "Concepts extraction from unstructured polish texts: a rule based approach," in *Proceedings of the 2015 Federated Conference on Computer Science and Information Systems*, ser. Annals of Computer Science and Information Systems, vol. 5. IEEE, 2015. doi: 10.15439/2015F280 pp. 355–364. [Online]. Available: http://dx.doi.org/10.15439/2015F280

[19] K. Jassem and L. Pawluczuk, "Automatic summarization of polish news articles by sentence selection," in *Proceedings of the 2015 Federated Conference on Computer Science and Information Systems*, ser. Annals of Computer Science and Information Systems, vol. 5. IEEE, 2015. doi: 10.15439/2015F186 pp. 337–341. [Online]. Available: http://dx.doi.org/10.15439/2015F186

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 513–516
DOI: 10.15439/2016F419
ACSIS, Vol. 8. ISSN 2300-5963

Word2vec Based System for Recognizing Partial Textual Entailment

Martin Víta
NLP Centre
Faculty of Informatics
Masaryk University
Botanická 68a, 602 00 Brno
Czech Republic
Email: info@martinvita.eu

Vincent Kríž
Faculty of Mathematics and Physics
Charles University
Malostranské nám. 25, 118 00 Prague
Czech Republic
Email: kriz@ufal.mff.cuni.cz

***Abstract*—Recognizing textual entailment is typically considered as a binary decision task – whether a text T entails a hypothesis H. Thus, in case of a negative answer, it is not possible to express that H is "almost entailed" by T. Partial textual entailment provides one possible approach to this issue.**

This paper presents an attempt to use word2vec model for recognizing partial (faceted) textual entailment. The proposed approach does not rely on language dependent NLP tools and other linguistic resources, therefore it can be easily implemented in different language environments where word2vec models are available.

I. Introduction and Preliminaries

NOWADAYS, textual entailment belongs to intensively and deeply studied notions in NLP, with potentially many practical applications including paraphrase detection, multi-document summarization, machine translation evaluation, plagiarism detection, etc. In this section we provide a brief description of textual entailment, partial textual entailment, we mention word2vec model and present the main aim of this work.

A. Textual Entailment

Different definitions of textual entailment (abbr. as TE) and a systematic overview of this area can be found in an older but comprehensive paper [1]. *Recognizing textual entailment* (RTE for short) is a corresponding decision problem whether a given (coherent) text T entails a given text H (in this context often called a hypothesis). Currently, there exist several systems for RTE problem: an up-to-date list of them can be found at ACLwikiWeb[1]. Some of them were created in order to participate SemEval challenges.

Since RTE is a binary decision problem, in case of a negative result of RTE, i. e., when T does not entail H, it is not possible to state "how distant" is H from another hypothesis H', such that H' is entailed by T. From a different point of view, it is not possible to express that H is "almost entailed" by T in this setting. Partial textual entailment is one possible approach to this issue. The key elements of the idea of partial textual entailment were introduced in [2], although the notion of partial textual entailment was not explicitly mentioned in the paper. The motivation for partial textual entailment has naturally arised from the problem of (automatic) analysis of student responses in educational process.

[1] http://aclweb.org/aclwiki

B. Partial and Faceted Textual Entailment

According to [3], we say that an ordered pair $(T; H)$ forms a partial textual entailment (abbr. as PTE) if *a fragment* of the hypothesis H is entailed by T. In this definition, the notion of a fragment of the hypothesis is no more specified. Hence, the key question is how to decompose the hypothesis into fragments.

In [2], facets were introduced as special fragments: a facet is an ordered pair of words (f_1, f_2) that are contained in the hypothesis – accompanied by a semantic relation binding these words together. A simplified version of this approach – used in SemEval 2013 challenge – deals only with a pair of words *without* explicitly metioned semantic relation.

For example, if the hypothesis has the form of a sentence "The water was evaporated, leaving the salt.", one of corresponding facets is: (evaporated, water). Starting now, we are going to use only this simplified model.

The problem of recognizing faceted entailment can be stated as follows: *"Does the given text T express the same semantic relationship between the words f_1 and f_2 exhibited in H?"*

II. Note on a Related Work – Existing System for Faceted Entailment

Currently, there are only a very few systems for recognizing faceted entailment. In SemEval 2013 Task 7, only one system was submitted – a system of Levy et al. [3].

It consists of three components: *Exact Match, Lexical Inference, Syntactic Inference*. Exact match checks whether lemmas of words contained in the facet appear both in the text. The Lexical Inference is based on Resnik similarity [4] over WordNet [5]. The idea behind this module is to find out whether words semantically related (semantically similar) to those contained in the facet, occur also in then text.

The Syntactic Inference module is based on BIUTEE entailment engine that deals with dependence trees. The dependency

978-83-60810-90-3/$25.00

tree corresponding to a given facet is obtained from the dependency tree of the whole hypothesis using lowest common ancestor (LCA) of facet-nodes: it is just the path from one facet node to the second one via LCA node. This inference component has no paralel in our approach.

The best results of Levy et al. system were achieved in "Majority" configuration (Exact $\vee$ (Lexical $\wedge$ Syntactic)). In terms of F_1-measure, the scores vary from 0.765 to 0.816 depending on different scenarios.[2]

A. Main Aim of the Work

In this paper we present a novel system for recognizing partial/faceted textual entailment that is based on word2vec representations of the words contained in the text T and words contained in the facets.

The results of this monolingual setting can provide rough estimations of overall accuracy and other measures for intended cross-lingual modification that is briefly described in the last section of this text, thus this work can also be viewed as a prerequisite to cross-lingual faceted entailment.

B. Word2vec Model

Word2vec model belong to a class of distributed representations of words. The main attribute of distributed representations (proposed relatively long time ago, in the second half of 80th in [6]) is, that the representations of (semantically) similar words are close in the vector space.

Word2vec model arises from the idea of predicting the neighbours of a word using a neural network. There are two possible modes of predicting: distributed Skip-gram or Continuous Bag-of-Words (CBOW), see [7]. The CBOW idea is to predict the word "in the middle" from the surrounding words, whereas in Skip-gram model the training objective is to learn predicting its context in the same sentence. The (real number) vector representations of words correspond with the weights between input and first hidden layer in used deep feedforward network. The dimension of the target word2vec space is a parameter of the model.

III. Task Definition, Algorithm Description and Used Data

Recognizing faceted entailment is a binary classification task. The inputs are the text T and the hypothesis H along with the facet (f_1, f_2) of words contained in H. The output classes are *Expressed* and *Unaddressed*[3] (which means the semantic relationship between f_1 and f_2 is expressed explicitly or implicitly in T, or not, respectively).

Let us assume we have a word2vec model, i. e. for (almost) each word w we have its vector representation $r(w)$ in word2vec space of a given dimension, a text T an a facet (f_1, f_2). Parameters of our algorithm is a threshold α from the $(0, 1)$ interval.

[2]The comparison of our proposed system with this one was not provided due to missing information about the data used in each scenario.

[3]In the context of faceted entailment, "Expressed" and "Unaddressed" labels are used instead of "Entailed" and "Not entailed".

A. Algorithm Description

The decision algorithm for faceted textual entailment (abbreviated as W2V in the following text) works in the following steps:

1) Split the text T into tokens $t_1, \ldots, t_n$.
2) Get the word2vec representations
$$r(t_1), \ldots, r(t_n), r(f_1), r(f_2)$$
whenever possible.
3) For f_1 select the word t_p such that $d(r(f_1), r(t_p))$ is equal to
$$\min\{d(r(f_1), r(t_k)) \mid 1 \leq k \leq n\},$$
where d is the standard cosine distance. For f_2 select analogously t_q. Roughly said, select two words in T that have the lowest distances to the facets in the sense of word2vec space.
4) If
$$\frac{d(r(f_1), r(t_p)) + d(r(f_2), r(t_q))}{2} \leq \alpha$$
than (f_1, f_2) is *Expressed* in T, otherwise (f_1, f_2) is *Unaddressed* by T. If some word of the facet is missing in the word2vec model, the result class is set to *Unaddressed*.

If the facet consists of more than two words (tokens), use this algorithm analogously for all of them.

The optimal value of α is obtained after experiments on training data – the selected value provides the best results of this algorithm in terms of overall accuracy. We will refer to this algorihtm as "W2V".

In addition, we will employ the trivial algorithm (that will be refered as "EXACTMATCH"): it returns *Expressed* in case that both words of the facet are contained in the text T, otherwise it returns *Unaddressed*. No lemmatization is taken into account since we are preparing a maximally language independent solution – in EXACTMATCH we deal only with word forms. This trival algorithm is used in order to treat with situations when a facet uses the same words as those contained in the text – but that are not contained in the word2vec model (for instance, correct words with a very low frequency).

B. Used Data and Word2vec Model

The evaluation was performed using a dataset derived from SciEntsBank corpus [8] that was used in the Joint Student Response Analysis and 8th Recognizing Textual Entailment Challenge at SemEval-2013 Task 7. This corpus is focused on previously mentioned domain of student response analysis. It contains scholar questions, reference answers and student responses. From the "practice point of view", the aim is to recognize whether the student's answer is at least partially correct. Transforming this issue to recognizing (partial) textual entailment environment models this situation: the role of the hypothesis H plays the reference answer and the role of the

text T is played by the student's answer. If H is (partially) entailed by T, than student's answer is (partially) correct.

Let us illustrate it on the example.

QUESTION: *You used several methods to separate and identify the substances in mock rocks. How did you separate the salt from the water?*

STUDENT ANSWER: *Let the water evaporate and the salt is left behind.*

REFERENCE ANSWER: *The water was evaporated, leaving the salt.*

FACET: *(evaporated, water)*

As already mentioned, T is the student answer and the task is to decide whether the semantic relationship between "evaporated" and "water" is expressed in T. In this case, the relationship is "Expressed", thus the student answer can be regarded as partially correct.

In contrast, when student answers "I don't know." the facet *(evaporated, water)* is obviously not expressed.

The advantage of using this corpus is that facet extraction was already done and the faceted entailments were manually annotated. The SemEval-2013 Task 7 corpus is divided in two parts, training and test collections. The training collection contains 13145 pairs, the test collection contains 16263 pairs text-hypothesis (i. e. facets). As the texts T, we have considered just the student answers in all cases, no other texts (like parts of questions) were taken into the account.

While in case of "standard" recognizing textual entailment there are several training/test sets, for faceted/partial textual entailment, annotated corpora are very rare.

Word2vec model was built using the original implementation[4] over the TC Wikipedia[5]. Standard preprocessing issues were performed (e. g. lowercasing, punctuation removal). The model was obtained with the following basic parameters: the dimension was set to 200, the window was set to 5, the mode was CBOW.

IV. Results

Since recognizing faceted textual entailment is a binary classification task, the performance is measured in a standard way – obtaining precision, recall and F_1-measure scores over the test collection of SciEntsBank corpus. F_1-measure was chosen in order to compare our results with [3]. The threshold α in W2V algorithm was set to 0.555 – this value of the parameter maximizes the overall accuracy over the training collection.

The results are summarized in Table I, Table II and Table III. They were obtained by "official SemEval scripts"[6].

EXACTMATCH achieves relatively high precision at positive class, nevertheless it provides low recall – these characteristics correspond with "common sense" expectations. The combination of these two approaches leads to better results in F_1-measure than the W2V approach used separately.

[4] `http://code.google.com/p/word2vec`

[5] `http://nlp.cs.nyu.edu/wikipedia-data/`

[6] `https://www.cs.york.ac.uk/semeval-2013/task7/data/uploads/datasets/semevaltask7code.zip`

TABLE I
W2V ∨ EXACTMATCH RESULTS

	Precision	**Recall**	F_1**-measure**
Expressed	0.661	0.811	0.729
Unaddressed	0.875	0.761	0.814

TABLE II
W2V RESULTS

	Precision	**Recall**	F_1**-measure**
Expressed	0.652	0.774	0.707
Unaddressed	0.854	0.761	0.805

V. Conclusion

We have presented a simple system for recognizing faceted textual entailment that is based on word embeddings: word2vec model in particular – other embeddings with similar characteristics (like GloVe) can be treatened in an analogous way.

Despite of its simplicity it provides reasonable results in terms of F_1-measure. The key features of this system are no need of other language resources in except of a relevant word2vec model and no usage of NLP tools. Thus it can be quickly implemented in any language where word2vec models can be created. The preparation of word2vec models requires only a collection of texts of a sufficient volume like Wikipedia in the corresponding language and/or a relevant web corpus (without any annotations).

Using word2vec model and our approach "simulates" the use of lemmatization in morphologically rich languages (since the cosine distance of a given word form and its lemma is usually very low – observed during experiments with Czech language), thus our approach would most likely achieve relatively comparable results also in other languages.

It can be straightforwardly implemented in different programming languages and environments – in our case, in `R` environment (enriched by `lsa` and `tm` packages) was used. Word2vec representations were stored in CSV format and were loaded into `R`.

Comparing to the mentioned approach of Levy et al. [3], our approach provides a comparable results in terms of overall accuracy – but it can be easily implemented also in "under-resourced" languages (where syntactic tools – as well as WordNet – are not available). Our proposed system approximately corresponds with the first two components of their system (Exact Match and Semantic Inference). The differences are summarized as follows: in [3], Exact Match contains lemmatization, in our approach lemmatization is not used.

TABLE III
EXACTMATCH RESULTS

	Precision	**Recall**	F_1**-measure**
Expressed	0.970	0.366	0.531
Unaddressed	0.731	0.993	0.842

Semantic Inference module is in our setting "replaced" by low distances in word2vec space.

VI. Further Work

Our proposed system can serve as a baseline for further experiments.

Since word2vec models are able to capture many linguistic regularities [9], it is intended to employ rule based transformations on facet representations and subsequently determining whether transformed representations are contained in representation of T (for example, dealing with representations of hyper/hyponyms of words that forms the given facet, similarly as in [10]). These extension can be viewed as a paralel of Syntactic inference module of previously mentioned system.

In our presented approach, the threshold α stays constant in all cases and the distances $d(f_1, t_p)$ and $d(f_2, t_q)$ were simply combined into the mean, which was followingly tested against α. Other way of improving our system will be based on employing more features, e.g.:

- raw distances $d(f_1, t_p)$ and $d(f_2, t_q)$,
- the number of words in text between t_p and t_q,
- the angle between vectors $r(f_1) - r(f_2)$ and $r(t_p) - r(t_q)$,
- features obtained from hypernymy/hyponymy, synonymy and other attributes derived from WordNet data.
- ...

and using ML methods like SVM etc. We suppose that employing features (especially those arising from semantic resources like WordNet will help to improve both precision and recall).

Another part of further work is application-oriented: we are going to employ recognizig faceted entailment system in text summarization task (like that described in [11]) etc.

Note on the Cross-lingual Approach

As already mentioned, the proposed approach will be extended for using in a cross-lingual environment. It has been demonstrated in [12], paralel word2vec models can be used for of generating and extending dictionaries and phrase tables. The underlying idea is simple (with little assumptions about the languages involved): unknown word translations can be obtained by learning language structures over large monolingual data and mapping between languages on a small domain (in terms of the mapping).

More formally, let us have n word pairs and their vector representation $(x_i, z_i)_{i=1}^n$, where $x \in R^{d_1}$ is a vector representation of i-th word in the source language and $z \in R^{d_2}$ a vector representation of its translation. The goal is to find a matrix W such that Wx_i approximates z_i. The matrix W is obtained as a solution of an optimization problem:

$$\min_W \sum_{i=1}^{n} \|Wx_i - z_i\|^2.$$

In [12], this problem is solved with stochastic gradient descent.

At this moment, the modification of our algorithm for cross-lingual faceted entailment is straightforward: Having a facet (f_1, f_2) in the source language and the text T in the target language, then we take the vector representations (x_1, x_2) in source word2vec space, compute $(z_1, z_2) = (Wx_1, Wx_2)$ and determine representations of words that are the closest to z_1 and z_2 in the sense of cosine similarity in the target language word2vec model. The rest will be identical to the monolingual case.

Acknowledgment

The second author was supported by SVV - specific academic research - reg. no.: 260 333, which is funded by the Charles University in Prague.

References

[1] I. Androutsopoulos and P. Malakasiotis, "A survey of paraphrasing and textual entailment methods," *Journal of Artificial Intelligence Research*, pp. 135–187, 2010.

[2] R. D. Nielsen, W. Ward, and J. H. Martin, "Recognizing entailment in intelligent tutoring systems," *Natural Language Engineering*, vol. 15, no. 04, pp. 479–501, 2009.

[3] O. Levy, T. Zesch, I. Dagan, and I. Gurevych, "Recognizing partial textual entailment." in *ACL (2)*, 2013, pp. 451–455.

[4] P. Resnik, "Using information content to evaluate semantic similarity in a taxonomy," in *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence, IJCAI 95, Montréal Québec, Canada, August 20-25 1995, 2 Volumes*, 1995, pp. 448–453.

[5] C. Fellbaum, *WordNet*. Wiley Online Library, 1998.

[6] D. R. G. H. R. Williams and G. Hinton, "Learning representations by back-propagating errors," *Nature*, pp. 523–533, 1986.

[7] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *arXiv preprint arXiv:1301.3781*, 2013.

[8] M. O. Dzikovska, R. D. Nielsen, and C. Brew, "Towards effective tutorial feedback for explanation questions: A dataset and baselines," in *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Association for Computational Linguistics, 2012, pp. 200–210.

[9] T. Mikolov, W.-t. Yih, and G. Zweig, "Linguistic regularities in continuous space word representations." in *HLT-NAACL*, 2013, pp. 746–751.

[10] J. A. Miñarro-Giménez, O. Marín-Alonso, and M. Samwald, "Applying deep learning techniques on medical corpora from the world wide web: a prototypical system and evaluation," *arXiv preprint arXiv:1502.03682*, 2015.

[11] K. Jassem and L. Pawluczuk, "Automatic summarization of polish news articles by sentence selection," in *2015 Federated Conference on Computer Science and Information Systems, FedCSIS 2015, Lódz, Poland, September 13-16, 2015*, 2015, pp. 337–341. [Online]. Available: http://dx.doi.org/10.15439/2015F186

[12] T. Mikolov, Q. V. Le, and I. Sutskever, "Exploiting similarities among languages for machine translation," *arXiv preprint arXiv:1309.4168*, 2013.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 517–525
DOI: 10.15439/2016F304
ACSIS, Vol. 8. ISSN 2300-5963

Exploration for Polish-* bi-lingual translation equivalents from comparable and quasi-comparable corpora

Krzysztof Wołk
Polish-Japanese Academy of Information Technology,
ul. Koszykowa 86, 02-008 Warszawa, Poland
Email: kwolk@pja.edu.pl

Krzysztof Marasek
Polish-Japanese Academy of Information Technology,
ul. Koszykowa 86, 02-008 Warszawa, Poland
Email: kmarasek@pja.edu.pl

Agnieszka Wołk
Polish-Japanese Academy of Information Technology,
ul. Koszykowa 86, 02-008 Warszawa, Poland
Email: awolk@pja.edu.pl

***Abstract*—In contemporary world, translation becomes a critical need of the time. Parallel dictionaries have now become a most accessible source by humans, but confines are there as they do not offer good quality translation function, because of neologisms and words that are out of vocabulary. To overcome this problem in the usage of statistical translation systems is becoming more and more important in maintaining the eminence and quantity of the training data. But due to the limitations in these systems they have very limited availability for few languages and very limited narrow text areas. The purpose of this research is to bring calculation time up gradation via GPU acceleration, tuning script introduction and the enhancement and improvements in the methodologies of the contemporary comparable corpora mining through re-implementation of analogous algorithms through Needleman-Wunch algorithm. Experiments have been conducted on multiple language data which were extracted on numerous domains from Wikipedia. For the sake of Wikipedia, multiple cross-lingual contrasts and comparison were established. Optimistic impact on the both quantity and quality of mined data was observed due to such changes and adaptation. The solution is language independent and highly practical especially for under-resourced languages.**

I. Introduction

THE purpose of the research is to organize the language models and parallel and comparable corpora. This process advances the quality of SMT through riddling of parallel corpora and it also works through extraction of supplementary parallel data via the resulting corpora. In order to improve the language spring of the SMT systems, alteration measures and interpolation methods are applied to the obtainable prepared data. For this, various experiments were led by using wide domain (TED presentations on variety of topics).

SMT system's assessment was functioned on random samples of analogous data by utilizing automated algorithms (BLEU metric) in order to assess the possible usage and standard of the SMT systems' output. In addition, human evaluation was conducted in order to measure the impact of newly obtained corpora on translation error reduction. [1]

While experiments are discussed, the utilization of the software Moses Statistical Machine Translation Toolkit [2] is done. Further, the symmetrisation is done using Berkeley Aligner [3] and translation models training is done using multi-threaded application the GIZA++ tool. Only from single language data base, the statistical models are shaped well by utilizing SRILM (SRI Language Modelling toolkit). Furthermore, the data from external domain is adapted. In the situation of parallel modelling, in-domain data collection is found using, Moore-Lewis Filtering [4] while single-language models are linearly interpolated [5].

Finally, methods recommended in the Yalign [6] parallel data mining tool are upgraded and critically evaluated. By using the tool in a multi-threaded way and by employing graphics processing units (GPUs), its speed was also amplified. Furthermore, by utilizing Needleman-Wunch [7] algorithm and by developing a tuning script that is used to regulate mining parameters to fix domain supplies, its quality is improved as well.

In the tests, the resultant SMT systems out-performed the baseline systems in terms of BLEU metric and error reduction.

II. Corpora Types

A corpus includes a large collection of texts stored up on a computer. These text compilations are known as corpora. Usually in linguistic fields, parallel corpus as a term is used with the reference to texts which are the source of translation of each other. In order to deal with the statistical machine translation, we are significantly considerate about parallel corpora. These are paired with text through another language. For the preparation of parallel texts for the call for statistical machine translation, it may require removing the text from HTML, web crawling and sentence structure [5] and performing document alignment.

Two major kinds of parallel corpora exist in two different languages. One is comparable corpus in which common texts are present and their content is also the same. Polish and English newspaper's articles are best example of comparable corpora. The second one is the translation corpus, in this type of corpus the text of first language (e) is the translation of text in second language (f). It is significant to recognize that the word "comparable corpora" signifies the texts in two different languages, they are common in the content but they are not common translations of one another. [5]

In order to assess a parallel text, pattern arrangement is used which mentions common texting sections (approximately, sentences), that is a significant requirement for examination.

Within first and second language machine translation algorithms for translation are often trained, using common fragments. This includes a first and second language corpus that is an element for element translation of the first language corpus. In such kind of training it may involve huge training sets which can be removed from huge corpora of common sources, like databases of the news articles written in the first and second languages while describing common events [5]. Due to this complicatedness, it is problematic to obtain high quality parallel data, significantly for uncommon languages. Comparable corpora are the key to the solution of problem of absence of data for rare language pairs that are under-resourced languages and other subject domains. It is easier and conceivable to use comparable corpora to achieve straight knowledge for the purposes of translations. This data is considered to be a precious foundation of knowledge for other information dependent and cross-lingual tasks. This data is not as rare as parallel, even for Polish-* languages. On the flip side, single language data is accessible in huge qualities [5].

While concluding, there are four key corpora kinds which are notable. Parallel corpora that are also very uncommon, can be explained as corpora which have the translations of the common file into two and more than two languages. For that such kind of data it is needed for it to be aligned, at the level of sentence as a minimum. Noisy-parallel corpus that contains bilingual structure sentences that are not excellently arranged and they can also have translations with bad quality. Yet, in most cases, bilingual translations of a precise document are present in it. A comparable corpus is structured from unstructured sentences and with un-translated bilingual documents, but the text or the documents need to be about the same topic. Seemingly quasi-comparable corpus also has very non-parallel bilingual and very mixed documents, they may or may not be structured according to the topic [8].

III. State of the art

If comparable corpora are concerned, numerous efforts (as for Wikipedia) have been observed in order to evaluate parallel data samples. Two core methods for building comparable corpora can be easily separated or illustrated. The most common method is founded on the notion of recovery of cross-lingual knowledge. Second approach is based on the fact that source texts or documents should be interrelated by using random translation systems. After that the translated documents can be compared with the texts that are written in the most targeted language and the basic purpose is to find out the pairs of common pairs within the documents.

An exciting idea for searching for parallel data inside the Wikipedia was mentioned and explained in [9]. Firstly, the idea is to utilize an online machine translation (MT) system to decode the language, using translating techniques to translate Dutch language into English language on Wiki pages, and after that try to compare original EN pages with translated ones. This idea, though seems computationally impractical, is interesting but perplexed problem. Their second method uses a dictionary generated from the hyperlinks and Wikipedia titles that are shared between documents. Inappropriately, the second method involves the generation of Wikipedia titles by using dictionary and the hyperlinks that are being shared between texts. The second method was improved in [9] by a range of spare confines of the communication between the portions of the concern document and with the help of the introduction of added measure on the bases of similarity. They report that in [9] the accuracy (number of correct translation pairs over total strength of the applicants) is approximately 21% and at this stage in the recommended method [10], the accuracy is around 43%.

Yasuda and Sumita [12] projected a MT bootstrapping structure on the basis of figures that helps to create a sentence-structured corpus. Sentence structure and alignment is accomplished by utilizing a bilingual lexicon which is spontaneously upgraded by the structured sentences automatically. They use corpus that has previously been structured for the early training session. Their recommendations showed that 10% of Japanese Wikipedia sentences have an equivalent on the English Wikipedia.

Tyers and Pienaar in [10] gave the idea to give lead to internal Wiki links. A bilingual dictionary is removed and evaluated on the bases of Wikipedia link structure. In the work of these authors, they actually calculated the normal disparity for numerous languages that are linked between Wikipedia pages. Results showed that 69-92% depends on the specific language according to the precision of the method.

The authors in [13] attempt to raise the best skill in parallel data mining with the help of modelling of document-level arrangement by utilizing the observational technique, so that parallel sentences can greatly and most commonly be found in propinquity. Authors also use explanation that is available on Wikipedia and a mechanically injected lexicon model. For that authors report 80% precision and 90% recall.

In [14] author introduces an instinctive arrangement methodology for parallel textual documentation fragments which utilizes a phrase-based SMT system and textual entailment method. The author mentions that important up gradation in SMT quality were adopted (BLEU increased by 1.73) by utilizing this arranged data between French and German languages.

M. Volk and M. Plamada also explained another method for discovering Wikipedia which was explained in [14]. Previously explained methods differ from their solutions. In these methods parallel data was restricted to the monotonically control of the arrange algorithms that were used in order to match the candidate sentences. Their algorithm pays no heed to the position of a candidate in the text and, instead, ranks candidates by means of modified measurements that combine contradictory similarity criteria [15]. Additionally, the authors limit the process of mining towards a specific domain and examine the semantic equality of took out pairs. Mining accuracy is 39% in the work of M. Plamada and M. Volk, while 26% are for loud parallel sentences, with other remained sentences misaligned. Reportedly they say that an up gradation of 0.5 points happened in the BLEU metric out of domain data, but no prominent improvement took place in-domain data.

In [16] the authors suggested to use titles and few meta-information only, like time for specific document and publication data, neglecting its full-fledged contents, to lessen the cost of development of the comparable corpora. The similarity of cosine of the title terminology as frequency vectors was utilized to match the contents and the contents of the matched pairs. In the research explained in [17], the two authors came up with a document of resemblance of measure which is based on the occasions. In order to count the values of this metric, they present documents as sets of occasions and events. These occasions are time based and are based on the geographical terms that are found in the texts. Documents that are targeted ranked and based on geographical orders. The writers in [18] also recommend a programmed method in order to build a similar corpus from the website by utilizing news web pages like Twitter and Wikipedia. They mine things, URLs of web pages, filtering within time limitations and the specified lengths of the documents as features for the congregation and the categorization of the similar data.

In the current research, a methodology that was inspired by the Yalign is being used. The method originally was far from perfect, but after the up gradation and improvements during research, it actually has provided us with excellent mining results.

IV. Parallel data mining

This study aims at developing new methodologies for obtaining parallel corpora from the sources that are not aligned likewise comparable corpora, quasi-comparable or noisy parallel. We have selected Wikipedia as a base of the data because of the huge number of texts it provides (4,524,017 on EN wiki approximately). Moreover, Wikipedia comprises not only similar documents, but it also includes some text documents which are the translations of one another. This approach can be qualified by the use of measurements in the translations systems of MT.

In data mining, TED corpora, ready for the IWSLT 2015 assessment by FBK, were selected. This domain is very wide and protects numerous subject areas. Data contains nearly 2.5M un-accessible words [19]. The tests were shown on PL-*.

Our idea can be separated and divided into three key steps. Firstly, comparable data is selected, then it is arranged at the article level, and lastly for parallel, the arranged results are mined. The last two steps are important; the reason is that there are huge numbers of differences between documents of Wikipedia. Sentences in the Wiki corpus are mainly not arranged, with translation lines whose assignment does not agree to any textual information in the foreign language. Furthermore, some sentences have no consistent translations within corpus at all. The alignment is very difficult for the correctness. For that sentence alignment should also be practicable with competency that is of practical use in variety of applications. Earlier, a mining tool precedes the data and the text should be ready. Initially, entire data is preserve in a relational database. In the second phase, our tool organizes article pairs and eradicates the articles that seem to be present in only one of the two languages. All these arranged articles at topic-level are checked in order to get rid of XML tags, HTML tags or other noisy data (figures, references, tables, etc.). Lastly, fluent documents are marked with a single ID as an aligned topic, similar corpus. In order to separate the parallel pair's sentence, a decision was taken in an attempt to develop strategy that was designed to program the parallel text mining process after finding the sentences that are near to the translation matches from comparable corpora. This offers chances for finding parallel corpora from bases, as not translated textual documents including the web, which are not limited to any specific language pair.

Though, alignment models for two languages that are of two designated languages particularly the first one is to be created as priority. By using a comparative sentence metric gave an unbalance estimate (a number that is somehow between 0 and 1). This approximation shows that how there is a possibility of being a translation of the two sentences. It also applies pattern alignment, which gave a sequence that increases the quantity of the unique thread (per sentence pair) that is same between the two texts [6]. In order to maintain order alignment, we at first used error friendly and very slow Yalign that used an A* search method [20] to find an ideal alignment between multiple sentences in two particular documents. The algorithm has a polynomial worst time complication. It cannot control alignments that cancel each other or such that form from two

sentences a single sentence [20]. After the sequence alignment, only sentences which have top probability of being translated are placed in the results. The output is checked to deliver top quality corpus. To accomplish this, an action is used: if the sentence has the similar score that is less than the threshold, at that stage pair will not be included. For similarity of sentence metric, the algorithm uses statistical techniques. The classifier must be skilled to find if the sentence pairs are translations between each other. In this research Support Vector Machine (SVM) classifier is used. SVM can give a distant outlook to the parting hyper plane during labelling. This distance can be simply adapted by utilizing a Sigmoid Function to return a value that is similar to similarity between 0 and 1 [21]. Usage of classifier means that the excellence of the rearrangement depends not just on the input but it also depends on the quality of the classifier. For the training of the classifier, high quality parallel data is necessary. For this, we utilized the TED talks [8] corpora. To get a dictionary, we used a phrase table and took 1-grams from it [22].

V. Development of the Mining Process

Much to our distress, the local Yalign instrument was not practical enough for matters related to calculation in terms of comprehensive and real life parallel data mining. Typical execution required input in simple text or web links and the RAM memory was re-loaded with classifier for every text pair. Moreover, the Yalign software is uniquely stranded. In an effort to speed up the process, this unique solution was developed to supply articles to the tool and load classifier only once per session. The newly developed system also used the multithreading and minimized the mining time by factor of 6.1x, utilizing the four cores and eight thread i7 CPU. The alignment algorithm was replaced to improve accuracy and to make best use of the strengths of the GPUs to fulfil the supplementary requirements of the computations.

A. Needleman-Wunsch algorithm (NW)

Major purpose of this algorithm is to line up two sequences together. At first, it is necessary to define the correspondence among the two elements. It can be explained by using the similarity matrix S in which N represents the number of elements in the first sequence and M represents the number of elements in the second sequence. The algorithm was designed to analyse the matters related to bio-informatics for the assessment of RNA and DNA. However, it can be modified to deal with assessment of textual data. In other words, the given algorithm combines real number and a pair of each element together in the matrix. As the similarity index rises, so is the similarity of the elements concern. For instance, if we have a similarity matrix S which is equal to the numbers between 0 and 1 than by 0 for the two expressions mean that their similarity index is zero whereas 1 means that the two given

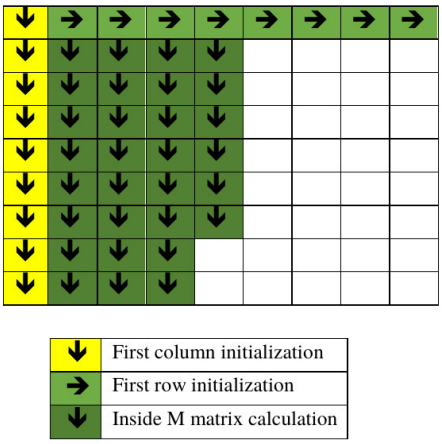


Fig. 1 Needleman-Wunsch S-matrix calculation

words are perfect translations of each other. The significance of similarity matrix index for the consequences of the algorithm is undeniable [20]. After that we will identify the consequences of gap penalty. It is essential particularly when one of the elements of the sequence is connected with the gap in another sequence.

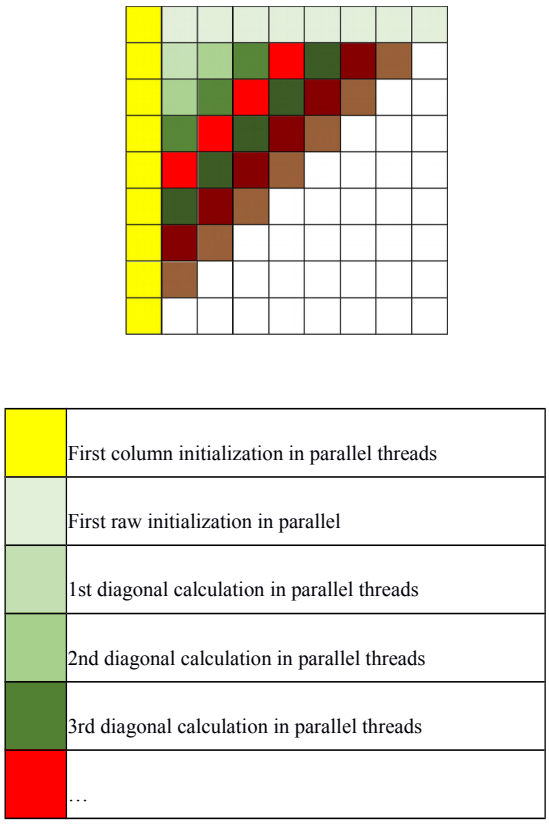


Fig. 2 Needleman-Wunsch S-matrix calculation with parallel threads

Though, such a stride will lead to a penalty (p). The calculation of the S matrix will be executed right from the start of S (0, 0) element which by definition is equal to 0. Once the process of initializing first and second row is started, the algorithm moves to the other elements of the S matrix, taking the cue from the upper left side, leading up to the bottom right side. All of these steps are illustrated in the Fig. 1.

The two (CPU and GPU) NW algorithms are theoretically same but the GPU has a leverage of better performance due to its hardware handicap up to max(n,m) times.

Though, this process is different when it comes to the calculation of the elements of S matrix. At this step forward, we will apply the multiple strands to an optimized level. These processes are so small so that that they can be processed easily by a huge number of Graphics Processing Units (ex. CUDA cores). The major purpose behind this is that we will calculate all the elements in predefined diagonals in parallel way which always starts from upper left and ends at the bottom right, as presented in the Fig. 2 [23].

In an attempt to explain the value of a cell of S(m,n), for all pairs of m and n, the values to its top S(m-1,n) , left S(m,n-1) and top left S(m-1,n-1) must be known in advance and filled with tokenized documents that are to be compared. Where, S(m,n) can be calculated with the help of following equation: [24]

$$S(m,n) = max\{S(m-1) \pm 1, \\ S(m-1,n)-2, S(m,n-1)-2\}$$

Regardless of the results of the A* algorithm, if the coherence between calculation and the gap penalty are defined on the similar patterns as they are defined in the NW algorithm, they will have similar results just if there are supplementary restraints on the way, these ways cannot be led to uphill or left-side in the matrix. Yalign does not compel such terms and conditions, therefore in some cases, expressions can be repeated more than once or misaligned. Most of the cases the algorithm keeps on moving backward and forward to the first two sequences in line. S matrix has been exemplified without any barriers in the Fig. 3.

	a	d	e	g	f
a	X				
d		X			
c	X				
d		X			
e			X	X	X

Fig. 3 S matrix pass-through without constraints

The alignment result in this scenario is:

a, d, a, d, e, g, f

a, d, c, d, e, −, −

In the same problem, the NW would react as presented in Fig. 4:

	a	d	e	g	f
a	X				
d		X			
c		X			
d		X			
e			X	X	X

Fig. 4 S matrix pass through with NW

The alignment result using NW would be:

a, d, −, −, e, g, f

a, d, c, d, e, −, −

For second example, we will assume that the very first sequence in a row is “Tablets make children very addicted” and second one says that “Tablets make people spoil children”. Fig. 5 given below presents a solution to this problem by the help of A* algorithm and Fig. 6 shows it with NW, without any restrains.

	Tablets	make	children	very	Addicted
tablets	X	-	-	-	-
make	-	X	-	-	-
people	X	-	-	-	-
spoil	-	-	-	-	-
children	-	-	X	X	X

Fig. 5 A* alignment without constraints

Due to lack of any restrains, repetitions and bad alignments are most likely to be made by envisaging the blemish A* algorithm, which is applicable in the Yalign program. However, some of the sentences can be easily passed over during the checking of the alignment. That is why NW with GPU optimization is the most preferred algorithm.

	Tablets	make	children	very	addicted
tablets	X	-	-	-	-
make	-	X	-	-	-
people	-	-	-	-	-
spoil	-	-	-	-	-
children	-	-	X	-	-

Fig. 6 NW alignment with constraints

B. Other improvements to data mining methodology

The SVM classifier is used to define the quality of the alignment; it creates a trade-off between the recall and the precision. Two configurable variables can be found in the classifier.

- Threshold: the acceptance of an alignment is 'Good' if the confidence threshold if high. For less recall and more precision, the value should be lowered. The probability estimated through the support vector machine is the 'confidence' that classifies it as 'is a translation' or 'is not a translation' [25].
- Penalty: the alignment allows control of the amount of 'slipping ahead' [6] if you are aligning subtitles. There would then be no extra or fewer paragraphs, and the alignment would be one-to-one while the penalty would be high. The penalty would be lower if the translations of the alignments are similar and there are no extra paragraphs.

These parameters are automatically selected during the training; however, they can be manually changed if necessary. A tuning algorithm is also introduced in our implementation of the solution[1] used. In this research, it allows adjustments for better accuracy. Random articles of the corpus must be extracted to perform the tuning; humans can manually align these random articles. Given the information provided, the tuning mechanism finds the classifier value through randomly selected parameters; it tries to find the output that would be as humanly similar as possible.

The improvements that were debated earlier, deal entirely with heuristics utilized in the mining tool and they can be implemented to any fluent textual data. Though, Wikipedia has huge extra sources of cross-lingual information that need to be utilized. Firstly, the theme domain of Wikipedia cannot be enclosed into a particular domain; this page covers approximately any topic in question. Due to the fact that these articles frequently contain complex vocabulary, and that is why approaches of statistical mining can skip many of the parallel sentences. The answer to this query can be extracted dictionary by utilizing the article titles from Wikipedia (Fig. 7) and moreover it can be implemented into web crawler tool.

Fig. 7 Sample of bi-lingual Wikipedia page title

In [11] authors say that the precision of this dictionary can be attained at 92%. For that this dictionary can be utilized not only for the delay of the parallel corpora but in the classifier phase of exercise as well.

Secondly, figures contain best quality parallel sentences and explanations. It is likely to attain pictures and graphics with the help of hyperlinks and analysis the pictures. Similarly, same is for any figures, maps, tables, videos, audio or even compound media on Wikipedia. Tactlessly not whole knowledge can be removed from Wikipedia dumps and it is essential to utilize a web crawler that is right for this assignment (Fig. 8). It also can be predicted that simply only cross-language knowledge which are marked with simple links can be removed.

Fig. 8 Specimen of bi-lingual character caption

On Wikipedia it is very likely for the sentences to be equal linguistically, if they are referenced with the similar publication. Such an analytical approach, combined with other comparative tactics, can move us forward to better accuracy in a parallel text sequence discovery task (Fig. 9).

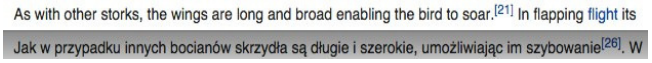


Fig. 9 Sample of bilingually referenced sentence

[1]https://github.com/krzwolk/yalign

VI. Assessment of obtained parallel corpora and conclusions

By the help of techniques explained earlier we were capable of creating comparable corpora for many PL-* language pairs and later, probe them for parallel phrases. We paired Polish (PL) with Arabic (AR), Czech (CS), German (DE), Greek (EL), English (EN), Spanish (ES), Persian (FA), French (FR), Hebrew (HE), Hungarian (HU), Italian (IT), Dutch (NL), Portuguese (PT), Romanian (RO), Russian (RU), Slovene (SL), Turkish (TR), Vietnamese (VI). Statistics of the resulting corpora are presented in Table I.

In order to assess the corpora quality and usefulness, we trained the baseline SMT systems by utilizing the WIT[2] data (BASE). We also augmented them with resulting mined corpora both as parallel data as well as the language models (EXT). The additional corpora were domain adapted through the linear interpolation and Modified Moore-Lewis filtering [26]. Tuning of the system was not executed during experiments due to the volatility of the MERT [27]. However, usage of the MERT would have an overall positive impact on MT system in general [27]. The results are showed in Table II.

The assessment was based on sets of official test sets from IWSLT 2013[3] conference. Bilingual Evaluation Understudy (BLEU) measurement was used to score the progress. As it was expected earlier, sets of supplementary data enhance the general quality of translation for each and every language.

In order to verify the importance of our results we conducted the significance tests for 4 divers languages. The decision was made to use the Wilcoxon test. The Wilcoxon test (also known as the signed-rank test or the matched-pairs test) is one of the most popular alternatives for the Student's t-test for dependent samples. It belongs to the group of non-parametric tests. It is used to compare two (and only two) dependent groups that is two measurement variables. The significance tests were conducted to evaluate how the improvements differ from each other. Changes with low significance were marked with *, significant changes were marked with ** and very significant with *** in presented Tables III and IV.

Bi-lingual sentence extraction has particular importance in dealing with unsubstantiated learning processes for multiple tasks involved. With the help of this methodology we can easily resolve the dissimilarities between Polish and other languages. It is a method that is independent from language matters, it is adaptable to new environments for any language pair. Our experiments validated the performance of the method. The corpora received as consequence of the experiments, can maximize the quality of MT in an under-resourced text domain. However, in few scenarios, only small differences have been observed in BLEU scores. Keeping that aside, it can be said that even such small differences, can influence the real life situations positively especially for infrequent translation cases. Moreover, the results of our work are freely accessible for research community (corpora is hosted at OPUS[4] and tools at GitHub[5]). In order to see things from sensible outlook, we can say that such methodology does not require large scale training or special language specific

Table I.

Results of mining after progress

Language Pair	Number of bi-sentences	Number of unique PL tokens	Number of unique foreign tokens
PL-AR	823,715	1,345,702	1,541,766
PL-CS	62,507	197,499	206,265
PL-DE	169,739	345,266	341,284
PL-EL	12,222	51,992	51,384
PL-EN	172,663	487,999	412,759
PL-ES	151,173	411,800	377,557
PL-FA	6,092	31,118	29,218
PL-FR	51,725	215,116	206,621
PL-HE	10,006	42,221	47,645
PL-HU	41,116	130,516	136,869
PL-IT	210,435	553,817	536,459
PL-NL	167,081	446,748	425,487
PL-PT	208,756	513,162	491,855
PL-RO	6,742	38,174	37,804
PL-RU	170,227	365,062	440,520
PL-SL	17,228	71,572	71,469
PL-TR	15,993	93,695	92,439
PL-VI	90,428	240,630	204,464

[2]https://wit3.fbk.eu/mt.php?release=2013-01

[3]Iwslt.org

[4] http://opus.lingfil.uu.se/Wikipedia.php

[5]https://github.com/krzwolk/yalign

TABLE II.

RESULTS OF MT EXPERIMENTS

LANGUAGE	SYSTEM	DIRECTION	BLEU	LANGUAGE	SYSTEM	DIRECTION	BLEU	LANGUAGE	SYSTEM	DIRECTION	BLEU
PL-AR	BASE	→PL	19.67	PL-FA	BASE	→PL	14.21	PL-PT	BASE	→PL	27.07
	EXT	→PL	21.78		EXT	→PL	14.32		EXT	→PL	29.14
	BASE	←PL	20.98		BASE	←PL	16.87		BASE	←PL	30.11
	EXT	←PL	23.12		EXT	←PL	17.03		EXT	←PL	31.33
PL-CS	BASE	→PL	12.21	PL-FR	BASE	→PL	19.07	PL-RO	BASE	→PL	22.16
	EXT	→PL	12.98		EXT	→PL	20.01		EXT	→PL	22.26
	BASE	←PL	13.44		BASE	←PL	21.13		BASE	←PL	25.01
	EXT	←PL	14.21		EXT	←PL	21.56		EXT	←PL	25.67
PL-DE	BASE	→PL	23.68	PL-HE	BASE	→PL	17.03	PL-RU	BASE	→PL	12.36
	EXT	→PL	24.91		EXT	→PL	17.65		EXT	→PL	13.51
	BASE	←PL	26.61		BASE	←PL	18.18		BASE	←PL	13.58
	EXT	←PL	26.87		EXT	←PL	18.54		EXT	←PL	14.32
PL-EL	BASE	→PL	14.27	PL-HU	BASE	→PL	14.62	PL-SL	BASE	→PL	12.11
	EXT	→PL	14.67		EXT	→PL	15.23		EXT	→PL	12.57
	BASE	←PL	17.22		BASE	←PL	17.18		BASE	←PL	14.26
	EXT	←PL	17.28		EXT	←PL	17.81		EXT	←PL	14.61
PL-EN	BASE	→PL	15.91	PL-IT	BASE	→PL	18.83	PL-TR	BASE	→PL	11.59
	EXT	→PL	17.01		EXT	→PL	19.87		EXT	→PL	12.68
	BASE	←PL	17.09		BASE	←PL	21.19		BASE	←PL	13.07
	EXT	←PL	18.43		EXT	←PL	21.34		EXT	←PL	13.44
PL-ES	BASE	→PL	16.35	PL-NL	BASE	→PL	18.29	PL-VI	BASE	→PL	12.66
	EXT	→PL	17.92		EXT	→PL	20.13		EXT	→PL	14.12
	BASE	←PL	18.34		BASE	←PL	20.79		BASE	←PL	14.11
	EXT	←PL	18.65		EXT	←PL	21.45		EXT	←PL	15.17

rammer resources, and despite of that they produce gratifying results.

Because statistically classified data contains some amounts of noisy data, in future we plan to develop precise filtering strategies for bi-lingual corpora. The results of current solution are highly related to SVM classifier. In other words, we plan to train more classifiers for different text domains in order to discover more bi-lingual sentences.

VII. ACKNOWLEDGEMENTS

Work financed as part of the investment in the CLARIN-PL research infrastructure funded by the Polish Ministry of Science and Higher Education and was backed by the PJATK legal resources.

VIII. REFERENCES

[1] K. Wołk, K. Marasek. „Real-Time Statistical Speech Translation." *In: New Perspectives in Information Systems and Technologies*, Volume 1. Springer International Publishing, 2014, p. 107-113. http://dx.doi.org/10.1007/978-3-319-05951-8_11

[2] K. Wołk, K. Marasek. „Polish–English Speech Statistical Machine Translation Systems for the IWSLT 2013". *In: Proceedings of the 10th International Workshop on Spoken Language Translation*, Heidelberg, Germany. 2013, p. 113-119. http://dx.doi.org/10.13140/RG.2.1.1128.9204

[3] A. Haghighi et al. "Better word alignments with supervised ITG models." *In: Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2*. Association for Computational Linguistics, 2009, p. 923931.

[4] P. Koehn. „Statistical machine translation." *Cambridge University Press*, 2009. http://dl.acm.org/citation.cfm?doid=1380584.1380586

[5] G. Berrotarán, R. Carrascosa, A. Vine „Yalign documentation", https://yalign.readthedocs.org - accessed 01/2015

[6] R. Dieny, J. Thevenon, J. Martinez-Delrincon, J. C. Nebel. „Bioinformatics inspired algorithm for stereo correspondence." *International Conference on Computer Vision Theory and Applications*, March 5–7, Vilamoura - Algarve, Portugal, 2011.

[7] G. Musso. „Sequence alignment (Needleman-Wunsch, Smith-Waterman)", http://www.cs.utoronto.ca/~brudno/bcb410/lec2notes.pdf.

[8] M. Cettolo, C. Girardi, M. Federico. "Wit3: Web inventory of transcribed and translated talks." *In: Proceedings of the 16th Conference of the European Association for Machine Translation (EAMT)*. 2012, p. 261-268.

[9] M. Mohammadi; N. Ghasemaghaee. „Building bilingual parallel corpora based on Wikipedia." *In: Computer Engineering and Applications (ICCEA)*, 2010 Second International Conference on. IEEE, 2010, p. 264-268. http://dx.doi.org/10.1109/ICCEA.2010.203

[10] F. M. Tyers, J. A. Pienaar. „Extracting bilingual word pairs from Wikipedia", *Collaboration: interoperability between people in the creation of language resources for less-resourced languages 19*, 2008, p. 19-22.

[11] J. R. Smith, C. Quirk, K. Toutanova. „Extracting parallel sentences from comparable corpora using document level alignment." *In:*

TABLE III.

SIGNIFICANCE TESTS PL ->*

Language Pair	Number of bi-sentences
PL-EN	0.0103**
PL-CS	0.0217**
PL-AR	0.0011***
PL-VI	0.0023***

TABLE IV.

SIGNIFICANCE TESTS * ->PL

Language Pair	Number of bi-sentences
PL-EN	0.0193**
PL-CS	0.0153**
PL-AR	0.0016***
PL-VI	0.0027***

Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics. Association for Computational Linguistics, 2010, p. 403-411.

[12] K. Yasuda, E. Sumita. „Method for building sentence-aligned corpus from wikipedia". *In: 2008 AAAI Workshop on Wikipedia and Artificial Intelligence (WikiAI08),* 2008, p.263-268.

[13] S. Pal, P. Pakray, S. K. Naskar. "Automatic Building and Using Parallel Resources for SMT from Comparable Corpora." *In: Proceedings of the 3rd Workshop on Hybrid Approaches to Translation (HyTra)@ EACL*, 2014. p. 48-57.

[14] M. Plamada, M. Volk. "Mining for Domain-specific Parallel Text from Wikipedia." *Proceedings of the Sixth Workshop on Building and Using Comparable Corpora*, ACL 2013, 2013, p.112-120. http://dx.doi.org/10.5167/uzh-80043

[15] A. Aker, E. Kanoulas, R.J. Gaizauskas. "A light way to collect comparable corpora from the Web". *In: LREC*, 2012, p. 15-20.

[16] J. Strötgen, M. Gertz, C. Junghans. "An event-centric model for multilingual document similarity." *In: Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval. ACM*, 2011, p. 953-962. http://dx.doi.org/10.1145/2009916.2010043

[17] M.L. Paramita et al. "Methods for collection and evaluation of comparable documents." *In: Building and Using Comparable Corpora*. Springer Berlin Heidelberg, 2013, p. 93-112. http://dx.doi.org/10.1007/978-3-642-20128-8_5

[18] D. Wu, P. Fung. "Inversion transduction grammar constraints for mining parallel sentences from quasicomparable corpora." *In: Natural Language Processing– IJCNLP 2005*. Springer Berlin Heidelberg, 2005, p. 257-268. http://dx.doi.org/10.1007/11562214_23

[19] J.H. Clark et al. "Better hypothesis testing for statistical machine translation: Controlling for optimizer instability." *In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2*. Association for Computational Linguistics, 2011, p. 176181.

[20] S. Adafre; M. De Rijke. „Finding similar sentences across multiple languages in Wikipedia." *In: Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics*, 2006, p. 6269.

[21] K. Wołk, K. Marasek. "A Sentence Meaning Based Alignment Method for Parallel Text Corpora Preparation." *In: New Perspectives in Information Systems and Technologies*, Volume 1. Springer International Publishing, 2014, p. 229-237. http://dx.doi.org/10.1016/j.protcy.2014.11.024

[22] A. Axelrod, X. HE, J. Gao. "Domain adaptation via pseudo in-domain data selection." *In: Proceedings of the Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics*, 2011, p. 355-362.

[23] K. Wołk, K. Marasek. "Tuned and GPU-accelerated parallel data mining from comparable corpora." *In: Text, Speech, and Dialogue. Springer International Publishing*, 2015, p. 32-40. http://dx.doi.org/10.1007/978-3-319-24033-6_4

[24] C. S. Khaladkar. "An Efficient Implementation of Needleman Wunsch Algorithm on Graphical Processing Units", PHD Thesis, School of Computer Science and Software Engineering, The University of Western Australia, 2009.

[25] https://github.com/machinalis/yalign/issues/3 accessed 10.11.2015

[26] R. Roessler. "A GPU implementation of NeedlemanWunsch, specifically for use in the program pyronoise 2." *Computer Science & Engineering*, 2010.

[27] T. Joachims. "Text categorization with support vector machines: Learning with many relevant features." *Lecture Notes in Computer Science vol 1398*, 2005, p. 137-142. http://dx.doi.org/10.1007/BFb0026683

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 527–532
DOI: 10.15439/2016F311
ACSIS, Vol. 8. ISSN 2300-5963

Towards increasing F-measure of approximate string matching in O(1) complexity

Adrian Boguszewski, Julian Szymański, Karol Draszawka
Department of Electronic, Telecommunication and Informatics
Gdańsk University of Technology
Gdańsk, Poland
E-mail: {adrbogus1, julian.szymanski, kadr}@eti.pg.gda.pl

***Abstract*—The paper analyzes existing approaches for approximate string matching based on linear search with Levenshtein distance, AllScan and CPMerge algorithms using cosine, Jaccard and Dice distance measures. The methods are presented and compared to our approach that improves indexing time using Locally Sensitive Hashing. Advantages and drawbacks of the methods are identified based on theoretical considerations as well as empirical evaluations on real-life dictionaries.**

***Index Terms*—approximate string matching, misspelling correction, LSH, CPMerge, AllScan, Levenshtein distance, neural net indexer**

I. Introduction

AN APPROXIMATE string matching is a task of finding a specific word in a given dictionary, that is similar to the word provided by the user to at least a certain degree.

This task is encountered in a wide spectrum of applications, especially for calculating similarity between texts [1]. One of the examples is misspelling detection in a written text and recommendation for a correct word. This solution is used in most interfaces where a user enters a text, e.g.: in web search engines, while she or he enters a query, or in mobile phones, while typing messages. Among other examples, there is plagiarism detection [2], [3] or spam filtering [4], that checks whether a text contains words intentionally modified in order to evade naive spam filters, e.g. vulgar words with added dots instead of some letters in internet posts.

The typical approaches for approximate string matching use some kind of edit distance, such as Levenshtein distance [5]. The two words are matched, if an edit distance between them is below a specified threshold. A useful algorithm should then return a number (possibly all) of words from a predefined dictionary that approximately match a given (input) word in a reasonable time.

In this paper we compare four different approaches to accomplish this, reporting their processing time as well as quality of matched words returned: linear search with Levenshtein distance, AllScan and CPMerge algorithms proposed in [6], and an approach based on Locality Sensitive Hashing. We identify their advantages and drawbacks.

The paper is constructed as follows. Section II describes Levenshtein, AllScan and CPMerge algorithms. Section III presents the approach based on Locally Sensitive Hashing. The experiments and the results of their evaluation in the approximate string matching task are provided in Section IV. Finally, Section V contains conclusions and an idea of further improvements.

II. Approximate string matching algorithms

At the very beginning we define basic symbols we use to describe the algorithms:

1) Σ - an alphabet, a finite set of symbols (letters)
2) Σ^* - a set of all possible words over Σ (e.g. a, b, ..., aa, ab, ...)
3) $V \subset \Sigma^*$ - a finite size dictionary
4) $sim(x, y)$ - a similarity function between words x and $y - f: \Sigma^* \times \Sigma^* \rightarrow [0, 1]$
5) $\alpha \in [0, 1]$ - similarity threshold
6) $|x|$ - length of word x
7) $|X|$ - the number of elements in set X

In general, a search for similar words, can be defined as constructing a set $Y_{x,\alpha}$ of certain words y from dictionary V, for which the similarity to a given word x is greater or equal to α:

$$Y_{x,\alpha} = \{y \in V | sim(x, y) \geq \alpha\}. \quad (1)$$

A. Linear search with Levenshtein distance

The most popular approaches for approximate string matching use edit distance. In general, an edit distance measures two strings dissimilarity by counting how many edit operations are required to change one word into the other. Levenshtein distance considers three single character modifications: insertion, deletion and substitution [7]. Each of them has an equal cost of 1.

Formally, Levenshtein distance $lev(x, y)$ between two words x and y of lengths $|x|$ and $|y|$, is equal to $lev_{x,y}(|x|, |y|)$, where $lev_{x,y}(i, j)$ is a discrete function of two non-negative arguments defined as:

$$lev_{x,y}(i, j) = \min \begin{cases} lev_{x,y}(i-1, j) + 1 \\ lev_{x,y}(i, j-1) + 1 \\ lev_{x,y}(i-1, j-1) + 1_{(x_i \neq y_j)} \end{cases}, \quad (2)$$

if $\min(i, j) \neq 0$,

$$lev_{x,y}(i, j) = \max(i, j), \quad \text{if} \min(i, j) = 0.$$

For example, the distance between words *written* and *writes* equals 2, because we need two modifications to transform first

string into the second one:
written $\rightarrow$ *writen* (deletion of *t*)
writen $\rightarrow$ *writes* (substitution of *s* for *n*)

In order to make use of (1) explicitly, a Levenshtein-derived similarity measure can be defined:

$$sim_{lev}(x,y) = \frac{\max(|x|,|y|) - lev(x,y)}{\max(|x|,|y|)}. \quad (3)$$

Because we need to browse the whole dictionary, the complexity of this algorithm is $O(n)$ where n denotes number of objects in the dictionary in term of dictionary size. This is the main main disadvantage of this straightforward approach. On the other hand, this method provides all matching words in the result, so that no one similar word is ever missed. Another useful property it has is that there is no need for preprocessing of the dictionary, that may require additional computations and storage space, which will be the case for methods described next. The approach finds many modifications e.g.: Damerau-Levenshtein distance [8] or others [9], [10], but the complexity usually remains linear. Thus, in experiments we use the basic Levenshtein implementation as a baseline.

B. Shingle word representation

The rest of compared algorithms for approximate string matching do not use edit distance to compare strings in their raw form. Instead, they all incorporate a preprocessing step, called *shingling* [11], which converts a string into a set of *n*-grams. The similarity of two strings are then determined indirectly by computing the similarity between the corresponding sets of *n*-grams.

In our experiments, we split words into letter trigrams and call them *features* that represent a particular word. For example, a word `rotation` is represented by a set `{$$r, $ro, rot, ota, tat, ati, tio, ion, on$, n$$}`, where `$` sign denotes 'no letter'.

It should be noticed that such a tri-gram representation is chosen arbitrarily and can be further improved [12]. However, to compare the approaches of approximate string matching we decided to stick to one fixed representation.

There is a number of similarity measures between feature sets than can be used for shingle-based strings representation. One of the most popular is cosine similarity, defined by (4).

$$sim_{cos}(x,y) = cos(X,Y) = \frac{|X \cap Y|}{\sqrt{|X||Y|}}, \quad (4)$$

where X and Y are *feature sets* (containing $|X|$ and $|Y|$ elements) of x and y respectively .

Other measures, like Jaccard or Dice distance, can also be successfully applied here [13]. They are defined by (5) and (6) respectively.

$$sim_{jacc}(x,y) = jaccard(X,Y) = \frac{|X \cap Y|}{|X \cup Y|} \quad (5)$$

$$sim_{dice}(x,y) = dice(X,Y) = \frac{2|X \cap Y|}{|X| + |Y|} \quad (6)$$

To provide an illustrative example let us assume two words: $x =$ `rotation` and $y =$ `aviation`. We then have $X =$ `{$$r, $ro, rot, ota, tat, ati, tio, ion, on$, n$$}` and $Y =$ `{$$a, $av, avi, via, iat, ati, tio, ion, on$, n$$}`. Hence $|X| = 10, |Y| = 10$. The cosine similarity of these words is equal to $cos(X,Y) = \frac{5}{\sqrt{10*10}} = 0.5$. Jaccard and Dice distances are as follows: $jaccard(X,Y) = \frac{5}{15} = 0.33$, $dice(X,Y) = \frac{2*5}{10+10} = 0.5$

C. AllScan algorithm

The AllScan algorithm first shingles word x obtaining a feature set X. All words in the vocabulary must also be shingled accordingly in a preprocessing phase. Then, AllScan computes lower and upper bounds for *length* of word y potentially similar to x to at least α. For cosine similarity, this bounds are determined by inequality:

$$\left\lceil \alpha^2 |X| \right\rceil \leq |Y| \leq \left\lfloor \frac{|X|}{\alpha^2} \right\rfloor. \quad (7)$$

Inequality (7) comes from (1) and (4), after observing that for minimal length of y we have $|X \cap Y| = |Y|$, while for maximal length of y, it is $|X \cap Y| = |X|$.

To give an example, if we assume $\alpha = 0.7$ and $x =$ `rotation`, $|X| = 10$. Therefore Y set size must be between 5 and 20, because $|Y| \geq \lceil 0.7^2 * 10 \rceil \geq 5$ and $|Y| \leq \lceil \frac{10}{0.7^2} \rceil \leq 20$

Additionally, there is a minimum *overlap* τ value defined for each possible word length from (7). The τ means minimal value of the same letter trigrams, which have to occur in both strings to exceed the similarity threshold value α. For cosine similarity, it is therefore:

$$\tau = \left\lceil \alpha \sqrt{|X||Y|} \right\rceil. \quad (8)$$

Let us assume $\alpha = 0.7$ and words `rotation` and `aviation` ($|X| = |Y| = 10$). Equation (8) gives 7. Hence the feature sets of the words have to contain at least seven shingles in common to be similar in 0.7. They have only four such shingles, so they do not satisfy the minimal overlap value.

Based on obtained $|Y|$ and τ values, the algorithm retrieves all words from the dictionary that satisfy a given matching criterion, i.e. it returns all the words, similarities of which exceed the threshold.

The main advantage of this approach is the retrieval of all matching words (that satisfy given similarity measure), but the problem is its performance. Although the search space is reduced due to y length bounds calculation, the overall complexity of AllScan algorithm is still $O(n)$ in term of dictionary size, both preprocessing the dictionary and searching.

D. CPMerge

CPMerge algorithm [6] extends AllScan by reduction of the dictionary that is searched during single retrieval. The improvement limits the size of the dictionary, removing words which certainly are not a result. It uses Property 1 to

determine which words are candidates to be in a result set. The result set is much smaller than the whole dictionary and it causes significant speedup.

Property 1 *Let there be a set X (of size $|X|$) of word x n-grams and a set (of any size) Y of word y n-grams. Consider any subset $Z \subseteq X$ of size $(|X| - \tau + 1)$. If $|X \cap Y| \geq \tau$, then $Z \cap Y \neq \varnothing$.*

Assume $x =$ `rotation` ($|X| = 10$) and word y with length 6 ($|Y| = 8$). If $\alpha = 0.5$ then $\tau = 5$ (from (8)). Hence, $|Z|$ must be 6 and if $|X \cap Y| \geq 5$ then $Z \cap Y$ must have at least one element.

As previous algorithms, CPMerge returns all matching words as a result. It is faster than AllScan algorithm, due to described improvements, but its complexity is still $O(n)$. Full explanation is available in [6].

III. LSH-BASED APPROACH

To improve the efficiency of dictionary indexing we propose an approach based on special type of constructing the hash indexes. The idea is to construct such an index that works in opposite way to MD5 signature [14]. If the source strings differ slightly, the output hash would be also modified slightly. One way to create such hash function can be based on Locality-Sensitive Hashing [15] used to reduce dimensionality of high-dimensional data. It can be used in many applications where nearest neighbors need to be effectively computed [16], such as in tasks of entity resolution, fingerprint comparison or finding similar documents.

The algorithm takes data and computes a hash, which is a lower-dimensional representation of a given input. The result must preserve similarity, i.e. if words are similar then their hashes must be similar as well. In contrast to conventional hashing functions, LSH tries to maximize probability of a collision for similar items.

The main goal of incorporating LSH idea to approximate string matching is to significantly improve time performance. The LSH-based approach to this task consists of three phases:

1) Shingling (described above)
2) Min-Hashing – converts large sets to short signatures, while preserving similarity
3) Locality-Sensitive Hashing – places similar words into the same bucket

Firstly, an empty set C of shingles is created. During shingling of each word from the dictionary, shingles are added to collection C, so that after this step C is a sorted set of all shingles that occur in the shingled dictionary. Every trigram has a corresponding unique number (index of the shingle in C).

Secondly, we take the indices of a word shingles and store them in a vector. This vector representation is an input for Min-Hash algorithm, which calculates the signature. It should be noticed that different strategies can be used here [17]. We used the most popular version of Min-Hash algorithm.

Min-Hash internally creates occurrence matrix (of size #shingles $\times$ #words) filled with ones at positions where a given string contains a specific shingle. An example of such a matrix is shown in Table I. Then, rows of the matrix are permuted n times and at every permutation, for each word (each column) an index of the first row containing 1 is saved. In result, each word has a signature, which is a vector of these indices.

TABLE I
SAMPLE OCCURRENCE MATRIX

	the	those	these
$$t	1	1	1
$th	1	1	1
the	1	0	1
he$	1	0	0
e$$	1	0	0
tho	0	1	0
hos	0	1	0
ose	0	1	0
se$	0	1	1
e$$	0	1	1
hes	0	0	1
ese	0	0	1

In the third step, we partition these signatures into b bands. Every band is hashed, using locality sensitive hashing, into one of k buckets. In this case, hash function is of the form: $f: \mathbb{Z}^{\lfloor \frac{n}{b} \rfloor} \rightarrow \mathbb{Z}$

We chose $n = 100, b = 20$ and we want each bucket contains about one hundred hashes, so k must be $\frac{dict\_size}{100}$.

Candidate words are in buckets, which contain at least one hash from word typed by user. Then, a similarity between given hash and all hashes in bucket is calculated. If the similarity exceeds the threshold, a string associated with the hash is added to the result list.

The complexity of LSH searching is sub-linear, better than that of previous approaches, although worse than O(1). In contrast to previously described algorithms, LSH has a big disadvantage – it does not guarantee that all matching words are in the result set. Bigger dictionary can cause worse results [18].

IV. EXPERIMENTS

In our experiments, we measured processing time depending on various settings of the compared algorithms. In the case of LSH algorithm, we also measured the quality of obtained results in terms of recall and precision.

In every test, we randomly choose words from polish dictionary for games containing over 2,700,000 words, taking into account words with length less or equal to 15. The dictionary is available online [19].

The processing times were measured ten times. Averages of them are reported below.

A. The time of constructing the search structure

At the very beginning, we tested the time of building searching structures needed for algorithms and how it depends on the number of words in a dictionary. The time for Levenshtein and CPMerge algorithms are the same as in the case of AllScan, because data preparation process is exactly the same.

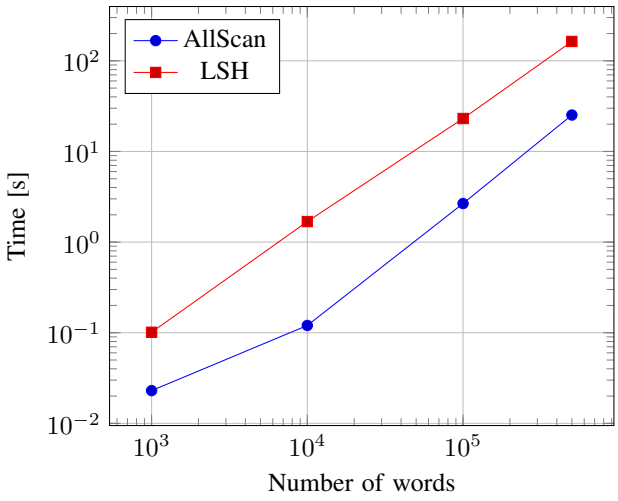


Fig. 1. Building time depending on the number of words

As we can see in Fig. 1, the time grows linearly in relation to the number of words with LSH having bigger coefficient.

B. Searching time

1) Dictionary size: The second test measured the search time of algorithms depending on the number of words in the dictionary. In this experiment we assumed the following values:

distance (allscan, cpmerge) = cosine
length of a given word ($|x|$) = 10
similarity threshold (α) = 0.7

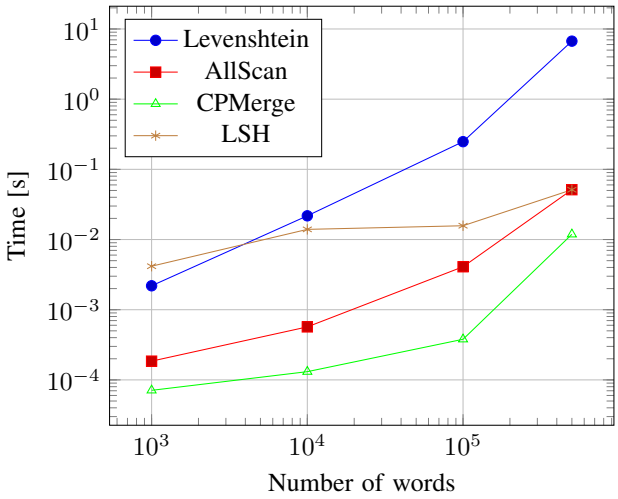


Fig. 2. Searching time depending on the number of words

Fig. 2 shows that the fastest algorithm is CPMerge. However, its search time grows squarely, what in comparison to LSH sub-linear growth allows us to state that LSH should be faster for big data.

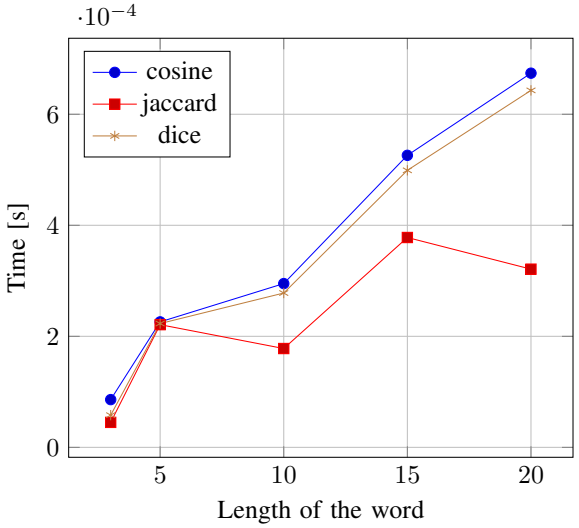


Fig. 3. Searching time depending on the length of a given word

2) Length of given words: To evaluate search time in the function of the length of given words we assumed the following values:

algorithm = cpmerge
number of words (V) = 100 000
similarity threshold (α) = 0.7

We measured the results for all three distance metrics.

Fig. 3 shows that processing time is the smallest for Jaccard distance, as the range of candidate words is the narrowest. All times grows linearly. Subtle deviations are caused by the number of candidate words changing stepwise in relation to word length.

3) Similarity threshold value: To evaluate search time depending on similarity threshold and distance metric we assumed:

searching algorithm = cpmerge
number of words (V) = 100 000
length of given words ($|x|$) = 15

As can be seen in Fig. 4, time falls in relation to similarity. This is because the number of candidate words is decreasing. The algorithm is the fastest for Jaccard distance due to the same reason as above, i.e. the smallest set of candidate words.

C. Number of words found

In this experiment we tested the impact of α similarity threshold on the number of returned similar strings. CPMerge algorithm had the following settings:

number of words ($|V|$) = 100 000
length of given words ($|x|$) = 10

Every given word had a typo in order to search for similar, but not identical, word.

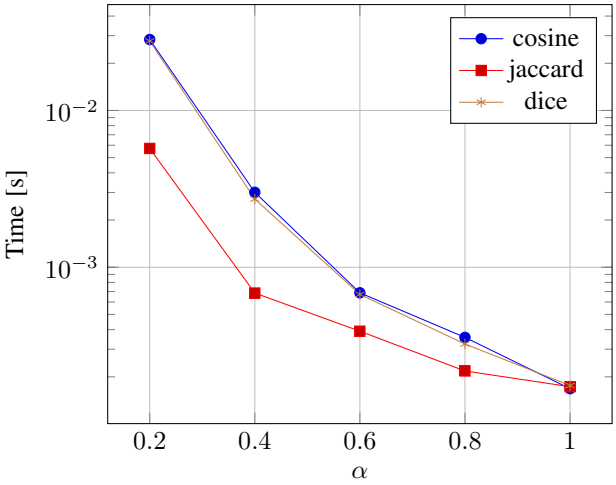


Fig. 4. Searching time depending on similarity threshold

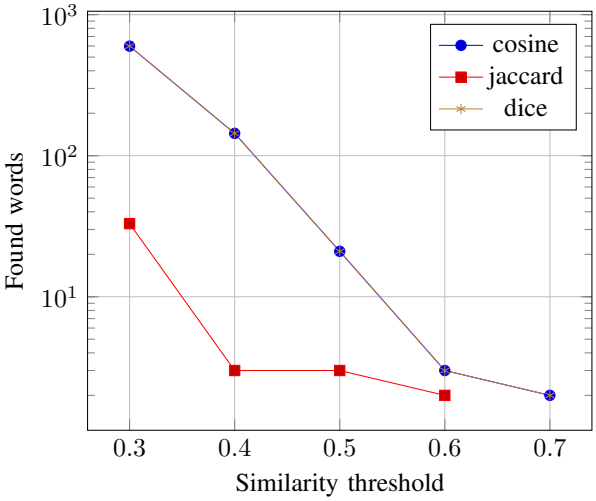


Fig. 5. Number of found word depending on similarity threshold

Fig. 5 shows that the number of found words falls exponentially in relation to assumed similarity threshold. As previously noticed, Jaccard distance returns the smallest set of candidates.

D. Recall and precision of LSH algorithm

We used the following measures for evaluating the quality of the approximate string matching of LSH algorithm: $recall = \frac{found\&relevant}{relevant}$ and $precision = \frac{found\&relevant}{found}$. Recall indicates what fraction of all matching words is returned in the result. Precision indicates what is the fraction of correct matches in the result set. In this task it is beneficial to maximize both recall and precision, therefore we also show their balanced harmonic mean, i.e. F-measure: $F_1 = \frac{2*P*R}{P+R}$.

1) Dictionary size impact: We assumed:
number of hash functions (n) = 100
number of bands (b) = 20
similarity threshold (α) = 0.7
number of buckets (k) = $\frac{number\_of\_words}{100}$

TABLE II
LSH RECALL AND PRECISION DEPENDING ON DICTIONARY SIZE

	100	1000	10000	100000
relevant	89	93	93	105
found	76	69	48	48
precision	1	1	1	1
recall	0.85	0.74	0.52	0.46

2) Similarity threshold impact: We assumed the following values:
number of words (V) = 100 000
number of hash functions (n) = 100
number of bands (b) = 20
number of buckets (k) = 1000

TABLE III
LSH RECALL AND PRECISION DEPENDING ON SIMILARITY THRESHOLD

	0.4	0.5	0.6	0.7
relevant	10911	1806	261	103
found	453	189	91	46
precision	1	1	1	1
recall	0.04	0.10	0.35	0.45

Table III shows that recall for low threshold is very low, which means LSH algorithm, in this configuration, is almost useless.

E. Wikipedia editors misspellings correction

In the following two tests we used a list of misspellings made by Wikipedia editors [20].

1) Comparison of algorithms: We set: number of words (V) = 1922
number of misspellings = 2455
similarity threshold (α) = 0.5
distance (cpmerge) = cosine
number of hash functions (lsh: n) = 100
number of bands (lsh: b) = 20
number of buckets (lsh: k) = 19

TABLE IV
COMPARED ALGORITHMS AT WIKI TYPOS CORRECTION TASK

	Levenshtein	AllScan & CPMerge	LSH
recall	0.89	0.89	0.74
precision	0.56	0.48	0.64
F-measure	0.69	0.63	0.69

Allscan results are the same like results of CPMerge algorithm, because they differ only in the processing time.

2) Similarity threshold impact on CPMerge: The setting were:

searching algorithm = CPMerge
distance (cpmerge) = cosine
number of words (V) = 1922
number of misspellings = 2455

TABLE V
CPMERGE AT WIKI TYPOS CORRECTION TASK DEPENDING ON SIMILARITY MEASURE

	0.2	0.3	0.4	0.5	0.6	0.7
recall	0.99	0.98	0.94	0.89	0.81	0.67
precision	0.03	0.12	0.29	0.48	0.64	0.63
F-measure	0.06	0.21	0.44	0.63	0.71	0.65

V. CONCLUSIONS

The most straightforward approach to approximate string matching is to incorporate some edit distance, like Levenshtein. Nevertheless, searching time of this method is the highest, because it has to browse the whole dictionary. $O(n)$ complexity disqualifies this approach for big dictionaries. AllScan and CPMerge reduces the size of the set of words to process, so they are faster than Levenshtein-based approach. However, their processing time still very much depends on dictionary size. All three methods have a significant advantage – they correctly (depending on similarity measure in use) return all words similar to a given string within assumed margins.

On the other hand, LSH provides results in acceptable time, but it degrades recall value. This is the main disadvantage of the algorithm, which causes that many similar words are not in the returned list.

We wish to have a solution, that is fast and that provides recall on very high level. To reach this goal we propose and test an initial solution that employs feed-forward neural network as indexer. The network takes a word as input and returns its index in the dictionary. With constant searching time, recall above 97% and high precision, this method can potentially beat all algorithms compared in this article in the task of approximate string matching. The idea extending the approach based on feedforward network is to use a denoising autoencoder. The approach based on on auto-associative network reconstructs its input given at the output layer through a bottleneck hidden layer, while the input additionally contains some noise, i.e. misspelled words in this case. In this approach, we would learn a network using one exact word and some words with typos and expect a correct word to be reconstructed.

REFERENCES

[1] W. H. Gomaa and A. A. Fahmy, "A survey of text similarity approaches," *International Journal of Computer Applications*, vol. 68, no. 13, 2013.
[2] A. Parker *et al.*, "Computer algorithms for plagiarism detection," 1989.
[3] Z. Su, B.-R. Ahn, K.-Y. Eom, M.-K. Kang, J.-P. Kim, and M.-K. Kim, "Plagiarism detection using the levenshtein distance and smith-waterman algorithm," in *Innovative Computing Information and Control, 2008. ICICIC'08. 3rd International Conference on*. IEEE, 2008, pp. 569–569.
[4] F. D. Garcia, J.-H. Hoepman, and J. Van Nieuwenhuizen, "Spam filter analysis," in *Security and Protection in Information Processing Systems*. Springer, 2004, pp. 395–410.
[5] T. Okuda, E. Tanaka, and T. Kasai, "A method for the correction of garbled words based on the levenshtein metric," *Computers, IEEE Transactions on*, vol. 100, no. 2, pp. 172–178, 1976.
[6] N. Okazaki and J. Tsujii, "Simple and efficient algorithm for approximate dictionary matching," *Proceedings of the 23rd International Conference on Computational Linguistics*, pp. 851–859, 2010.
[7] V. I. Levenshtein, *Binary codes capable of correcting deletions, insertions, and reversals*. Soviet Physics Doklady, 1966.
[8] F. J. Damerau, "A technique for computer detection and correction of spelling errors," *Communications of the ACM*, vol. 7, no. 3, pp. 171–176, 1964.
[9] G. V. Bard, "Spelling-error tolerant, order-independent pass-phrases via the damerau-levenshtein string-edit distance metric," in *Proceedings of the fifth Australasian symposium on ACSW frontiers-Volume 68*. Australian Computer Society, Inc., 2007, pp. 117–124.
[10] G. Navarro, "A guided tour to approximate string matching," *ACM computing surveys (CSUR)*, vol. 33, no. 1, pp. 31–88, 2001.
[11] K. Monostori, R. Finkel, A. Zaslavsky, G. Hodász, and M. Pataki, "Comparison of overlap detection techniques," in *Computational Science—ICCS 2002*. Springer, 2002, pp. 51–60.
[12] C. Badica, N. T. Nguyen, and M. Brezovan, Eds., *Computational Collective Intelligence. Technologies and Applications - 5th International Conference, ICCCI 2013, Craiova, Romania, September 11-13, 2013, Proceedings*, ser. Lecture Notes in Computer Science, vol. 8083. Springer, 2013.
[13] S.-S. Choi, S.-H. Cha, and C. C. Tappert, "A survey of binary similarity and distance measures," *Journal of Systemics, Cybernetics and Informatics*, vol. 8, no. 1, pp. 43–48, 2010.
[14] B. Kaliski and M. Robshaw, "Message authentication with md5," *CryptoBytes (RSA Labs Technical Newsletter)*, vol. 1, no. 1, 1995.
[15] A. Gionis, P. Indyk, R. Motwani *et al.*, "Similarity search in high dimensions via hashing," in *VLDB*, vol. 99, no. 6, 1999, pp. 518–529.
[16] M. Slaney and M. Casey, "Locality-sensitive hashing for finding nearest neighbors [lecture notes]," *Signal Processing Magazine, IEEE*, vol. 25, no. 2, pp. 128–131, 2008.
[17] L. Paulevé, H. Jégou, and L. Amsaleg, "Locality sensitive hashing: A comparison of hash function types and querying mechanisms," *Pattern Recognition Letters*, vol. 31, no. 11, pp. 1348–1358, 2010.
[18] J. Leskovec, A. Rajaraman, and J. D. Ullman, *Mining of massive datasets*. Cambridge University Press, 2014.
[19] SJP, http://sjp.pl/slownik/growy/, 2016, [Online; 01-07-2016].
[20] Wikipedia, https://en.wikipedia.org/wiki/Wikipedia:Lists_of_common_misspellings, 2016, [Online: 01-07-2016].

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 533–540
DOI: 10.15439/2016F391
ACSIS, Vol. 8. ISSN 2300-5963

Grammatical Case Based IS-A Relation Extraction with Boosting for Polish

Paweł Łoziński, Dariusz Czerski, Mieczysław A. Kłopotek
Institute of Computer Science
Polish Academy of Sciences
ul. Jana Kazimierza 5, 01-248 Warsaw, Poland
Email: {pawel.lozinski, dariusz.czerski, mieczyslaw.klopotek}@ipipan.waw.pl

***Abstract*—Pattern-based methods of IS-A relation extraction rely heavily on so called Hearst patterns. These are ways of expressing instance enumerations of a class in natural language. While these lexico-syntactic patterns prove quite useful, they may not capture all taxonomical relations expressed in text. Therefore in this paper we describe a novel method of IS-A relation extraction from patterns, which uses morpho-syntactical annotations along with grammatical case of noun phrases that constitute entities participating in IS-A relation. We also describe a method for increasing the number of extracted relations that we call *pseudo-subclass boosting* which has potential application in any pattern-based relation extraction method. Experiments were conducted on a corpus of about 0.5 billion web documents in Polish language.**

I. Introduction

RELATION extraction is a necessary step of any ontology induction or taxonomy induction task. Typically it takes as input morpho-syntactically annotated text and produces a set of triples (E_1, R, E_2), where E_1 and E_2 are entities and R is a relation in which E_1 and E_2 participate as a pair. In case of ontology induction or information extraction in open domain (as described, e.g., in [1], [2], [3], [4]) no restrictions are imposed on R. There are many types of relations that can be extracted this way, such as *quality*, *part* or *behavior* [5]. In case of taxonomy induction the main interest is in the IS-A (hyponym-hypernym) relation. Approaches to IS-A extraction described in literature rely on evidence from pattern extraction and statistical information (cf. [6], [7], [8]). In methods that are based solely on statistical information it is not uncommon to assume (cf. [7]), that relation extraction is performed only for a predefined list of concepts extracted earlier with a different method (e.g. [9]). Pattern-based methods rely heavily on so called Hearst patterns, first described in [10]. These are ways of expressing instance enumerations of a class in natural language. Typical forms are „c *such as* i1, i2 or i3" or „c, *for example* i1, i2 or i3". Terms extracted with such patterns may serve as input for elaborate taxonomy and ontology construction methods as, e.g., [11]. While these lexico-syntactic patterns prove quite useful, they may not capture all taxonomical relations expressed in text. Therefore in this paper we describe a novel method of IS-A relation extraction from patterns, which uses morpho-syntactical annotations along with grammatical case of noun phrases that constitute entities participating in IS-A relation. As it will be shown in the paper, the method allows for extraction of additional knowledge from text, that is often not expressed with Hearst patterns. The method is unsupervised, as it is based on hand-crafted patterns, dictionary filtering and manually adjusted support level. Precision of this method, understood as the ratio of correct extracted IS-A relations to all extracted relations is estimated using manual scoring of about 110 relations randomly selected from the method's output. Based on an internet corpus of documents, the method produces a big number of IS-A relations. Most of them (roughly 90%) occur only once in the corpus introducing a high level of noise. We show in conducted experiments that even for a slight increase of support (given as a number of occurrences), the estimated precision of this method increases strongly. We also describe a new method for increasing the number of extracted relations for any support level bigger than 1. The method is based on very simple heuristic for detection of hyponymy between class part of extracted relations, thus we call it *pseudo-subclass boosting* (*PSC* in short). It is worth mentioning that this boosting approach can be applied in any pattern-based relation extraction method. Experiments were conducted on a corpus of about 0.5 billion web documents in Polish language crawled in NEKST project (http://www.nekst.pl) and maintained up to date. These include primarily HTML documents, but also other formats found on websites like PDFs and DOCs. In order to process such high volume of data it was implemented using MapReduce framework [12] implemented in Apache Hadoop project (http://hadoop.apache.org) and Hive (http://hive.apache.org). All examples mentioned in the article are real data, taken from working instance of NEKST system.

II. Our approach

It is known that languages that have inflection and free word order are much harder for automatic analysis[1] than, e.g., English. As pointed out in [14, pp. 100], free word order implies non-projective grammar. It is shown in [15] and [16] that dependency parsing for non-projective grammars is NP-hard, apart from a very narrow subclass called edge-factored grammars. This challenge is addressed, among others, by transition-based dependency parsing [17] used in the pre-processing step for the algorithm described in this paper. We argue that inflection in a language is not only a drawback but

The study is cofounded by the European Union from resources of the European Social Fund. Project PO KL „Information technologies: Research and their interdisciplinary applications", Agreement UDA-POKL.04.01.01-00-051/10-00.

[1]See e.g. [13] for problems with relation mining in German, in which the word order is much less free than in Polish; note that they use an initial lexicon while we do start from scratch when extracting relations.

TABLE I. SUFFIXES IN INSTRUMENTAL CASE FOR POLISH

	masculine	neuter	feminine
singular	-em		-ą
plural	-ami (-mi)		

can also be a great advantage. Typical constructs that express the hypernymy relation explicitly in Polish language are:

$$NP_1^{Nom} \text{ to } NP_2^{Nom}, \tag{1}$$

$$NP_1^{Nom} \text{ jest } NP_2^{Abl}. \tag{2}$$

Both of them are a way of saying NP_1 *is* NP_2 and in both cases noun phrase NP_1 is expressed in nominative. They differ in grammatical case of NP_2, where in the first construct we have nominative and in the second: instrumental. The second pattern has its equivalent for past tense:

$$NP_1^{Nom} \text{ był/była/było } NP_2^{Abl}. \tag{3}$$

Obviously in case of past tense construction it is possible that IS-A relation no longer holds[2]. The problem exists to a lesser extent also in present tense, which for example can be a consequence of outdated web documents. Assessment of correctness with respect to a given point in time is, in our opinion, a research direction of its own, thus it is out of scope of this paper.

As will be shown later, combination of word and grammatical case pattern allows for relation extraction with quite high precision. It is possible partially thanks to the fact that instrumental case in Polish language is *regular* for nouns and has unique suffixes shown in Table I (after [18, pp. 145, 148]). This makes automatic analysis of sentence tokens easy for this case.

We propose a rule-based approach for IS-A relation extraction with the following procedure:

- run each sentence in corpus through POS-tagger and dependency parser,
- select dependency trees with promising structure,
- apply dictionary filtering for the head of NP_2,
- apply a set of construction rules to dependency tree in order to build instance name out of NP_1 and class name out of NP_2,
- apply a set of filtering rules.

This method is additionally extended with a technique that we call *pseudo-subclass boosting* which increases the number of extracted relations.

It is worth noting that *automatic* detection of IS-A patterns is possible. Experiments described in [19] show that handcrafted ontologies like WordNet can be used successfully as a training set for such pattern discovery task. However, our problem setting differs from that research significantly. Apart from the already mentioned inflection challenge and free word order language, our corpus consists of about 11 billion sentences, which is four orders of magnitude more than the Reuters corpus used in [19] and imposes efficiency limitations. On the other hand, the gain in size comes at the price of quality – Internet documents tend to have much more noisy content than printed journal articles. We have no knowledge of any research on IS-A patterns detection in similar setting (that is web-scale), which leads us to first tackle a more realistic problem of extracting IS-A relations with *known* patterns. Nevertheless, this is a task worth trying given experience gained from research reported here.

[2]The relation was valid in the past only

A. POS tagging and dependency parsing

For part-of-speech tagging we use the Apache OpenNLP (http://opennlp.apache.org) tagger trained with Maximum Entropy classifier on NKJP [20] corpus. Additionally, for known words, we optimized the tag disambiguation process by narrowing tags that can be chosen by information taken from the PoliMorf dictionary [21]. For Polish language, whose tagset contains around 1000 tags [22], this simple optimization gives an improvement of tagging in terms of accuracy and processing speed at the same time. To give an example, the word *artykułów* (inflected form of the word *article*) has only two possible tags `subst:pl:gen:m3` and `subst:pl:gen:p3`. Using this knowledge in OpenNLP tagger reduces search space for this word 500 times. Dependency parsing is based on MaltParser framework [23] trained on Polish Dependency Bank that consists of 8030 sentences [24]. To obtain high processing speed (essential for such large volume of text data) the liblinear classification model has been used.

B. Promising dependency tree structure selection

By *promising* structure of a dependency tree we mean one that matches any of the patterns depicted in Figures 1, 2 and 3, where **form**, **dep** and **pos** mean: token form, dependency relation type (as described in [24]) and part-of-speech tag (as described in [20]) respectively.

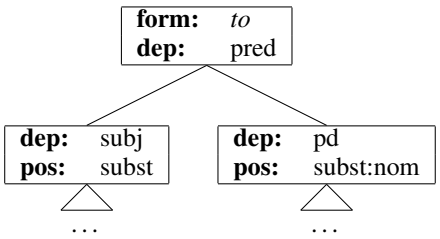


Figure 1. Dependency tree structure for construct (1)

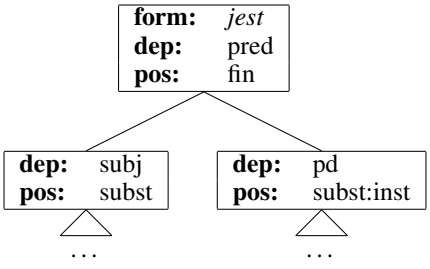


Figure 2. Dependency tree structure for construct (2)

In both nominative and instrumental case, the base structure has a predicate word with outgoing dependency arcs to

two other words with subjective and predicative complement relation type. The difference between structure 1, 2 and 3 is in the grammatical case of the predicative complement and part of speech of the predicate. Our intuition is that selected structures are natural sources of IS-A relation. This claim is supported by the estimated precision obtained in conducted experiments.

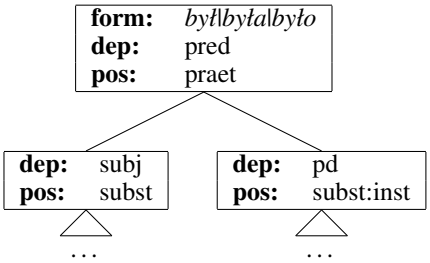


Figure 3. Dependency tree structure for construct (3)

Figure 4 illustrates an example of sentence that matches pattern 2, parsed with our dependency parser and printed in CoNLL [25] format. It is worth noting that in this case the part-of-speech tagger made an error in assigning a case to the adjective *myśliwski* (hunting), where instrumental instead of locative should appear. This may happen because singular masculine adjective suffixes for instrumental (as noted in [18, p. 160]) are not unique as with nouns. That's why in our analysis we focus only on the grammatical case of the head of noun phrase and assume the same case for its dependent adjective tokens. This assumption is justified by the fact, that for Polish language agreement exists between noun and adjective in a noun phrase [26, p. 174]. POS tag in the example is repeated twice because CoNLL format specifies CPOSTAG and POSTAG allowing for coarse-grained and fine-grained part-of-speech tagsets which are the same for Polish language. The following steps illustrate how pattern 2 applies to the example sentence from figure 4:

- find a root word of the sentence (*jest* in our case), and check its dependency relation (must be *pred*) and a POS tag (must be *fin*),
- if the root word has two descendants, then test if:
  - its left descendant (*golden*) has correct dependency relation (must be *subj*) and a POS tag (must be *subst*),
  - its right descendant (*pies*) has correct dependency relation (must be *pd*) and a POS tag (must be *subst:inst*),
- if all requirements are fulfilled, the sentence is moved to the phase of dictionary filtering (section II-C) and instance and class name construction (section II-D).

Given a sentence whose dependency tree matches one of above-mentioned patterns, we construct NP_1 from its left sub-tree and NP_2 from its right sub-tree. Head (or root) of left and right sub-tree will be denoted N_1^H and N_2^H respectively.

C. Dictionary filtering for the head of NP_2

Preliminary experiments showed that many of sentences matching constructs (1) and (2) contain very general, ambiguous nouns in NP_2 like *problem*, *aspect*, *element* or *outcome*. Those nouns cannot be considered proper classes in the sense of IS-A relation, rather they are catch-all phrases used to express various thoughts about what is contained in NP_1.

We eliminated those nouns by manually evaluating a random sample of about 1000 experiment results and creating a dictionary of such meaningless „classes". In this step of our extraction procedure we filter extractions with this dictionary. This process was repeated in three iterations. Size of the dictionary started with 95 catch-all phrases increased by 50, and 20 reaching the level of about 170.

D. Construction rules for NP_1 and NP_2

We construct both instance name (from NP_1) and class name (from NP_2) out of lemmatized tokens. The first step is to serialize tokens present in both dependency sub-trees with operators *leftOffspring* and *rightOffspring*, which operate as follows:

1) put all nodes of dependency sub-tree in a list L,
2) sort L by CoNLL token id descending (for *leftOffspring* operator)/ascending (for *rightOffspring* operator),
3) find index i_H of sub-tree head in L,
4) create sub-list L' from i_H to the first occurrence of interpunction or end of L,
5) in case of *leftOffspring*: sort L' by CoNLL token id ascending,
6) concatenate lemmas of tokens in L' and return.

Computational complexity of this algorithm is $O(n)$, where n is the sentence length. Actual sorting of tokens in case of steps 2. and 5. is not necessary and was introduced to simplify the description[3]. Boundaries detection of instance name is quite simple because it is typically directly defined by left sub-tree of all considered dependency structures (Figures 1, 2 and 3). Therefore it is constructed as concatenation:

$$leftOffspring(N_1^H) + N_1^H + rightOffspring(N_1^H)$$

Creation of class name is more complicated as it is often preceded by degrees of comparison and followed by the rest of the sentence which may be loosely coupled with the class itself. Consider the following sentences:

> Trójmorski Wierch jest jedyną polską górą, z której spływają wody aż do trzech mórz.

(*Trójmorski Wierch is the only Polish mountain, from which waters flow to as many as three seas.*)

> Korona norweska to waluta oznaczana międzynarodowym kodem – NOK.

(*Norwegian krone is a currency marked with the international code – NOK.*)

In the first example, the word *jedyną* (the only) cannot be considered as part of class name. Likewise, anything that comes after word *waluta* (currency) in the second example is merely a description of Norwegian krone, not part of a class name. To address such issues construction rules for class name simply omit the output of *leftOffspring* operator and truncate

[3]It unifies the procedure for left and right part of the sentence.

1	Golden	golden	subst	subst	sg:nom:m3	3	subj
2	retriever	retriever	subst	subst	sg:nom:m2	1	app
3	jest	być	fin	fin	sg:ter:imperf	0	pred
4	psem	pies	subst	subst	sg:inst:m2	3	pd
5	myśliwskim	myśliwski	adj	adj	sg:loc:m3:pos	4	adjunct
6	.	.	interp	interp	_	3	punct

Figure 4. Tree pattern match example in CoNLL format for the Polish sentence "Golden retriever jest psem myśliwskim" (*Golden retriever is a hunting dog*). Note that the parser did not produce fully correct dependency tree (e.g. "Golden" is tagged as noun and linked directly with "jest"). This does not affect our extraction process.

rightOffspring output: it is iterated from left to right only as long as the tokens have POS tag from set {adj, subst, ger} *and* dependency type from set {adjunct, app, conjunct, obj}. So the class name results from concatenating:

$$N_2^H + truncate(rightOffspring(N_2^H))$$

This forces extraction of shorter phrases, which increases the probability of observing a given instance-class pair more than once. As we show in section III, this highly influences the precision of the method. Extraction results for above examples are: *Trójmorski Wierch* `IS-A` *góra* [*Trójmorski Wierch IS-A mountain*] and *Korona norweska* `IS-A` *waluta* [*Norvegian crown IS-A currency*], while from such sentence:

> Narodowy Bank Belgijski jest bankiem centralnym od 1850 roku.[4]

we acquire *Narodowy Bank Belgijski* `IS-A` *bank centralny* [*Belgian National Bank IS-A central bank*].

E. Final filtering rules

It is common that NP_1 contains reference to earlier parts of text. Two types of such reference can be distinguished:

1) explicit:
   > Ten wikipedysta jest numizmatykiem.[5]
2) implicit:
   > Pisarka jest członkiem Związku Pisarzy Białorusi.[6]

In both cases NP_1 typically contains a class of referenced entity, not the entity itself which leads to erroneous extractions. As long as this reference is explicit, we filter such cases with a dictionary of referencing words (pronouns and textual references like *above-mentioned*). The case where reference is implicit is much harder, and at this point left for further research, as described later in section VI.

F. Pseudo-subclass (PSC) boosting

Our experiments showed that the number of extracted relations drops significantly with increase of support level t. To compensate this loss we designed a boosting method that is based on the following intuition: if *I* `IS-A` *C* and *I* `IS-A` *C'* are extracted relations and *C* is a substring of *C'*, then there is high chance that *C'* is a way of describing *I* more precisely than *C*, i.e., *C'* is a pseudo-subclass of *C*. If so, we can boost our confidence in the fact that *I* `IS-A` *C* is properly extracted. To give an example:

> Kraków to najchętniej odwiedzane miasto przez turystów w Polsce. Kraków – dawna stolica Polaków jest miastem magicznym.[7]

Above two sentences allow for boosting confidence in extraction *Kraków* IS-A *miasto* (*Cracow* IS-A *city*). From the first sentence we get the relation *Kraków* IS-A *miasto* and from the second *Kraków* IS-A *miasto magiczne* (*Cracow* IS-A *magic city*). As "miasto magiczne" is a superstring of "miasto", the second sentence supports the first extracted relation. In general, to detect class/pseudo-subclass matches for each extraction *R* = *I* `IS-A` *C* we generate a list *L* of

- prefix lists of tokens from *C*,
- suffix lists of tokens from *C* that don't include leading adjectives.

In Map phase of MapReduce job, we emit the pair *(I, C)* with *R*'s occurrence count and pairs *(I, c)* (with the same count) for each $c \in L$. Reduce phase aggregates our data by matched pairs and here we acquire knowledge about pseudo-subclasses' occurrence count and type of constructs they were discovered in. Figure 5 illustrates a more elaborate case of pseudo-subclass boosting. Each numbered row represents a relation *mukowiscydoza* IS-A ... extracted from text. Row 13 is an example of suffix list boosting with *wieloukładowa* being an adjective removed at the stage of creating list *L*. Rows 2-12 boost relation *mukowiscydoza* IS-A *choroba*, additionally rows 4-7 boost *mukowiscydoza* IS-A *choroba genetyczna*, etc.

III. Experiments

Experiments were conducted on a corpus of about 0.5 billion web documents in Polish language with roughly 11 billion sentences. Tables II, III and V present the results of passing the entire collection through the algorithm described in section II.

Method evaluation was conducted for four levels of the value of t, which, as earlier described, is the minimal IS-A relation occurrence count acceptance threshold. Precision evaluation was based on manual scoring of about 110 randomly selected relations from given experiment's results. Estimated precision was calculated by the formula 4.

$$\hat{Pr} = \frac{TP}{TP + FP} \tag{4}$$

[4]Belgian National Bank is the central bank since 1850.
[5]This wikipedian is a numismatist.
[6]The writer is a member of Union of Belarus Writers.
[7]Cracow is the most visited city by tourists in Poland. Cracow – the former capital of the Poles is a magical city.

```
mukowiscydoza (cystic fibrosis) IS-A
 1. choroba (disease)
 2.   choroba dziedziczna (hereditary disease)
 3.   choroba genetyczna (genetic disease)
 4.     choroba genetyczna ludzi rasy białej
        (genetic disease of white race people)
 5.     choroba genetyczna ogólnoustrojowa (systemic genetic disease)
 6.     choroba genetyczna rasy białej (genetic disease of white race)
 7.     choroba genetyczna układu pokarmowego
        (genetic disease of the digestive system)
 8.   choroba monogenowa (monogenic disease)
 9.   choroba nieuleczalna (incurable disease)
10.   choroba przewlekła (chronic disease)
11.   choroba wielonarządowa (multiorgan disease)
12.   choroba wieloukładowa (multisystem disease)
13.   wieloukładowa choroba (multisystem disease)
14.     wieloukładowa choroba monogenowa
        (multisystem monogenic disease)
15. przyczyna wykonywania (cause of performing)
16.   przyczyna wykonywania przeszczepu płuca
      (cause of performing lung transplant)
17. schorzenie (disease - synonym)
18.   schorzenie genetyczne (genetic disease - synonym)
```

Figure 5. Tree representation of pseudo-subclass boosting.

where TP is the number of relations scored as correct and FP is the number of relations scored as erroneous. Note that we cannot compute other traditional measures as accuracy, recall or F-measure. This is due to the fact, that in Open Relation Extraction setting the number of false negatives (relations incorrectly left out in the extraction process) is not known.

Tables II, III and IV show results of these experiments. Column *nom* contains number of unique IS-A relations extracted only from nominative construct, *inst* is the number of unique relations only from instrumental constructs, *nom∩inst* refers to count of relations extracted from nominatives and instrumentals. Table III refers to the number of relations that were additionally accepted only thanks to pseudo-subclass boosting which helped to observe a given relation more than t times or with both grammar cases.

Total number of extracted IS-A relations, for either nominative or instrumental construction, is slightly above 4 milion (table II). Increase of support level results in drop of accepted relations (up to 1 order of magnitude between consecutive levels). Final count of relations (for $t = 4$) does not exceed 90000, which is almost 2 orders of magnitude lower than the total.

Pseudo-subclass boosting method allows to extract around 86000 more relations at support level 2. Nominal number of additional relations decreases for higher support levels, but increases in terms of relative gain (as shown in the last column of table III).

Estimated precision of our method is 61% at the lowest support level, and achieves 87% for level 4 (table IV). Increasing the number of accepted relations with pseudo-subclass boosting comes at the cost of lower estimated precision. At support level 2 this loss is 1%, but for 3 and 4 jumps to several percent. Estimated precision of our method, equipped with pseudo-subclass boosting, increases with the increase of t, saturating at the level of about 80%. Table IV contains also estimated precision of our implementation of Hearst patterns which is substantially lower (from 14% to 29%).

Experiments were performed on a cluster of 70 machines with total of 980 CPU cores and 4.375TB of RAM. Total processing time of raw web documents: lemmatization, POS tagging, dependency parsing and IS-A relation extraction was under 24 hours.

IV. Relation to Hearst patterns

In order to compare our method with the most popular approach, we implemented Hearst patterns extraction algorithm as follows:

- Detect enumeration phrase R (one of „*taki jak*", „*taki jak na przykład*", „*taki jak np.*" which are special cases of phrase "*such as*" in English) in a sentence, based on lexical constructions proposed in [10].
- Check if words from R to the end of the sentence form a comma separated list of phrases (with the last element optionally separated by conjunction: „*i*" or „*oraz*"). The list is assumed to represent instances of a class.
- Detect the class name in words left to R with a Conditional Random Field model [27]. Words in this part of sentence are labeled with either „1" or „0". The sequence of „1" nearest to R is assumed to represent the class. The model was trained on manually annotated set of around 600 sentences. Its precision calculated on 10-fold cross validation is 93.89%.

Table V shows the number of extracted Hearst patterns and overlap between this method and our approach (percentage

TABLE II. NUMBER OF EXTRACTED RELATIONS FOR DIFFERENT VALUES OF MANUALLY ADJUSTED ACCEPTANCE SUPPORT LEVELS t. NUMBER OF RELATIONS EXTRACTED ARE GIVEN IN COLUMNS: "**NOM**" FOR NOMINATIVE CONSTRUCT AND "**INST**" FOR INSTRUMENTAL CONSTRUCTS. COLUMN "**NOM∩INST**" CONTAINS THE NUMBER OF RELATIONS EXTRACTED WITH BOTH NOMINATIVE AND INSTRUMENTAL CONSTRUCTS.

	nom	**inst**	**nom∩inst**	**total**
$t = 1$	1647500	2380021	39865	4027521
$t = 2$	138877	264764	9895	403641
$t = 3$	52430	100320	4938	152750
$t = 4$	29210	55232	3154	84442

TABLE III. NUMBER OF ADDITIONAL RELATIONS EXTRACTED THANKS TO PSEUDO-SUBCLASS BOOSTING (FOR DIFFERENT VALUES OF SUPPORT LEVEL t). COLUMN "**NOM**" CONTAINS RESULTS FOR NOMINATIVE CONSTRUCT AND "**INST**" FOR INSTRUMENTAL CONSTRUCTS. COLUMN "**NOM∩INST**" CONTAINS THE NUMBER OF ADDITIONAL RELATIONS EXTRACTED WITH BOTH NOMINATIVE AND INSTRUMENTAL CONSTRUCTS.

	nom	**inst**	**nom∩inst**	**total**	**PSC gain**
$t = 1$	0	0	0	0	0%
$t = 2$	24335	61244	2931	85579	21.20%
$t = 3$	13122	38004	2116	51126	33.47%
$t = 4$	8726	26702	1521	35428	41.95%

TABLE IV. ESTIMATED PRECISION ($\hat{Pr}$ – SEE EQUATION 4) OF EXTRACTION FOR DIFFERENT ACCEPTANCE SUPPORT LEVELS. "PSC" STANDS FOR PSEUDO-SUBCLASS BOOSTING. OUR APPROACH IS MARKED WITH "**NOM∩INST**", WHILE "**HRST**" STANDS FOR HEARST PATTERNS.

t	**nom+inst** (no PSC)	**nom+inst** (with PSC)	**hrst**
1	0.61	0.61	0.47
2	0.71	0.72	0.56
3	0.87	0.79	0.58
4	0.87	0.81	0.62

values in brackets are calculated relative to the number of Hearst patterns-based extractions). The overlap varies from 0.57% to 1.02% for nominative scheme and from 1.19% to 2.65% for instrumental. Relations detected in all three methods constitute from 0.25% to 0.58% of relations extracted with the basic method. This suggests that our method allows for extraction of new relations, not expressed in language constructs described by Hearst, with even higher precision.

V. DISCUSSION

Experiments lead to interesting conclusions. Firstly, there is little intersection between IS-A relations extracted by the three methods: Hearst traditional method and our methods, one based on nominative, the other based on instrumental case. The IS-A relation space seems too sparse for such methods to produce overlapping results. Nominative construction produces less relations than instrumental, which presumably is a consequence of the fact that this construct is only applicable for present tense. Decrease in total extractions count is much bigger going from support level 1 to 2 (9.98 times) than when in other cases ($2 \rightarrow 3$: ~2.64 times, $3 \rightarrow 4$: ~1.81 times). It can be connected to the natural model of language, where distribution of word frequencies has power law probability distribution [28]. There is a lot of particular, domain specific taxonomical information that is infrequent in textual resources accessible on the Internet. On the other hand more common knowledge that can be found multiple times in text is substantially less frequent.

Of course pseudo-subclasses don't give any boost when $t = 1$ and do not affect precision, because we simply accept everything that passes the final filtering rules. In other cases PSC increases the number of extractions significantly (the higher t the better), although not as much as to eliminate the effect of increased t. This boosting method is very beneficial for support level 2 as it increases extractions count by 23% with no observable loss in precision (see Table IV). For $t = 3$ and $t = 4$ the gain in extractions count comes at the price of significantly lower precision.

Analysis of false-positive extractions reveal several types of errors made by this method:

1) Implicit reference – which leads to errors like
   - *autor* `IS-A` *dyrektor jednostki* (*author* `IS-A` *director of the unit*),
   - *sobota* `IS-A` *dzień koncertu głównego* (*Saturday* `IS-A` *main concert day*).
2) Wrong decision about phrase begin/ending point[8]:
   - *trening funkcjonalny* `IS-A` *rodzaj (. . . czego?)* (*functional training* `IS-A` *kind (. . . of what?)*),
   - ~~*zdecydowana większość*~~ *kandydatów do Parlamentu* `IS-A` *członek określonej partii politycznej* (~~vast majority of~~ *candidates to Parliament* `IS-A` *member of a particular political party*).
3) Ever growing dictionary mentioned in section II-C. After each iteration of catch-all phrases eliminations new such phrases emerge in result samples. Above-mentioned experiments revealed such false-positive classes as: *result*, *an essential element* and *something amazing*. The number of such phrases decreased in each dictionary-construction iteration, which allows us to assume that this set is relatively small. Nonetheless, we are aware that manual construction of this set doesn't take evolution of the language's vocabulary into account.

VI. FUTURE WORK

Plans for future development include dealing with issues detected in above-mentioned experiments. The problem of detecting implicit references to earlier parts of text is known in natural language processing as coreference resolution and

[8]Missing parts are added in brackets, unwanted parts are striked out.

TABLE V. NUMBER OF RELATIONS EXTRACTED WITH HEARST PATTERNS FOR DIFFERENT VALUES OF MANUALLY ADJUSTED ACCEPTANCE SUPPORT LEVELS t.

	hrst	**nom∩hrst**	**inst∩hrst**	**nom∩inst∩hrst**
$t = 1$	4007927	23044 (0.57%)	47953 (1.19%)	10222 (0.25%)
$t = 2$	781419	6492 (0.83%)	15567 (1.99%)	3434 (0.44%)
$t = 3$	356873	3488 (0.98%)	8728 (2.45%)	1899 (0.53%)
$t = 4$	224200	2295 (1.02%)	5939 (2.65%)	1298 (0.58%)

constitutes an independent field of research as described in [29, p. 614] or specifically for Polish: [30]. It is planned to adapt selected coreference resolution methods to our BigData environment and verify their effectiveness in increasing precision of our extraction method.

We plan to achieve better detection of phrase begin/ending points by replacing construction rules described in section II-D with Conditional Random Field classifier trained on sentences scored in our experiment with manually annotated proper phrase boundaries. Creating of such golden standard set of sentences with IS-A relations is of course more time consuming than the approach proposed in this paper. In case of Hearst patterns it turned out to be a necessity. Sentences with Hearst-like enumerations contain more complicated dependency structures which are harder to parse correctly.

Better catch-all phrases elimination can be done as a post-processing step. Membership in these classes should be uniformly distributed over instances and subclasses in the taxonomy, so there should be no significant correlation between membership in these classes and proper classes. Filtering methods based on such correlation will be investigated.

Taking into account the number of filtered out IS-A relations (starting from support level 2) it is worthwhile to consider development of other ways of assessing their correctness. The support level criterion (frequency based) effectively increases quality of extracted information, but at the same time significantly reduces its quantity. It would be interesting to choose one of the most popular classification methods (ea. Support Vector Machine or Random Forest classifier) and check its ability to learn a more sophisticated filtering criterion of incorrect IS-A relations. The feature space for this classification problem could be much richer than simple information about occurrence frequency. One can use more sophisticated characteristics of IS-A relation like for example: size of class and instance phrase (count in number of words), type of sources (nominative, instrumental), popularity of instance and class phrase independently (expressed in number of occurrences among all extracted IS-A relations).

It would be also interesting to compare precision of Hearst patterns implemented with pseudo-subclass boosting.

VII. CONCLUSIONS

This paper presents a novel method of IS-A relation extraction from patterns for Polish that is different from so popular Hearst patterns and is applicable in inflected languages with free word order. Thanks to this method we were able to extract knowledge that may not be expressed in enumeration constructs defined by Hearst. Additionally, a method for boosting relation extractions count is introduced. As mentioned at the beginning, thanks to its simplicity it has potential application in any pattern-based IS-A relation extraction method. As experiments showed, the algorithm achieves satisfactory precision [9] (although there is still room for improvement) and is capable of generating high number of taxonomical relations. This makes it a valuable input source of data for any taxonomy induction task.

It is needless to say that experiments described in this paper do not provide a full statistical overview of millions of IS-A relations extracted from the corpus of Polish Internet documents. We focus on an assessment of precision of the proposed IS-A relation extraction method. In-depth statistical analysis of such a dataset is desirable and remains as a task to be accomplished in the next publication devoted to the research path outlined in the previous section.

REFERENCES

[1] H. Poon and P. Domingos, "Unsupervised ontology induction from text," in Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, 2010, pp. 296–305.

[2] A. Fader, S. Soderland, and O. Etzioni, "Identifying relations for open information extraction," in Proceedings of the Conference on Empirical Methods in Natural Language Processing, ser. EMNLP '11. Stroudsburg, PA, USA: Association for Computational Linguistics, 2011, pp. 1535–1545.

[3] M. Banko, M. J. Cafarella, S. Soderland, M. Broadhead, and O. Etzioni, "Open information extraction from the Web," in Proceedings of the 20th International Joint Conference on Artifical Intelligence, ser. IJCAI'07. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2007, pp. 2670–2676.

[4] O. Etzioni, A. Fader, J. Christensen, S. Soderland, and M. Mausam, "Open information extraction: The second generation," in Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence - Volume Volume One, ser. IJCAI'11. AAAI Press, 2011, pp. 3–10.

[5] E. Barbu, "Property type distribution in wordnet, corpora and wikipedia," Expert Systems with Applications, vol. 42, no. 7, 2015, pp. 3501 – 3507.

[6] W. Wu, H. Li, H. Wang, and K. Zhu, "Probase: A probabilistic taxonomy for text understanding," in ACM International Conference on Management of Data (SIGMOD), May 2012.

[7] T. Fountain and M. Lapata, "Taxonomy induction using hierarchical random graphs," in Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics, 2012, pp. 466–476.

[8] P. Cimiano, A. Hotho, and S. Staab, "Learning concept hierarchies from text corpora using formal concept analysis." J. Artif. Intell. Res.(JAIR), vol. 24, 2005, pp. 305–339.

[9] P. Szwed, "Concepts extraction from unstructured Polish texts: A rule based approach," in Computer Science and Information Systems (FedCSIS), 2015 Federated Conference on, Sept 2015, pp. 355–364.

[10] M. A. Hearst, "Automatic acquisition of hyponyms from large text corpora," in Proceedings of the 14th Conference on Computational Linguistics - Volume 2, ser. COLING '92. Stroudsburg, PA, USA: Association for Computational Linguistics, 1992, pp. 539–545.

[9] 60-80% precision seems to be achieved by other researchers too, see e.g. [31] Figure 4 or [32] table 5.

[11] Z. Kozareva, "Simple, Fast and Accurate Taxonomy Learning," in Text Mining. Springer International Publishing, 2014, pp. 41–62.

[12] J. Dean and S. Ghemawat, "Mapreduce: simplified data processing on large clusters," Commun. ACM, vol. 51, no. 1, Jan. 2008, pp. 107–113. [Online]. Available: http://doi.acm.org/10.1145/1327452.1327492

[13] F. Xu, D. Kurz, J. Piskorski, and S. Schmeier, "Term extraction and mining of term relations from unrestricted texts in the financial domain," in Proceedings of the 5th International Conference on Business Information Systems, Poznan, Poland, 2002.

[14] J. Nivre, J. Hall, J. Nilsson, A. Chanev, G. Eryigit, S. Kübler, S. Marinov, and E. Marsi, "Maltparser: A language-independent system for data-driven dependency parsing," Natural Language Engineering, vol. 13, no. 02, 2007, pp. 95–135.

[15] R. McDonald and F. Pereira, "Online learning of approximate dependency parsing algorithms," in In Proc. of EACL, 2006, pp. 81–88.

[16] R. McDonald and G. Satta, "On the complexity of non-projective data-driven dependency parsing," in Proceedings of the 10th International Conference on Parsing Technologies, ser. IWPT '07. Stroudsburg, PA, USA: Association for Computational Linguistics, 2007, pp. 121–132.

[17] M. Kuhlmann and J. Nivre, "Transition-based techniques for non-projective dependency parsing," Northern European Journal of Language Technology, vol. 2, no. 1, 2010, pp. 1–19.

[18] A. Nagórko, Zarys gramatyki polskiej. Warszawa: Wydawnictwo Naukowe PWN, 2007.

[19] R. Snow, D. Jurafsky, and A. Y. Ng, "Learning syntactic patterns for automatic hypernym discovery," in Advances in Neural Information Processing Systems (NIPS 2004), November 2004.

[20] A. Przepiórkowski, M. Bańko, R. L. Górski, and B. Lewandowska-Tomaszczyk, Eds., Narodowy Korpus Języka Polskiego. Warszawa: Wydawnictwo Naukowe PWN, 2012.

[21] M. Woliński, M. Miłkowski, M. Ogrodniczuk, and A. Przepiórkowski, "Polimorf: a (not so) new open morphological dictionary for Polish," in Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12), N. Calzolari (Conference Chair), K. Choukri, T. Declerck, M. U. Doğan, B. Maegaard, J. Mariani, A. Moreno, J. Odijk, and S. Piperidis, Eds. Istanbul, Turkey: European Language Resources Association (ELRA), may 2012.

[22] A. Przepiórkowski, The IPI PAN Corpus: Preliminary version. Warsaw: Institute of Computer Science, Polish Academy of Sciences, 2004.

[23] J. Nivre, J. Hall, J. Nilsson, A. Chanev, G. Eryigit, S. Kübler, S. Marinov, and E. Marsi, "Maltparser: A language-independent system for data-driven dependency parsing," Natural Language Engineering, vol. 13, 6 2007, pp. 95–135.

[24] A. Wróblewska, "Polish Dependency Bank," Linguistic Issues in Language Technology, vol. 7, no. 2, 2012.

[25] S. Buchholz and E. Marsi, "Conll-x shared task on multilingual dependency parsing," in Proceedings of the Tenth Conference on Computational Natural Language Learning, ser. CoNLL-X '06. Stroudsburg, PA, USA: Association for Computational Linguistics, 2006, pp. 149–164.

[26] Z. Saloni and M. Świdziński, Składnia współczesnego języka polskiego. Warszawa: Wydawnictwo Naukowe PWN, 2011.

[27] J. D. Lafferty, A. McCallum, and F. C. N. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," in Proceedings of the Eighteenth International Conference on Machine Learning, ser. ICML '01. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2001, pp. 282–289.

[28] C. D. Manning and H. Schütze, Foundations of Statistical Natural Language Processing. Cambridge, MA, USA: MIT Press, 1999.

[29] A. Clark, C. Fox, and S. Lappin, The Handbook of Computational Linguistics and Natural Language Processing. Wiley-Blackwell, 2010.

[30] M. Ogrodniczuk, A. Wójcicka, K. Głowińska, and M. Kopeć, "Detection of nested mentions for coreference resolution in Polish," in Advances in Natural Language Processing: Proceedings of the 9th International Conference on NLP, PolTAL 2014, Warsaw, Poland, September 17–19, 2014, ser. Lecture Notes in Artificial Intelligence, A. Przepiórkowski and M. Ogrodniczuk, Eds. Heidelberg: Springer International Publishing, 2014, vol. 8686, pp. 270–277.

[31] P.-M. Ryu and K.-S. Choi, "Automatic acquisition of ranked is-a relation from unstructured text," 2007.

[32] D. Ravichandran, P. Pantel, and E. Hovy, "The Terascale Challenge," in Proceedings of KDD Workshop on Mining for and from the Semantic Web (MSW-04), 2004, pp. 1–11.

9th International Workshop on Computational Optimization

MANY real world problems arising in engineering, economics, medicine and other domains can be formulated as optimization tasks. These problems are frequently characterized by non-convex, non-differentiable, discontinuous, noisy or dynamic objective functions and constraints which ask for adequate computational methods.

The aim of this workshop is to stimulate the communication between researchers working on different fields of optimization and practitioners who need reliable and efficient computational optimization methods.

TOPICS

The list of topics includes, but is not limited to:

- unconstrained and constrained optimization
- combinatorial optimization
- continues optimization
- global optimization
- multiobjective optimization
- optimization in dynamic and/or noisy environments
- large scale optimization
- parallel and distributed approaches in optimization
- random search algorithms, simulated annealing, tabu search and other derivative free optimization methods
- nature inspired optimization methods (evolutionary algorithms, ant colony optimization, particle swarm optimization, immune artificial systems etc)
- hybrid optimization algorithms involving natural computing techniques and other global and local optimization methods
- computational biology and optimization
- distance geometry and applications
- optimization methods for learning processes and data mining
- application of optimization methods on real life and industrial problems
- computational optimization methods in statistics, econometrics, finance, physics, chemistry, biology, medicine, engineering etc.

EVENT CHAIRS

- **Fidanova, Stefka,** Bulgarian Academy of Sciences, Bulgaria
- **Mucherino, Antonio,** INRIA, France
- **Zaharie, Daniela,** West University of Timisoara, Romania

PROGRAM COMMITTEE

- **Bartl, David,** University of Ostrava, Czech Republic
- **Bonates, Tibérius,** Universidade Federal do Ceará, Brazil
- **Breaban, Mihaela,** "Alexandru Ioan Cuza" University, Iasi, Romania
- **Chira, Camelia,** Technical University of Cluj-Napoca, Romania
- **Fidanova, Stefka,** Bulgarian Academy of Science
- **Gonçalves, Douglas,** Universidade Federal de Santa Catarina, Brazil
- **Hosobe, Hiroshi,** Hosei University, Japan
- **Iiduka, Hideaki,** Kyushu Institute of Technology, Japan
- **Lavor, Carlile,** IMECC-UNICAMP, Brazil
- **Marinov, Pencho,** Bulgarian Academy of Science, Bulgaria
- **Muscalagiu, Ionel,** Politehnica University Timisoara, Romania
- **Ninin, Jordan,** ENSTA-Bretagne, France
- **Parsopoulos, Konstantinos,** University of Ioannina, Greece
- **Pintea, Camelia,** Tehnical University Cluj-Napoca, Romania
- **Roeva, Olympia,** Institute of Biophysics and Biomedical Engineering, Bulgaria
- **Siarry, Patrick,** Universite Paris XII Val de Marne, France
- **Stefanov, Stefan,** South-West University ""Neofit Rilski, Bulgaria
- **Stuetzle, Thomas,** Université Libre de Bruxelles (ULB), Belgium
- **Tamir, Tami,** The Interdisciplinary Center (IDC), Israel
- **Zilinskas, Antanas,** Vilnius University, Lithuania

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 543–546
DOI: 10.15439/2016F459
ACSIS, Vol. 8. ISSN 2300-5963

Computational Optimizations in wildland fires for Bulgarian test cases

Nina Dobrinkova
Institute of Information and Communication Technologies – Bulgarian Academy of Sciences, acad. Georgi Bonchev bl. 2, Bulgaria, Email: ninabox2002@gmail.com

***Abstract*—In this article we are going to present the optimizations that has been done through different types of modeling actions on wildland fires for Bulgarian test cases. We will present approaches where meteorological data along with terrain specific relief and vegetation coverage are modeled in a way to present credible scenarios for wildland propagation used for calibration purposes of the different approaches. This work aims to prove that the used modeling tools can be used also in real time decision support for the responsible authorities when it comes to wildland fire propagation and the measures corresponding to limitation of its devastating consequences for the nature and human lives.**

I. Introduction

THE work presented in this paper is a year's long efforts which has been started because of an accident, that happened in Pirin Mountain near by the city of Razlok. In the year 2003 a helicopter with water tank flew very low trying to depress the rapidly burning wildland fire in the mountain. Unfortunately the engine oxygen has been vacuumed because of the flames, which caused helicopter's crash with four people crew dead [1]. This accident was very problematic for the Bulgarian society. That is why in the Bulgarian scientific community has been launched in the beginning of 2007 a pilot PhD program dedicated to the wildland propagation and its modeling opportunities as first attempts for computer based simulations on wildland propagations in Bulgaria.

In 2007 small team from Bulgarian Academy of Sciences (BAS) started adaptation of a US model, which was running in parallel mode. The model was called WRF-Fire (in 2010 renamed SFIRE). The input data for the model was needed to be first collected for specific test area and second preprocessed for model calibration.

The area of interest for the BAS team was first nearby Sofia, where idealized case has been run and second for real case calibration – test area near by the village of Leshnikovo, region of Harmanli has been chosen.

In this paper we will show the basis of the mathematical calculations and optimizations outlined from the research efforts and the achieved results.

This work has been supported by the Bulgarian Academy of Sciences Program for support of young researchers No: ДФНП-95-А1 and by the National Science Fund of the Bulgarian Ministry of Education, Youth and Science under Grant FNI I02/20.

II. WRF-Fire (SFIRE) mathematical basis

The mathematical background of the WRF-Fire model (SFIRE) is as position in the (x, y) plane. The model is semi-empirical and it represents the spread of the fire in direction of the fire line. This is the so called Rothermel modified formula. The burning region is represented as Ω for time t, which is represented with the point coordinates (x, y). The formula itself is:

$$\widetilde{S} = \min\{B_0, R_0 + \phi_w + \phi_s\}, \quad (1)$$

where B_0 is the fire spread against the wind direction, R_0 is the fire spread in absence of wind, $\phi_w = a\,(\vec{v}\cdot\vec{n})^b$ is the wind correction and $\phi_s = d\nabla z\cdot\vec{n}$ is the terrain correction, $\vec{v}$ is wind, ∇z is terrain variable along the normal $\vec{n}$ of the fire line, *a*, *b* и *d* are constants. In this case WRF-Fire use:

$$S = \begin{cases} 0, \text{ако } \tilde{S} < 0 \\ S_{max}, \text{ако } \tilde{S} > S_{max} \\ \tilde{S}, \text{ако } 0 \le \tilde{S} \le S_{max} \end{cases}, \quad (2)$$

where S_{max} is max fire spread. After the burning materials are burnt the model decrease them in the points (x, y) exponentially and that is represented with the formula:

$$F(x, y, t) = F_0(x, y)e^{-(t-t_i(x,y))/W(x,y)}, \quad (3)$$

where *t* is the time, t_i is the time for the burning, F_0 is the initial quantity of the burning materials (before they started to burn) and *W(x,y)*does not depend on the time, but from the burning materials. The heat transfer released by the fire, is represented in the atmosphere model as layer above the surface, which is situated in height [3]. The burning material quantity is represented by:

$$\Phi = -A(x, y)\frac{\partial}{\partial t}F(x, y, t). \quad (4)$$

This representation is needed because the atmosphere model WRF, does not support border values for heat transfer. The coefficients B_0, R_0, S_{max}, a, b, d, W and A, which describe the burning materials are measured in laboratory with experiments. For every surface point in the plane the coefficients of the burning materials are represented using the 13 Anderson categories [4]. These categories are developed for US originally and they have been defined by usage of the different sea levels on the surface. WRF-Fire has internally representation of every category and all additional characteristics, which gives opportunity for modifications when the fire is outside US.

WRF-Fire use also level-set functions for the spread of the fire [5]. This approach set as function $\psi = \psi(x, y, t)$, which define for Ω subregions using the rule:

$$\Omega(t) = \{(x, y) \in \Omega : \psi(x, y, t) < 0\}. \quad (5)$$

These subregions are burned and the fire line is defined as curve:

$$\Gamma(t) = \{(x, y) \in \Omega : \psi(x, y, t) = 0\}. \quad (6)$$

The function $\psi(x, y, t)$ satisfy the equation:

$$\frac{\partial \psi}{\partial t} + S(x, y) |\nabla \psi| = 0, \quad (7)$$

which can be solved numerically.

Formulas (1) - (7) are general description how mathematically the fire spread is represented inside the WRF-Fire (SFIRE) model. In the beginning the atmosphere model is interpolating the wind in order to get into the bigger domain of the atmosphere the fire changes. Afterwards is applied numerical method for the level-set function. The next step is to apply quadratic formulas for evaluation of the burnt material. In parallel it is evaluated also the released heat transfer into the atmosphere layers. The last step gives atmospheric change and that trigger the repetition of the model starts again.

III. Experimental results with WRF-Fire (SFire)

The experimental results which were obtained after evaluation of the WRF-Fire (SFIRE) model will be presented in this section as a brief summary where we will try to make as much as possible the use of the achieved results.

The first runs with the model were on ideal cases in order to see how the model correspond with the meteorological data and terrain data for the selected zone in south Bulgaria. The run used as inputs coordinates and information for village Leshnikovo, were the domain was set of size 4 by 4 km, with horizontal resolution of 50 m, for the atmosphere mesh, we used 80 by 80 grid cells and with 41 vertical levels from ground surface up to 100hPa. We didn't use nesting to keep the ideal case as basic as possible in order to evaluate the model capacity.

The domain, which we set was located 4 km west from village Zheleznitsa in the south-east part of Sofia district. The domain was covering the lower part of the forest part of Vitosha mountain.

The ignition line which we used was set in the center of the domain and to ignite it we set 345 m long line. The model does not consider ignition from point, because the atmospheric model does not cover such measurements. The ignition in parallel has been set to start 2 seconds after the simulation has begun. The results from this first simulation gave us idea how the model can be initialized and what the input data will be if we start simulation with real case forest fire for calibration of the model.

That is why we selected from the national data base in the ministry of forests, food and agriculture fire which has been burning in the period 14-17 August 2009.

For the initialization of the model with real case we had to use algorithm for implementation of the real data in a way WRF-Fire (SFIRE) to recognize it. We set two domains the first was covering area of 48 km^2 with resolution 300m (160x160). This domain was producing boundary and initial meteorological conditions for the inner domain and in this domain there were no fire simulations.

The inner domain was located in the middle of the coarse domain. The resolution in Domain 2 we set as 60m and the area covered is 9.6 km^2 (161x161). Domain 2 was centred on the fire ignition line and it was covering the areas of villages Ivanovo, Leshnikovo and Cherna Mogila. This area was located in South-East Bulgaria close to the Bulgarian-Greece border.

Following the description in [6] we get the intermediate fails for topography and fuel. The only difference is in the geogrid program, where the output fail has 2 extra varaibles – NFUEL_CAT and ZSF. NFUEL_CAT is the variable containing the data for the 13th categories of fuel available to bur and ZSF is the variable containing data for the detailed topography. The result as burnt simulated area compared to the real burnt area can be seen on the figures 1 and 2.

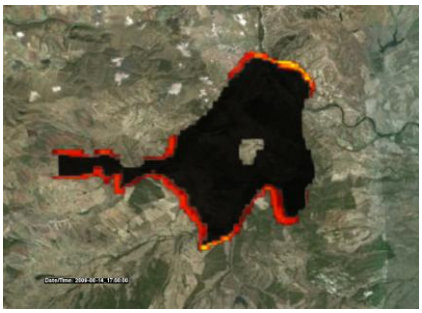


Fig. 1 The simulation fire burnt area

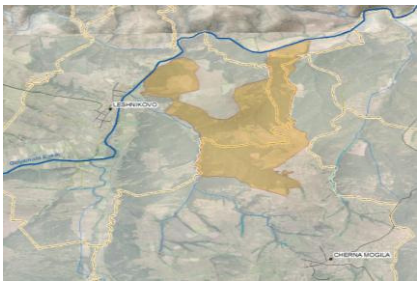


Fig. 2 The real fire burnt area

The simulation showed on figures 1 and 2 was done on the supercomputer at the University of Denver by distant connection. In Table 1 the simulation results are presented according to the number of the cores used.

TABLE 1: THE TIME REQUIRED FOR THE SIMULATION PRESENTED IN SECONDS DEPENDING ON THE NUMBER OF PROCESSORS RUNNING THE PARALLEL EXECUTION OF PROCESSES SHOWING THAT IN 120 CORES THE SIMULATIONS RUN AS REAL TIME. EVERYTHING ABOVE IS FASTER THAN REAL FIRE PROPAGATION.

Cores	6	12	24	36	60	120	240	360	480	720	960	1200
Fire line propagation in km.	1.91	1.08	0.50	0.34	0.22	0.13	0.08	0.06	0.06	0.04	0.10	0.04
Region 1	6.76	7.05	2.90	2.06	1.20	0.73	0.45	0.32	0.26	0.23	0.24	0.17
Region 2	0.00	0.00	0.00	0.02	0.02	0.04	0.04	0.06	0.06	0.08	0.07	0.15
Total sec. which is the coeff. for real time	10.59	9.21	3.91	2.75	1.64	0.99	0.61	0.44	0.37	0.31	0.44	0.26

With this simulations for the test site nearby Harmanli town has been elaborated a methodology for collection, procession and implementation of real data for test sites on Bulgarian territory. The selected model was having as input meteorological data, DEM and only 13 burning classes which led to the idea that we can experiment also with different models like BEHAVE Plus and FARSITE for our next tests.

IV. EXPERIMENTAL IMPLEMENTATION OF BEHAVE PLUS AND FARSITE SIMULATIONS IN THE TEST CASES OF ZLATOGRAD, MADAN AND NEDELINO MUNICIPAL AREAS IN BULGARIA

In the framework of bilateral cooperation program between Greece and Bulgaria 2007-2013 our team was having the opportunity to work in the Zlatograd forestry department located on the territories of Zlatograd, Madan and Nedelino municipal areas in Bulgaria.

The data we were working on was about fifteen wildfires that occurred in 2011 to 2012 within the Zlatograd municipal territory were provided by the Zlatograd forestry department. Based on initial BehavePlus results using standard fuel models, custom fuel models were developed for some vegetation types not well represented by the US fuel models.

Following evaluation of fuel models with BehavePlus, we performed analyses in FARSITE, a spatial fire growth system that integrates fire spread models with a suite of spatial data and tabular weather, wind and fuel moisture data to project fire growth and behavior across a landscape. We defined our test landscapes using a 500 m buffer zone around each of the fifteen Zlatograd fires.

Input for FARSITE consists of spatial topographic, vegetation, and fuels parameters compiled into a multi-layered “landscape file” format. Topographic data required to run FARSITE include elevation, slope, and aspect. Using the aforementioned 30 m DEM, we calculated an aspect layer, and then clipped elevation, aspect, and slope rasters to the extent of our fifteen test landscapes. Required vegetation data include fuel model and canopy cover. Fuel models within the 500 m buffered analysis area for each individual fire were assigned based on our BehavePlus analyses; fuel model assignments were tied to the dominant vegetation for each polygon based on the Zlatograd forestry department’s vegetation data. Canopy cover values were visually estimated from orthophoto images and verified with stand data from the Zlatograd forestry department. Additional canopy variables (canopy base height, canopy bulk density, and canopy height) that may be included in the landscape file were omitted, as these variables are most important for calculating crown fire spread or the potential for a surface fire to transition to a crown fire. None of the fifteen fires analyzed experienced crown fire.

Tabular weather and wind files for FARSITE were compiled using the weather and wind data from TV Met, Bulgarian meteorological company that included hourly records. Tabular fuel moisture files were created using the fine dead fuel moisture values calculated for the BehavePlus analyses for 1-hr timelag fuels. The 10-hr fuel moisture value was estimated by adding 1% to the 1-hr fuel moisture and the 100-hr fuel moisture was generally calculated by adding 3% to the 1-hr fuel moisture. The live fuel moisture values previously estimated for BehavePlus analyses were used to populate live herbaceous and live woody moisture values.

All simulations performed in FARSITE used metric data for inputs and outputs. An adjustment value was not used to alter rate of spread for standard fuel models, rather custom fuel models were created. Crown fire, embers from torching trees, and growth from spot fires were not enabled.

As an example of one of our successful FARSITE runs, we present the results from a single wildfire that burned in grassland vegetation, for which we developed custom fuel models. This fire occurred on

August 30, 2011, starting at 1400 and ending around 1800, and burned a total area of 0.3 ha. We used the following input parameters to model this small grassland fire in FARSITE:

Fuel moisture values: 6% (1-hr), 7% (10-hr), 9% (100-hr), 45% (live herbaceous), and 75% (live woody);

Daily maximum temperatures: 17-21°C;

Daily minimum relative humidity: 24-50%;

Winds: generally from the west-southwest at 1-2 k h-1

The fire size as calculated using FARSITE was 0.5 ha, which seems reasonable considering the modeled size would not have included the suppression actions that most likely occurred given the close proximity of a village to this fire figure 3.

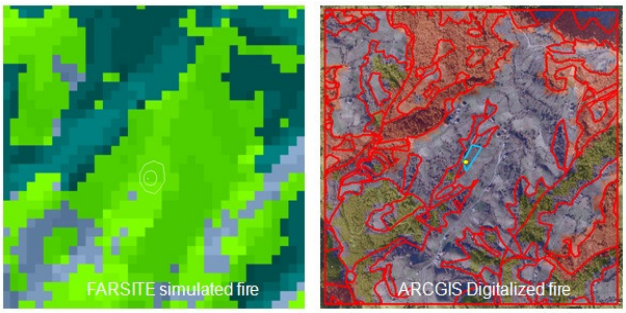


Figure 3: FARSITE run for a grassland fire, where size of the fire is very close to the real one, but the shape is different, because of wind information discrepancies

From this modeled fire we were able to estimate that grasslands and any grass and shrub covered areas will probably have the need to be further modeled before using as inputs for standard simulations with FARSITE or any similar tool in future. However FARSITE and BehavePlus provided reasonable outputs for future work in the field of fire behavior fuel modeling on the Bulgarian territory. The work done in more details is described in [7],[8],[9],[10].

III. Conclusion

The presented paper was having as main aim to provide a broader view on the tested modeling options for the Bulgarian wild land fires and the achieved results. There are still a lot of issues to be solved in the data preparation phase and the accuracy of the meteorological inputs. However the achieved results after all computations give promising options for future implementation of this modeling tools for more operational use in the responsible authorities first response facilities. With the presented test is given an idea what kind of simulations are more accurate when past wildland fires are modeled. Also the burning materials representations is very crucial which can be seen through the presented examples. This field still need to be more elaborated, but the first results gave good basis on how we can continue our future work.

Acknowledgment

N. D. thanks to the program for career development of young scientists, BAS with contract No: ДФНП-95-А1 and to the National Science Fund of the Bulgarian Ministry of Education, Youth and Science under Grant FNI I02/20.

References

[1] http://news.ibox.bg/news/id_1888536796 (In Bulgarian)

[2] Rothermel, R. C. (1972) A mathematical model for predicting fire spread in wildland fuels. Research Paper INT-115. Ogden, UT: US Department of Agriculture, Forest Service, Intermountain Forest and Range Experiment Station, pp. 1-40.

[3] Patton, E.G., Coen, J.L.: WRF-Fire: A coupled atmosphere-fire module for WRF. In: Preprints of Joint MM5/Weather Research and Forecasting Model UsersWorkshop, Boulder, CO, June 2225. NCAR (2004) 221223 http://www.mmm.ucar.edu/mm5/workshop/ws04/Session9/PattonEdward.pdf.

[4] Anderson, H.E.: Aids to determining fuel models for estimating fire behavior.USDA Forest Service, Intermountain Forest and Range Experiment Station, Research Report INT-122 (1982)http : //www.fs.fed.us/rm/pubsint/intgtr122.html.

[5] http://ccm.ucdenver.edu/wiki/Jan_Mandel/Blog/2010_Dec_2011_Jan

[6] http://www.openwfm.org/wiki/How_to_run_WRF-Fire_with_real_data#Downloading_high_resolution_elevation_data

[7] Dobrinkov G., Dobrinkova N., "Input Data Preparation for Fire Behavior Fuel Modeling of Bulgarian Test Cases (Main Focus on Zlatograd Test Case).", 10th International Conference on "Large-Scale Scientific Computations" LSSC'15, Sozopol 8-12 June 2015, Lecture Notes in Computer Science 9374, ISSN 0302-9743, ISSN 1611-3349 (electronic), ISBN: 978-3-319-26519-3, DOI 10.1007/978-3-319-26520-9, Springer Germany, pp. 335–342, 2015.

[8] Dobrinkova N., Hollingsworth L., Heinsch F.A., Dillon G., Dobrinkov G., "Bulgarian fuel models developed for implementation in FARSITE simulations for test cases in Zlatograd area". (E-proceeding: http://www.treesearch.fs.fed.us/pubs/46778) Wade DD & Fox RL (Eds), Robinson ML (Comp) (2014) 'Proceedings of 4th Fire Behavior and Fuels Conference', 18-22 February 2013, Raleigh, NC and 1-4 July 2013, St. Petersburg, Russia. (International Association of Wildland Fire: Missoula, MT), p.513 - p.521.

[9] Dobrinkov G., Dobrinkova N., "Wildfire behavior modeling data preparation for FARSITE simulations in Bulgarian test cases", 5th International Conference on Cartography & GIS & Seminar with EU cooperation on Early Warning and Dissaster/Crisis Management 15-21 June 2014, Proceedings Vol.2, ISSN:1314-0604, 2014, Riviera, Bulgaria, p.763- p. 770.

[10] N.Dobrinkova, G. Dobrinkov, "Farsite and WRF-Fire models, Pros and Cons For Bulgarian Cases", 9th International Conference on "Large-Scale Scientific Computations" LSSC'13, Sozopol 3-7 June 2013, Lecture Notes in Computer Science 8353, ISBN: 978-3-662-43879-4, Springer Germany, pp. 382–389, 2014.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 547–550
DOI: 10.15439/2016F137
ACSIS, Vol. 8. ISSN 2300-5963

InterCriteria Analysis of ACO Start Startegies

Stefka Fidanova
IICT BAS
Sofia, Bulgaria
E-mail: stefka@parallel.bas.bg

Olympia Roeva
IBBE BAS
Sofia, Bulgaria
E-mail: olympia@biomed.bas.bg

Pawel Gepner
Intel Corporation
Swindon, UK
E-mail: pawel.gepner@intel.com

Marcin Paprzycki
SRI PAS
Warsaw, Poland
E-mail: marcin.paprzycki@ibspan.waw.pl

***Abstract*—In combinatorial optimization, the goal is to find the optimal object from a finite set. Since such problems are hard to be solved, usually some metaheuristics is applied. One of the most successful techniques for a number of classes of problems is Ant Colony Optimization (ACO). Some start strategies can be applied, to the ACO algorithms, to improve their performance. Here, the InterCriteria Analysis (ICrA) is applied to the ACO algorithm. On the basis of the ICrA, we examine and analyse the ACO performance according to the different start strategies.**

I. Introduction

THE IDEA for the ACO arises fro the way that real ants look for food. The ACO algorithm was proposed, by Marco Dorigo, more than 20 years [6]. Later, several variants and supplements were added, to improve its performance [6]. In [7], various ACO start strategies, which lead to finding better solutions, were proposed.

The InterCriteria Analysis (ICrA) is aiming at going beyond the nature of the criteria involved in a process of evaluation of multiple objects against multiple criteria, and, thus to discover some dependencies between the ICrA criteria themselves [2]. Initially, the ICrA has been applied in temporal, threshold and trends analyses of an economic case-study of EU member states' competitiveness [4]. Further, the ICrA has been used to discover dependencies between parameters in mathematical models and performance criteria for metaheuristics such as GAs and ACO [1], [9]. Here, the ICrA is applied for analysis of an ACO algorithm, with various start strategies. The Multiple Knapsack Problem (MKP) is used as a test problem. The goal is to analyze the dependence between start strategies and algorithm performance, and correlations between strategies.

II. ACO Algorithm with Start Strategies

Let us consider the ACO algorithm applied to the solution of a problem represented by a graph. Here, the feasible solutions are represented by path in that graph. The solution process "compares the length" of available paths. In every iteration, an ant starts from a random node and creates a solution. If the last selected node is u, the ant selects the next node (v), to be included in the path, by applying probabilistic rule called transition probability.

$$p_{uv} = \tau_{uv}^{\alpha}\eta_{uv}^{\beta} / \left[\sum_{(u,w)\in E_S\,:\,w\not\subset X} \left(\tau_{uw}^{\alpha}\eta_{uw}^{\beta}\right)\right], \qquad (1)$$

Here, α and β are transition probability parameters, $\tau \in (0,1)$ is a numerical information called pheromone, and η is an heuristic information related to the problem. At the beginning, the value of the pheromone across elements is the same. In each iteration, we update the pheromone, on selected elements of the graph, according to the value of the objective function (elements belonging to the better solutions receive more pheromone than others). The random start is very important for good performance of the ACO algorithm, but for some classes of problems (e.g. subset problems), selection of the starting node can be significant. For better managing the search process we include semi-random start of the ants. Here, the nodes are divided into several subsets. An estimation of how good and how bad is to start from some subset is introduced according the number of good and bad solutions that started from this subset. Assume that D_j is the estimation of how good it is to start from subset j, and E_j is the estimation how bad it is to start from subset j [7].

$$D_j(i) = \phi.D_j(i-1) + (1-\phi).F_j(i), \qquad (2)$$

$$E_j(i) = \phi.E_j(i-1) + (1-\phi).G_j(i), \qquad (3)$$

where $i \geq 1$ is the current process iteration and, for each j $(1 \leq j \leq N)$:

$$F_j(i) = \begin{cases} f_{j,A}/n_j & \text{if } n_j \neq 0 \\ F_j(i-1) & \text{otherwise} \end{cases}, \qquad (4)$$

$$G_j(i) = \begin{cases} g_{j,B}/n_j & \text{if } n_j \neq 0 \\ G_j(i-1) & \text{otherwise} \end{cases}, \qquad (5)$$

$f_{j,A}$ is the number of the solutions among the best $A\%$, $g_{j,B}$ is the number of the solutions among the worst $B\%$, where $A+B \leq 100$, $i \geq 2$ and $\sum_{j=1}^{N} n_j = n$, where n_j $(1 \leq j \leq N)$

is the number of solutions obtained by ants starting from nodes subset j, n is the number of ants. Here, initial values of the weight coefficients are: $D_j(1) = 1$ and $E_j(1) = 0$. Parameter ϕ, $0 \leq \phi \leq 1$, shows the weight of the information from the previous iterations and from the last iteration. Several start strategies and combinations between them are proposed. If thresholds for good estimation D and for bad estimation E are fixed, the proposed start strategies are as follows [7]:

1) If $E_j(i)/D_j(i) > E$ then, for current iteration, the subset j is forbidden. The starting node is randomly chosen from $\{j \mid j \text{ is not forbidden}\}$;
2) If $E_j(i)/D_j(i) > E$ then, for current simulation, the subset j is forbidden. The starting node is randomly chosen from $\{j \mid j \text{ is not forbidden}\}$;
3) If $E_j(i)/D_j(i) > E$ then, for K_1 consecutive iterations, the subset j is forbidden. The starting node is randomly chosen from $\{j \mid j \text{ is not forbidden}\}$;
4) Let $r_1 \in [\frac{1}{2}, 1)$ and $r_2 \in [0, 1]$ to be random numbers. If $r_2 > r_1$, a node from subset $\{j \mid D_j(i) > D\}$ is randomly chosen, otherwise a node from the not forbidden subsets is randomly chosen. r_1 is chosen and fixed at the beginning.
5) Let $r_1 \in [\frac{1}{2}, 1)$ and $r_2 \in [0, 1]$ to be random numbers. If $r_2 > r_1$, a node from subset $\{j \mid D_j(i) > D\}$ is randomly chosen, otherwise a node from the not forbidden subsets is randomly chosen. r_1 is chosen at the beginning and increase with r_3 every iteration, $r_3 \in (0, 1)$ is a parameter.

Here $K_1, K_1 \in [0, \textit{number of iterations}]$ is a parameter.

We can apply one of start strategies, or combine some of them. Strategies 1, 2, and 3 can be combined with strategies 4 and 5. When an ant chooses a start node, first applied strategy is one of 1, 2, or 3, after that, strategy 5 or 6 is used. Thus, together with a completely random start, there are 12 strategies that can be applied (see, also [7]).

III. Multiple Knapsack Problem

The start node selection is very important for the subset problems, because only some nodes of the graph of the problem belong to the solution. The Multiple Knapsack Problem (MKP) is a representative of the class of subset problems. It also arises as a sub-problem in a group of more complex problems. Some of important applications that can be formulated as a MKP are cargo loading problems, cutting stock, bin-packing, budget control and financial management. The MKP is also used in a fault tolerance problem [11]. Authors of [5] designed a public cryptography scheme whose security is based on the difficulty of solving the MKP. In [8] two-processor scheduling problems are proposed to be solved as a MKP. Other applications include industrial management, naval, aerospace, computational complexity theory, etc. The MKP can be formulated as follows:

$$\begin{array}{c} \max \sum_{j=1}^{n} p_j x_j \\ \text{subject to} \sum_{j=1}^{n} r_{ij} x_j \leq c_i \quad i = 1, \ldots, m \\ x_j \in \{0, 1\} \quad j = 1, \ldots, n \end{array} \tag{6}$$

$$x_j = \begin{cases} 1 \text{ iff the object } j \text{ is chosen,} \\ 0 \text{ otherwise.} \end{cases}$$

where m are the resources (the knapsacks), n are the objects, p_j is a profit of every object j, c_j (knapsack capacity) is the resource budget, r_{ij} is the consumption of resource i by object j. The goal is maximizing the sum of profits for a limited budget.

There are m constraints in this problem, so the MKP is also called m-dimensional knapsack problem. Let

$$I = \{1, \ldots, m\}, J = \{1, \ldots, n\},$$

with $c_i \geq 0$ for all $i \in I$. A well-stated MKP assumes that $p_j > 0,\ r_{ij} \leq c_i \leq \sum_{j=1}^{n} r_{ij}$ for all $i \in I,\ j \in J$. Note that the $[r_{ij}]_{m \times n}$ matrix and the $[c_i]_m$ vector are both non-negative.

IV. InterCriteria Analysis

Let us be given an Index Matrix (IM, [3]) whose index sets for rows consist of the names of the criteria and for columns of objects. We will obtain an IM with index sets consisting of names of the criteria both for rows and for columns. Elements of this IM correspond to the degrees of "agreement" and degrees of "disagreement" of the considered criteria. The following two points are assumed [10]. (1) All criteria provide an evaluation for all objects and all these evaluations are available. (2) All the evaluations of a given criteria can be compared among themselves. Further, by O, we denote the set of all objects $O_1, O_2, \ldots, O_n$ being evaluated, and by $C(O)$ the set of values assigned by a given criteria C to the objects. Let $x_i = C(O_i)$. Then the following set can be defined:

$$C^*(O) \stackrel{\text{def}}{=} \{\langle x_i, x_j \rangle | i \neq j \,\&\, \langle x_i, x_j \rangle \in C(O) \times C(O)\}.$$

In order to find the degrees of "agreement" of two criteria, the vector of all internal comparisons of each criteria is constructed. This vector fulfills exactly one of the following three relations: R, $\overline{R}$ and $\tilde{R}$. For a fixed criterion C, and any ordered pair $\langle x, y \rangle \in C^*(O)$ it is required that: $\langle x, y \rangle \in R \Leftrightarrow \langle y, x \rangle \in \overline{R}$, $\langle x, y \rangle \in \tilde{R} \Leftrightarrow \langle x, y \rangle \notin (R \cup \overline{R})$, $R \cup \overline{R} \cup \tilde{R} = C^*(O)$

For a criterion C, let us define a preference matrix between objects $1, 2, \ldots, n$ so that C_{ij} is 1 if i is better than j, -1 if i is worse than j, and 0 if i and j are equivalent or incomparable over criterion C. We determine the degree of "agreement" ($\mu_{C,C'}$) between the two criteria as the proportion of matching components. This can be done in several ways, e.g. by counting the matches or by taking the complement of the Hamming distance. The degree of "disagreement" ($\nu_{C,C'}$) is the proportion of components of opposing signs in the two vectors. The difference $\pi_{C,C'} = 1 - \mu_{C,C'} - \nu_{C,C'}$ is considered as a degree of "uncertainty".

V. Numerical Results

We combined the ICrA with the ACO for different start strategies applied to the MKP [7]. Ten test problems from the "OR-Library" with 100 objects and 10 constraints (available from

http://people.brunel.ac.uk/~mastjjb/jeb/orlib) were used. The ACO algorithm is applied to various number of nodes in node subsets. The node subsets consist of the same number of nodes, namely 1, 2, 4, 5 and 10. The average results over the 10 test problems and 30 runs of every problem with every strategy was obtained. The ranking is from 10 to 100. The ICrA objects $(O_1, O_2, \ldots, O_{20})$ are the different conditions, namely nodes 1, 2, 4, 5 and 10, in four cases of ϕ, $\phi = [0\ 0.25\ 0.5\ 0.75]$. The ICrA criteria $(C_1, C_2, \ldots, C_{12})$ are 12 different start strategies for the ACO. The ICrA resulted in two IM ,with the relations between considered 12 criteria. The resulting IMs, for $\mu_{C,C'}$ and $\nu_{C,C'}$, are shown in Table I and Table II.

TABLE I: Index matrix for $\mu_{C,C'}$

	C_1	C_2	C_3	C_4	C_5	C_6
C_1	1.000	0.042	0.016	0.079	1.000	1.000
C_2	0.042	1.000	0.642	0.742	0.042	0.042
C_3	0.016	0.642	1.000	0.670	0.016	0.016
C_4	0.079	0.742	0.670	1.000	0.079	0.079
C_5	1.000	0.042	0.016	0.079	1.000	1.000
C_6	1.000	0.042	0.016	0.079	1.000	1.000
C_7	0.037	0.826	0.653	0.758	0.037	0.037
C_8	0.037	0.826	0.653	0.758	0.037	0.037
C_9	0.026	0.637	0.889	0.705	0.026	0.026
C_{10}	0.026	0.621	0.884	0.700	0.026	0.026
C_{11}	0.026	0.737	0.695	0.821	0.026	0.026
C_{12}	0.026	0.737	0.695	0.821	0.026	0.026

	C_7	C_8	C_9	C_{10}	C_{11}	C_{12}
C_1	0.037	0.037	0.026	0.026	0.026	0.026
C_2	0.826	0.826	0.637	0.621	0.737	0.737
C_3	0.653	0.653	0.889	0.884	0.695	0.695
C_4	0.758	0.758	0.705	0.700	0.821	0.821
C_5	0.037	0.037	0.026	0.026	0.026	0.026
C_6	0.037	0.037	0.026	0.026	0.026	0.026
C_7	1.000	1.000	0.695	0.679	0.800	0.800
C_8	1.000	1.000	0.695	0.679	0.800	0.800
C_9	0.695	0.695	1.000	0.984	0.753	0.753
C_{10}	0.679	0.679	0.984	1.000	0.747	0.747
C_{11}	0.800	0.800	0.753	0.747	1.000	1.000
C_{12}	0.800	0.800	0.753	0.747	1.000	1.000

For better understanding of the results, the values of the $\mu_{C,C'}$, $\nu_{C,C'}$, $\pi_{C,C'}$ of the criteria pairs, are sorted by the value of the $\mu_{C,C'}$. The list is presented in Tables III and IV. Table III shows the criteria pair with high degrees of "agreement" ($\mu_{C,C'}$) and low value for the degree of "disagreement" ($\nu_{C,C'}$). Table IV shows the criteria pair with high degree of "uncertainty". Regarding Tables III and IV we observe that relations between criterion C_1 and criteria C_5 and C_6 have the highest value of $\mu_{C,C'}$ ($\mu_{C,C'} = 1$), i.e. these criteria are in strong positive consonance. Henceforth, the ACO algorithm performs in a similar way with random start and start strategies 4 and 5. In strategies 4 and 5 there are no forbidden regions (as in the random start). In these cases, only the probability to choose the next element in the solution is different. Other pairs that have the highest value of $\mu_{C,C'}$ ($\mu_{C,C'} = 1$) are $C_7 - C_8$ and $C_{11} - C_{12}$. These strategies (Strategies 1-4, 1-5, 3-4 and 3-5) show also very similar performance.

TABLE II: Index matrix for $\nu_{C,C'}$

	C_1	C_2	C_3	C_4	C_5	C_6
C_1	0.000	0.000	0.000	0.000	0.000	0.000
C_2	0.000	0.000	0.300	0.137	0.000	0.000
C_3	0.000	0.300	0.000	0.237	0.000	0.000
C_4	0.000	0.137	0.237	0.000	0.000	0.000
C_5	0.000	0.000	0.000	0.000	0.000	0.000
C_6	0.000	0.000	0.000	0.000	0.000	0.000
C_7	0.000	0.095	0.295	0.137	0.000	0.000
C_8	0.000	0.095	0.295	0.137	0.000	0.000
C_9	0.000	0.079	0.079	0.189	0.000	0.000
C_{10}	0.000	0.084	0.084	0.195	0.000	0.000
C_{11}	0.000	0.263	0.263	0.084	0.000	0.000
C_{12}	0.000	0.263	0.263	0.084	0.000	0.000

	C_7	C_8	C_9	C_{10}	C_{11}	C_{12}
C_1	0.000	0.000	0.000	0.000	0.000	0.000
C_2	0.095	0.095	0.295	0.311	0.195	0.195
C_3	0.295	0.295	0.079	0.079	0.263	0.263
C_4	0.137	0.137	0.189	0.195	0.084	0.084
C_5	0.000	0.000	0.000	0.000	0.000	0.000
C_6	0.000	0.000	0.000	0.000	0.000	0.000
C_7	0.000	0.000	0.242	0.258	0.137	0.137
C_8	0.000	0.000	0.242	0.258	0.137	0.137
C_9	0.242	0.242	0.000	0.016	0.205	0.205
C_{10}	0.258	0.258	0.016	0.000	0.211	0.211
C_{11}	0.137	0.137	0.205	0.211	0.000	0.000
C_{12}	0.137	0.137	0.205	0.211	0.000	0.000

The criteria pairs still in a consonance, are pairs of criteria C_2, C_3, C_4 and C_7, C_8, ..., C_{12}. They correspond to Strategies 1, 2 and 3, combined with 4 and 5. In all this strategies there are forbidden regions, therefore the ACO performs in a similar way when we apply "any of them". The criteria pairs with value of $\mu_{C,C'} = [0.75 - 0.25)$ are in dissonance, i.e. there are no dependencies between these criteria (they are independent). The ACO algorithm with random strategies and the ACO algorithm with strategies with forbidden regions perform in a very different way, thus we can not find relations between them.

VI. Conclusion

In this paper, an ICrA is used with the ACO algorithm, to establish the relations and dependencies between the ACO performance and the start strategies. Twelve start strategies are studied. Part of them disallow some regions of the search space for one or more iterations. We can conclude that criteria corresponding to the strategies without forbidden regions are in positive consonance, as well as the criteria corresponding to the strategies with forbidden regions. The criteria corresponding to the strategies with forbidden regions are in dissonance with criteria corresponding to the strategies without forbidden regions.

Acknowledgment

Work presented here is partially supported by the National Scientific Fund of Bulgaria under grants DFNI-I02/5 and DFNI I02/20, and by the Polish-Bulgarian collaborative grant "Parallel and Distributed Computing Practices".

TABLE III: Criteria pairs-I

Criteria pairs	$\mu_{C,C'}$	$\nu_{C,C'}$	$\pi_{C,C'}$
$C_1 - C_5$	1.000	0.000	0.000
$C_1 - C_6$	1.000	0.000	0.000
$C_5 - C_6$	1.000	0.000	0.000
$C_7 - C_8$	1.000	0.000	0.000
$C_{11} - C_{12}$	1.000	0.000	0.000
$C_9 - C_{10}$	0.984	0.016	0.000
$C_3 - C_9$	0.889	0.079	0.032
$C_3 - C_{10}$	0.889	0.079	0.032
$C_2 - C_7$	0.826	0.095	0.079
$C_2 - C_8$	0.826	0.095	0.079
$C_4 - C_{11}$	0.821	0.084	0.095
$C_4 - C_{12}$	0.821	0.084	0.095
$C_7 - C_{11}$	0.800	0.137	0.063
$C_7 - C_{12}$	0.800	0.137	0.063
$C_8 - C_{11}$	0.800	0.137	0.063
$C_8 - C_{12}$	0.800	0.137	0.063
$C_4 - C_7$	0.758	0.137	0.105
$C_4 - C_8$	0.758	0.137	0.105
$C_9 - C_{11}$	0.753	0.205	0.042
$C_9 - C_{12}$	0.753	0.205	0.042
$C_{10} - C_{11}$	0.747	0.211	0.042
$C_{10} - C_{12}$	0.747	0.211	0.042
$C_2 - C_4$	0.742	0.137	0.121
$C_2 - C_{11}$	0.737	0.195	0.068
$C_2 - C_{12}$	0.737	0.195	0.068
$C_4 - C_9$	0.705	0.189	0.105
$C_4 - C_{10}$	0.700	0.195	0.105
$C_3 - C_{11}$	0.695	0.263	0.042
$C_3 - C_{12}$	0.695	0.263	0.042
$C_7 - C_9$	0.695	0.242	0.063
$C_8 - C_9$	0.695	0.242	0.063
$C_8 - C_{10}$	0.679	0.258	0.063
$C_7 - C_{10}$	0.679	0.258	0.063
$C_3 - C_4$	0.670	0.237	0.084
$C_3 - C_7$	0.653	0.295	0.053
$C_3 - C_8$	0.653	0.295	0.053
$C_2 - C_3$	0.642	0.300	0.058
$C_2 - C_9$	0.637	0.295	0.068
$C_2 - C_{10}$	0.621	0.311	0.068

TABLE IV: Criteria pairs- II

Criteria pairs	$\mu_{C,C'}$	$\nu_{C,C'}$	$\pi_{C,C'}$
$C_1 - C_4$	0.079	0.000	0.921
$C_4 - C_5$	0.079	0.000	0.921
$C_4 - C_6$	0.079	0.000	0.921
$C_1 - C_2$	0.042	0.000	0.958
$C_2 - C_5$	0.042	0.000	0.958
$C_2 - C_6$	0.042	0.000	0.958
$C_1 - C_7$	0.037	0.000	0.963
$C_1 - C_8$	0.037	0.000	0.963
$C_5 - C_7$	0.037	0.000	0.963
$C_5 - C_8$	0.037	0.000	0.963
$C_6 - C_7$	0.037	0.000	0.963
$C_6 - C_8$	0.037	0.000	0.963
$C_1 - C_9$	0.026	0.000	0.974
$C_1 - C_{10}$	0.026	0.000	0.974
$C_1 - C_{11}$	0.026	0.000	0.974
$C_1 - C_{12}$	0.026	0.000	0.974
$C_5 - C_9$	0.026	0.000	0.974
$C_5 - C_{10}$	0.026	0.000	0.974
$C_5 - C_{11}$	0.026	0.000	0.974
$C_5 - C_{12}$	0.026	0.000	0.974
$C_6 - C_9$	0.026	0.000	0.974
$C_6 - C_{10}$	0.026	0.000	0.974
$C_6 - C_{11}$	0.026	0.000	0.974
$C_6 - C_{12}$	0.026	0.000	0.974
$C_1 - C_3$	0.016	0.000	0.984
$C_3 - C_5$	0.016	0.000	0.984
$C_3 - C_6$	0.016	0.000	0.984

REFERENCES

[1] M. Angelova, O. Roeva, T. Pencheva, InterCriteria Analysis of Crossover and Mutation Rates Relations in Simple Genetic Algorithm, Proceedings of the Federated Conference on Computer Science and Information Systems, Vol. 5, 419-424, 2015.

[2] K. Atanassov, D. Mavrov, V. Atanassova, *Intercriteria Decision Making: A New Approach for Multicriteria Decision Making*, Based on Index Matrices and Intuitionistic Fuzzy Sets. Issues in IFSs and GNs 11, 1–8 (2014)

[3] K. Atanassov,V. Atanassova, G. Gluhchev, InterCriteria Analysis: ideas and problems. Notes on Intuitionistic Fuzzy Sets 21(1), 81–88 (2015)

[4] V. Atanassova, L. Doukovska, D. Karastoyanov, F. Capkovic, *InterCriteria Decision Making Approach to EU Member States Competitiveness Analysis: Trend Analysis*, In: Angelov, P., et al. (eds.) Intelligent Systems'2014, Advances in Intelligent Systems and Computing, Vol. 322, 2014, 107–115.

[5] W. Diffe, M. E. Hellman, New direction in cryptography. IEEE Transactions of Information Theory. IT-36, 1976, 644-654.

[6] M. Dorigo, T. Stutzler, Ant Colony Optimization, MIT press, 2004.

[7] S. Fidanova, K. Atanassov, P. Marinov, Generalized Nets and Ant Colony Optimization, Academic Publishing House, Bulgarian Academy of Sciences, 2011.

[8] S. Martello, P. Toth, *A mixtures of dynamic programming and branch-and-bound for the subset-sum problem*, Management Science 30, 1984, 756-771.

[9] O. Roeva, S. Fidanova, M. Paprzycki, *InterCriteria Analysis of ACO and GA hybrid algorithms*, Studies in Computational Intelligence 610, 2016, 107–126.

[10] O. Roeva, S. Fidanova, P. Vassilev, P. Gepner, InterCriteria Analysis of a Model Parameters Identification using Genetic Algorithm, Proceedings of the Federated Conference on Computer Science and Information Systems, Vol. 5, 2015, 501-506.

[11] A. Sinha, A. A. Zoltner, *The multiple-choice knapsack problem*, Journal of Operational Research 27, 1979, 503-515.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 551–559
DOI: 10.15439/2016F19
ACSIS, Vol. 8. ISSN 2300-5963

Partitioning the Data Domain of Combinatorial Problems for Sequential Optimization

Christian Hinrichs, Jörg Bremer, Sönke Martens, Michael Sonnenschein
Department of Computing Science
University of Oldenburg
26129 Oldenburg, Germany
<first name>.<last name>@uni-oldenburg.de

***Abstract*—Following the long-term goal of substituting conventional power generation with cleaner energy will lead to an integration of a large share of small energy generation units imposing large problem sizes for coordination. The expected huge number of entities leads to a need for new techniques reducing the computational effort for coordination. Predictive scheduling is a frequent task in energy grid control. For a number of energy resources, schedules have to be found that fulfill several objectives at the same time. Considering day-ahead scenarios with 96-dimensional schedules imposes additional challenges to this already hard combinatorial problem. We explore the effects of reducing complexity by partitioning the data domain of the optimization problem for a sequential approach that integrates energy models for constraint handling directly into the optimization process. We explore the effects of different partitioning schemes and evaluate the trade-off between accuracy and effort with several simulation studies.**

I. Introduction

DESPITE being environmentally friendly and sustainable, the increasing amount of renewable electricity generation has a major drawback. In contrast to conventional power plants, the generation from e. g. solar and wind power can neither be predicted with high accuracy nor scheduled precisely. Furthermore, as storage of electrical energy is a rather difficult and expensive task, balancing supply and demand in the grid in real-time is one of the most important functions of power system control centers. Thus, to incorporate renewables accordingly, methods have to be established that can compensate for the missing flexibility of those energy sources. For instance, controlling flexible loads to use electrical power in times of high availability (i. e. high wind or solar radiation) can help using renewable power more efficiently [1].

From an algorithmic perspective, the task of scheduling energy units can be seen as combinatorial optimization problem: For each unit (i. e. controllable loads and generators), an optimal schedule has to be found such that for every time interval of a predefined planning horizon, a specific amount of electrical power (positive or negative) is assigned. A combination of schedules is optimal if the aggregated power equals a target profile that is given by the use case. For example, given the inverse of a predicted feed-in time series for wind and photovoltaic power plants as target profile, an optimal schedule assignment for the controllable energy units would lead to a perfect balancing of supply and demand in the considered system at each interval of the prediction horizon. Another use case is the operation of a virtual power plant (VPP): Given a target power profile that is to be offered in an energy market, the members of the VPP must collaborate in such a way that the VPP as a whole will produce the target profile. From the outside perspective, no difference between a VPP and a classical power plant would be evident [1].

However, the schedule optimization task becomes hard to solve in the presence of device-specific restrictions. Many flexible generators and loads are controllable in principle, but at the same time have to obey specific individual constraints. For instance, a cogeneration plant (e. g. a combined heat and power plant, CHP) produces thermal and electrical power simultaneously. As the generation of those two forms of power are strictly coupled within the unit and the use of the heat is subject to further restrictions such as the size of an attached thermal buffer storage, the electrical generation is severly confined as well [2], [3]. Due to such constraints, many established optimization algorithms cannot be applied to this task. For instance, meta-heuristics like evolutionary algorithms or simulated annealing are not able to cope with constraints per se and would have to be tailored specifically for the actual use case and the involved energy units.

In [4], a method has been introduced that is able to transform a problem with restrictions into a restriction-free representation using a machine learning approach. This so-called *support vector decoder model* allows generic optimization algorithms to operate in a restriction-free representation of the constrained search space of the original optimization problem. The method has been successfully applied to the schedule optimization problem [3]. In this context, the influence of the length of the planning horizon on solution quality became apparent: Usually, the method is applied to representations of the planning horizon as a whole by interpreting feasible schedules of energy units as elements to the combinatorial problem. However, the longer the planning horizon (and the schedules, consequently), the lower the solution quality of the employed optimization algorithms. At first glance, this may seem like an inherent restriction of the problem to solve. But interestingly, preliminary experiments indicated a potentially increasing solution quality when the optimization algorithm is applied in a successive manner to sequential partitions of the planning horizon. Thus, the objective of this paper is to explore the potential benefit of partitioning the search space

978-83-60810-90-3/$25.00

of the given combinatorial problem in the data domain in combination with sequential optimization of the individual data partitions.

In Section II, the motivating optimization problem as well as the support vector decoder model are briefly recapped from previous works. Following, Section III first revisits relevant related work in the field of high-dimensionality optimization strategies before describing the introduced concept of data partitions for the considered combinatorial problem in more detail. Section IV then evaluates the approach by employing a simulation study in the aforementioned application domain. Finally, Section V concludes the paper.

II. Methodical Background

We start with some preliminary definitions. First, let $\mathcal{U}$ be the set of DER units in the VPP and Z_U be the set of operational states of unit U. We regard the schedule of an energy unit as a vector $\boldsymbol{p} = (p_1, \ldots, p_d) \in \mathbb{R}^d$ of mean power p_i generated (or consumed) during the ith time interval. The starting time and the width of a time interval (today usually 15 minutes) are defined separately and have no effect on this representation. For the used support vector decoder it is advantageous to use schedules with scaled power values [5]. Scaling is done according to respective minimum (p_{min}) and maximum (p_{max}) nominal active power output (or input):

$$\begin{aligned} \rho : \mathbb{R}^d &\to \mathcal{X} \subset [0,1]^d \\ \boldsymbol{p} &\mapsto \boldsymbol{x} = \rho(\boldsymbol{p}), \text{with } x_i = \frac{p_i - p_{min}}{p_{max} - p_{min}}; \end{aligned} \tag{1}$$

For this paper we go with the example of predictive scheduling for active power planning in day-ahead scenarios (not necessarily 24 hours but for some given future period).

One of the crucial challenges in operating a VPP arises from the complexity of the scheduling task due to the large amount of (small) energy units in the distribution grid [6]. In the following, we consider predictive scheduling, where the goal is to select exactly one schedule $\boldsymbol{x}_i$ for each energy unit U_i from a search space of feasible schedules with respect to a future planning horizon, such that a global objective function (e. g. a target power profile for the VPP) is optimized by the sum of individual contributions [7]. A basic formulation of the scheduling problem is given by

$$\delta \left(\sum_{i=1}^{m} \boldsymbol{x}, \boldsymbol{\zeta} \right) \to min \tag{2}$$

such that

$$\boldsymbol{x}_i \in \mathcal{F}^{(U_i)} \ \forall U_i \in \mathcal{U}. \tag{3}$$

In equation (2) δ denotes an (in general) arbitrary distance measure for evaluating the difference between the aggregated schedule of the group and the desired target schedule $\boldsymbol{\zeta}$. W.l.o.g., in this contribution we use the Euclidean distance $\| \cdot \|_2$. To each energy unit U_i exactly one schedule $\boldsymbol{x}_i$ has to be assigned. The desired target schedule is given by $\boldsymbol{\zeta}$. $\mathcal{F}^{(U_i)}$ denotes the individual set of feasible schedules that are operable for unit U_i without violating any (technical) constraint. Solving this problem without unit independent constraint handling leads to specific implementations that are not suitable for handling changes in VPP composition or unit setup without having changes in the implementation of the scheduling algorithm [8].

In [9] a so called support vector decoder has been introduced. Basically, a decoder is a constraint handling technique that gives an algorithm hints on where to look for feasible solutions. It imposes a relationship between a decoder solution and a feasible solution and gives instructions on how to construct a feasible solution [10]. For example, [11] proposed a homomorphous mapping between an n-dimensional hyper cube and the feasible region in order to transform the problem into an topological equivalent one that is easier to handle. In order to be able to derive such a decoder mapping automatically from any given energy unit model, [9] developed an approach based on a support vector model [5]. We will briefly describe this method.

The basic idea is to start with a set $\mathcal{X} = \{\boldsymbol{x}_i\}_n$ of feasible example schedules derived from the simulation model of an energy unit and use this sample as a stencil for the region (the sub-space in the space of all schedules) that contains only feasible schedules. The set $\mathcal{X}$ can be easily generated after a sampling method from [12]. The schedule sample is then used as a training set for a support vector based machine learning approach [13] that derives a geometrical description of the sub-space that contains the given data (in our case: the feasible schedules). Given a set of data samples, the inherent structure of the scope of action of a unit where the data resides in can be derived as follows: After mapping the data to a high dimensional feature space by means of an appropriate kernel, the smallest enclosing ball in this feature space is determined. When mapping back this ball to data space, it forms a set of contours (not necessarily connected) enclosing the given data sample. An in-depth discussion can be found e. g. in [13].

At this point, the set of alternatively feasible schedules of a unit is represented as pre-image of a high-dimensional ball $\mathcal{S}$. Figure 1 shows the situation. This representation has some advantageous properties. Although the pre-image might be some arbitrary shaped non-continuous blob in $\mathbb{R}^d$, the high-dimensional representation is still a ball and thus geometrically easier to handle (right hand side of figure 1). The relation is as follows: If a schedule is feasible, i.e. can be operated by the unit without violating any technical constraint, it lies inside the feasible region (grey area on the left hand side in figure 1). Thus, the schedule is inside the pre-image (that represents the feasible region) of the ball and thus its image in the high-dimensional representation lies inside the ball. An infeasible schedule (e. g. $\boldsymbol{x}$ in Fig. 1) lies outside the feasible region and thus its image $\hat{\Psi}_x$ lies outside the ball. But we know some relations: the center of the ball, the distance of the image from the center and the radius of the ball. Hence, we can move the image of an infeasible schedule along the difference vector towards the center until it touches the ball. Finally, we calculate the pre-image of the moved image $\tilde{\Psi}_x$ and get a schedule at the boundary of the feasible region: a

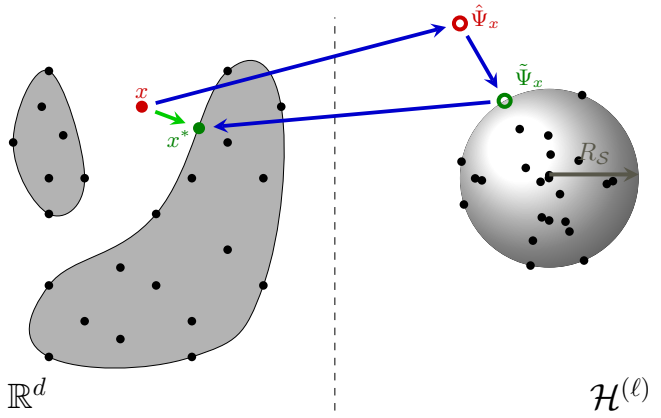


Fig. 1. General support vector model and decoder scheme for solution repair and constraint handling.

repaired schedule $\boldsymbol{x}^*$ that is now feasible. We do not need a mathematical description of the original feasible region or of the constraints to do this. The decoder that does the trick is derived directly from the training set $\mathcal{X}$ generated from the respective simulation model. More sophisticated variants of transformation are e. g. given in [4]. For a detailed description of the support vector decoder approach we refer to [4]. Formally, we have a mapping (the decoder γ)

$$\begin{aligned} \gamma : [0,1]^d &\rightarrow \mathcal{F}_{[0,1]} \subseteq [0,1]^d \\ \boldsymbol{x} &\mapsto \gamma(\boldsymbol{x}) \end{aligned} \tag{4}$$

that transforms any given (maybe in-feasible) schedule into a feasible one. Thus, we are able to transform the scheduling problem given Eq. (2) into an unconstrained formulation.

With these preliminaries in constraint handling we can now reformulate our optimization problem as

$$\delta\left(\sum_{i=1}^{m} \rho_i^{-1} \circ \gamma(\boldsymbol{x}_i), \boldsymbol{\zeta}\right) \rightarrow min, \tag{5}$$

where γ_i is the decoder function of unit i that produces feasible, scaled schedules from $\boldsymbol{x} \in [0,1]^d$ and σ_i^{-1} scales them unit specific entrywise to correct active power values (inverse to Eq. (1)) resulting in schedules that are operable by that unit. Please note, that this is a constraint free formulation. With this problem formulation, many standard algorithms for optimization can be easily adapted as there are no constraints (apart from a simple box constraint $\boldsymbol{x} \in [0,1]^d$) to be handled and no domain specific implementation (regarding the energy units and their operation schedules) has to be integrated. Equation (5) is used as a surrogate objective to find the solution to the constrained optimization problem equation (2).

Using a decoder fairly eases the implementation of a solver because no complex constraints have to be considered. On the other hand, such a decoder may introduce additional complexity into the optimization problem by the transformation. For this reason, we scrutinized the fitness landscapes of both problems (untransformed and transformed) to gain insight into the problem structure with means from standard fitness landscape analysis [14]. Indeed, our findings indicate a slightly growing in complexity by an increased ruggedness with a growing number of local minima [15]. But, this situation can be easily countered by using a heuristics that copes well with rugged non-linear problems like Simulated Annealing (SA).

Simulated Annealing [16] is an established Markov Chain Monte Carlo Method (MCMC) for non-linear optimization. It mimics a physical cooling process. In general, MCMC methods are an effective tool for statistical sampling applied to optimization problems [17]. The basic idea is a Markov Process that samples a target probability distribution $\pi(x) = \frac{1}{z} e^{-E(x)}$ with z as a problem specific normalization parameter and E measuring the error of the optimization objective. Originally, the method has been mainly applied to physical problems finding a minimum energy state and thus E is sometimes still written Hamiltonian $\mathcal{H}$, e.g. in [18]. We will use the term E. In this process a new state σ_{t+1} is generated from σ_t by drawing from a proposal transition distribution $Q(\sigma_{t+1}|\sigma_t)$ [19], [20]. The new state is accepted with probability

$$A(\sigma_t \rightarrow \sigma_{t+1}) = \min\left(1, \frac{\pi(\sigma_{t+1}) Q(\sigma_{t+1}|\sigma_t)}{\pi(\sigma_t) Q(\sigma_t|\sigma_{t+1})}\right). \tag{6}$$

The proposal distribution Q is a free parameter and must be adjusted to the individual problem at hand. Starting from a random initial state σ_0, the process needs a while to reach equilibrium and independence from σ_0. After this burn-in phase the samples represent the target distribution π.

In systems with deep local minima the process can be trapped without escape in reasonable time. This waiting time dilemma [21] is due to a stringent requirement for equilibrium. To escape, the process must generate subsequent states with higher energy and the probability for such a move declines roughly exponentially with the energy differences that has to be overcome. Thus, the expected waiting time for such escape grows also exponentially. For high-dimensional problems like the one that we scrutinize here, this problem is even more prevalent [21]. Several techniques have been proposed to overcome the problem of getting trapped, e.g. [22], [21], [23]; one is the concept of Simulated Annealing (SA).

SA introduces a variable temperature T into the target distribution: $\pi(x) = \frac{1}{z} e^{-E(\boldsymbol{x})/T}$. The effect is that the Markov Chain may escape local minima easier at a higher temperature. The general idea of Simulated Annealing is to interpret the fitness landscape of an optimization problem as a thermodynamic system with the objective function $E(\boldsymbol{x})$ denoting the error interpreted as the energy level of a proposed solution $\boldsymbol{x}$. Initially, the system is at a high temperature. During the Markov process, the system is gradually cooled down to the ground state with the global energy minimum.

Algorithm 1 shows the basic flow within our SA with integrated decoder. This integration has first been proposed in [15]. By mimicking a cooling process, temporarily worse solutions are allowed – depending on temperature and difference in solution quality – in order to escape local minima. In our approach, a solution is described by two matrices $\boldsymbol{X}_{ij}$ and $\boldsymbol{M}_{ij}$ denoting for each energy unit i and for each time interval

j of the schedule a scaled active power value in $[0, 1]$. In many objective scenarios, indicator values that describe the schedule with respect to different objectives might additionally be prevalent. For demonstration purposes, we stick with the single objective case here. In this sense, each row within the matrix is the schedule for one of the units. $\boldsymbol{X}$ contains schedules from the unconstrained search space (hypercube $[0, 1]^d$ not further constrained by technical issues from the units' operations). $\boldsymbol{X}$ is initialized with random values. $\boldsymbol{M}$ concurrently holds the respective feasible values generated by the support vector decoder: $\boldsymbol{M}_i = \gamma_i(\boldsymbol{X}_i)$. Thus, $\boldsymbol{M}$ always represents a feasible (scaled) solution to the problem.

$\boldsymbol{X}$ and $\boldsymbol{M}$ represent the genotype and phenotype of a solution respectively. In each iteration of the SA exactly one schedule $\boldsymbol{x}$ from $\boldsymbol{X}$ is randomly chosen and mutated. Modification is done at a randomly chosen element x_k by adding a random value $p \sim N(0, 1)$:

$$x_k \leftarrow \begin{cases} x_k + p - 1 & \text{if } x_k + p > 1 \\ x_k + p + 1 & \text{if } x_k + p < 0 \\ x_k + p & \text{else.} \end{cases} \tag{7}$$

Additionally, it can be useful especially for high-dimensional schedules to allow mutations at more than one element at a time. Only this mutated schedule has to be mapped by the respective decoder in order to keep $\boldsymbol{M}$ consistent with $\boldsymbol{X}$.

The system evolves as follows: at each temperature level T^t a Markov chain samples $E(\boldsymbol{x})$. $\boldsymbol{M}$ always represents a feasible, mutated solution that can be evaluated by Eq. (5). The new proposal solution part $\boldsymbol{x}^{t+1}$ is accepted (according to the Metropolis-Hastings criterion) with probability

$$A(\boldsymbol{x}^t \to \boldsymbol{x}^{t+1}) = \min\left(1, e^{\frac{-\Delta E}{T^t}}\right), \tag{8}$$

with $\Delta E = E(\boldsymbol{x}^{t+1}) - E(\boldsymbol{x}^t)$. In each iteration, temperature T^t is updated with with cooling rate $\lambda \in [0, 1[$: $T^{t+1} \leftarrow \lambda \cdot T^t$.

Algorithm 1 Basic scheme for the Simulated Annealing step (with integrated support vector decoder).

$\boldsymbol{X}_{ij} \leftarrow \boldsymbol{x}_i \sim U(0, 1)^d,\ 1 \le i \le n$
$\boldsymbol{M}_{ij} \leftarrow \gamma_i(\boldsymbol{X}_i),\ 1 \le i \le n$
$\vartheta \leftarrow \vartheta_{start}$
while $\vartheta < \vartheta_{min}$ **do**
  choose random $k;\ 1 \le k \le n$
  $\boldsymbol{x}^* \leftarrow \boldsymbol{X}_k$
  mutate($\boldsymbol{x}^*$)
  $\boldsymbol{M}^* \leftarrow \boldsymbol{M};\ \boldsymbol{M}^*_k \leftarrow \gamma_k(\boldsymbol{x}^*)$
  if $e^{-\frac{E(\boldsymbol{M}^*) - E(\boldsymbol{M})}{T}} > r \sim U(0, 1)$ **then**
    $\boldsymbol{M} \leftarrow \boldsymbol{M}^*;\ \boldsymbol{X}_k \leftarrow \boldsymbol{x}^*$
  end if
  $T \leftarrow \text{cooling}(T)$
end while

A major advantage of this approach is the anytime property: at any time, a feasible solution exists. The Markov chain may evolve in $[0, 1]^{d \cdot n}$ without taking care of technical constraints of the individual energy units. The decoder guarantees (apart from minor inaccuracies that might easily be corrected [4]) the feasibility of the solution.

III. Partitioning the Search Space

By employing the support vector decoder approach in combination with a heuristic solver for the optimization problem as described in the previous section, we are able to solve the scheduling problem for energy units efficiently without needing to adapt any part of the process to unit-specific properties such as technical constraints. The whole process is visualized in Algorithm 2. The resulting matrix $\boldsymbol{M}$ comprises m rows and d columns, where the i^{th} row vector represents the chosen schedule for energy unit U_i (for the remaining symbol definitions refer to Section II).

Algorithm 2 Predictive Scheduling

$m \leftarrow$ amount of energy units
$n \leftarrow$ sample size per energy unit
$d \leftarrow$ length of planning horizon
for all energy unit $U_i \in \mathcal{U}$ **do**
  $s_i \leftarrow$ predicted state of U_i at the beginning of the planning horizon
  repeat
    initialize simulation model for U_i with s_i
    simulate feasible schedule of length d
  until $\mathcal{F}^{(U_i)}$ contains n feasible schedules
  scale sample $\mathcal{F}^{(U_i)}$ using ρ_i
  calculate support vector model $\mathcal{S}_i$
  build support vector decoder γ_i
end for
return $\boldsymbol{M} \leftarrow \left(\text{solve } \delta\left(\sum_{i=1}^{m} \rho_i^{-1} \circ \gamma(\boldsymbol{x}_i), \boldsymbol{\zeta}\right) \to min\right)$

In the considered application domain, predictive planning is commonly done for *day-ahead* planning horizons, i. e. d corresponds to 24 hours with a schedule resolution of 15 minutes. In our problem formulation, this yields a 96-dimensional search space for each energy unit. Due to the curse of dimensionality [24], this may introduce significant negative effects. For instance, with larger problem dimensions, the required amount of training data for the support vector model increases exponentially [25]. This affects both the generation of feasible schedule samples via simulation, as well as learning the support vector models from these samples. Moreover, solving the optimization problem itself gets more time-consuming due to combinatorial explosion. Finally, as the support vector decoder model is based on approximation, mapping accuracy deteriorates with larger dimensions. This may lead to infeasible schedules being misleadingly recognized as feasible.

According to [26], strategies to circumvent the curse of dimensionality in such a case can be categorized as follows:

- *Decomposition:* Given that the problem is separable, decomposition subdivides the problem into smaller parts that are easier to solve.

- *Screening:* Less significant and redundant decision variables/dimensions are pruned from the problem description in order to reduce dimensionality.
- *Mapping:* The problem is mapped to a representation comprising less dimensions. For example, by exploiting correlations between variables in the original space, a mapping can be designed that yields a correlation-free space with less dimensions.
- *Space Reduction:* Using expert knowledge, parts of the search space are excluded from optimization.
- *Visualization:* An expert prunes insignificant parts of the search space using visualization techniques for high-dimensional data. In contrast to *Space Reduction*, this is done interactively during the optimization process.

For the considered support vector decoder approach, the strategies Screening, Visualization, and Space Reduction with expert's help are inappropriate, as they rely on specific knowledge about the individual problem instance to solve, which contradicts the main motivation for our approach. Because neigbouring values in the unit schedules are often quite similar (i. e. the gradient between two time intervals is usually rather small) and thus show some correlation, Mapping might be applicable. After optimization, however, the resulting low-dimensional power profile would have to be inversely mapped to a feasible high-dimensional schedule again, which would introduce further problems.

Finally, Decomposition offers a viable solution. We cannot split the problem along the m axis with respect to the result matrix $\boldsymbol{M}$ in Algorithm 2 (i. e. by optimizing over disjunct sets of energy units), because in each time step along the d axis, the schedule selections of *all* participating units have to be regarded in order to minimize δ. On the other hand, the problem formulation might allow us to optimize over each time step along the d axis independently: If the employed distance measure δ is a *metric*, it gets minimal if the individual distances along the d axis are minimal. This holds true for the Euclidean distance $\|\cdot\|_2$ we are using in this paper. Therefore, from the optimization point of view, the given problem seems to be separable along the d axis. Formally, we define such a *partitioning* of the search space as

$$\begin{aligned} \pi &: \mathbb{N}^2 \to \mathbb{N} \\ (d, j, k) &\mapsto l = \pi(d, j, k), \end{aligned} \tag{9}$$

where $l = \pi(d, j, k)$ denotes the length of the j^{th} partition along the d axis. The parameter k may hold arbitrary implementation-specific values (cf. the equidistant partitioning below). For a partitioning to be valid, the concatenation of all generated partitions must yield the whole planning horizon:

$$\sum_{j=1}^{\infty} \pi(d, j, k) = d \tag{10}$$

Moreover, for convenience we require

$$\forall i: \ \pi(d, j, k) = 0 \ \Rightarrow \ \pi(d, j+1, k) = 0, \tag{11}$$

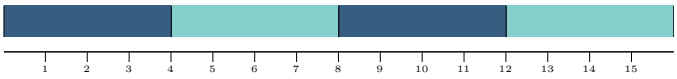


Fig. 2. Equidistant partitioning for $d = 16$ and $k = 4$.

i. e. as soon as the partitioning function yields the first zero partition, every following partition must be zero as well. Using this rather general definition of π, we may now define different partitioning strategies. For example, the *equidistant partitioning* subdivides the planning horizon into k partitions of equal size

$$\pi_{\text{eq}}(d, j, k) = \begin{cases} \lceil \frac{d}{k} \rceil & \text{if } j \leq k \ \wedge \ j \leq d \bmod k, \\ \lfloor \frac{d}{k} \rfloor & \text{if } j \leq k \ \wedge \ j > d \bmod k, \\ 0 & \text{else.} \end{cases} \tag{12}$$

Figure 2 shows an example for this partitioning with $d = 16$ and $k = 4$. There are many other possible partitioning strategies, ranging from simple arithmetic fragmentations to more sophisticated strategies involving expert knowledge about the use case at hand (i. e. the structure of the target profile or the δ function). A particular promising approach is the *entropy partitioning*, which exploits the entropy in the feasible schedule samples to determine intervals of high vs. low flexibility in the units' scopes of actions, and partitions the search space accordingly. But in order to remain maximally independent from such expert knowledge, we go with the example of equidistant partitioning in the remainder of this paper.

In order to implement a partitioning scheme like e. g. the equidistant partitioning in our approach, we have to extend Algorithm 2. Special care has to be taken regarding the simulation of feasible schedules: Originally, in Algorithm 2, each simulation model was initialized with the state of the energy unit right at the beginning of the planning horizon, and was executed for d time steps, such that each schedule sample exactly covers the planning horizon. Using partitions, however, schedule samples cannot be generated beforehand for the whole planning horizon. In order to identify a unit's flexibility for a certain partition, the exact state of the unit at the beginning of this partition has to be known. Thus, before being able to process a partition, we have to assign fixed schedules to the units for the preceding partition. As a consequence, the overall process ranging from schedule simulation to solving the optimization problem has to be executed for each partition separately. This ensures that, after the process finished for all partitions, the concatenated result schedules are feasible overall. On the other hand, with this approach we achieve a reduction of the design space (without expert knowledge as proposed in [26]) as every subsequent optimization process is already tackled to a fixed operational state of each unit at the beginning of a partition. The resulting process is visualized in Algorithm 3.

IV. Evaluation

The objective of this paper is to explore the potential benefit of partitioning the search space of the given combinatorial

Algorithm 3 Predictive Scheduling with Partioning

1: $m \leftarrow$ amount of energy units
2: $n \leftarrow$ sample size per energy unit
3: $d \leftarrow$ length of planning horizon
4: **for all** energy unit $U_i \in \mathcal{U}$ **do**
5: $s_i \leftarrow$ predicted state of U_i at the beginning of the planning horizon
6: **end for**
7: $j \leftarrow 1$
8: $k \leftarrow$ implementation specific value
9: **while** $\pi(d, j, k) \neq 0$ **do**
10: **for all** energy unit $U_i \in \mathcal{U}$ **do**
11: **repeat**
12: initialize simulation model for U_i with s_i
13: simulate feasible schedule of length $\pi(d, j, k)$
14: **until** $\mathcal{F}^{(U_i)}$ contains n feasible schedules
15: scale sample $\mathcal{F}^{(U_i)}$ using ρ_i
16: calculate support vector model $\mathcal{S}_i$
17: build support vector decoder γ_i
18: **end for**
19: $\boldsymbol{M}^j \leftarrow \left(\text{solve } \delta\left(\sum_{i=1}^{m} \rho_i^{-1} \circ \gamma(\boldsymbol{x}_i), \boldsymbol{\zeta}\right) \rightarrow min\right)$
20: **for all** energy unit $U_i \in \mathcal{U}$ **do**
21: run simulation model for U_i using schedule $\boldsymbol{x}_i$
22: $s_i \leftarrow$ predicted state of U_i after running $\boldsymbol{x}_i$
23: **end for**
24: $j \leftarrow j + 1$
25: **end while**
26: **return** $\boldsymbol{M} \leftarrow \left[\boldsymbol{M}^j\right]$

problem in the data domain, followed by sequential optimization of the individual partitions. In the previous section, a partitioning framework has been introduced for this, along with a detailed description of the according optimization process chain. In order to evaluate the proposed approach with respect to the objective, a simulation study has been conducted.

A. Simulation Setup

Following the considered example use case, we set up a simulated virtual power plant for active power planning in day-ahead scenarios, comprising CHP units with an 800 l thermal buffer store each. We used the simulation model of an EcoPower CHP as described in [3]. For each of those devices, the thermal demand for a four-family house during winter was simulated. The devices were operated in heat driven operation and thus primarily had to compensate the simulated thermal demand. Additionally, after shutting down, a device would have to stay off for at least two hours. However, due to their thermal buffer store and the ability to modulate the electrical power output within the range of $[1.3,\ 4.7]$, the devices still have some flexibility available.

For the generation of feasible schedule samples, a *successive sampling strategy* was employed: Instead of guessing whole schedules and checking feasibility afterwards (using a device's simulation model), which leads to large rejection rates, a period-wise guessing in combination with partial feasibility checks is applied repeatedly to construct feasible schedules in a successive manner, cf. [12]. Preliminary experiments indicated 200 as an adequate size for $\mathcal{F}^{(U_i)}$, so we set $n = 200$ in the present study.

The planning horizon was set to $d = 96$ time intervals, i. e. 24 hours in 15 minute resolution, which is a common use case in the application domain. As motivated in the previous section, we employ the equidistant partitioning function π_{eq} in this study. Regarding the parameter k, which defines the length of the partitions and thus inversely determines the number of partitions to be generated according to (12), several experiments with $k \in [1, 96]$ have been conducted. For instance, $k = 1$ yields 96 partitions of length 1, while $k = 96$ corresponds to a single partition of length 96, i. e. no partitioning at all. This way, the influence of a partitioning on the optimization can be explored in a structured manner.

While k represents the primary influence factor in our study, other parameters may cause relevant interaction effects. Here, especially the magnitude of the problem size along the m axis (i. e. the number of energy units in the VPP, cf. Section III) is of particular interest, as it affects the problem complexity for each partition likewise. Similarly, different target profiles ζ have to be examined with respect to the units' available flexibilities. For example, a target profile might turn out to be easily realizable due to well matching schedule options in the units' search spaces, or vice-versa. The question arises whether this influences the potential benefit of a partitioning, and how a partitioning should be done in order to gain optimal results.

In all experiments, we used the Simulated Annealing solver as outlined in Section II. Each examined parameter configuration was simulated 100 times, so that the results can be interpreted with statistical soundness.

B. Results

The evaluation focuses on solution quality, which is calculated as remaining error after optimization:

$$\delta\left(\sum_{i=1}^{m} \rho_i^{-1}(\boldsymbol{x}_i), \boldsymbol{\zeta}\right), \quad \boldsymbol{x}_i \in \boldsymbol{M} \tag{13}$$

where $\boldsymbol{M}$ denotes the $m \times d$ schedule matrix after all partitions have been processed (line 26 in Algorithm 3). In the following, results are visualized as box-charts, where the box spans from the upper to the lower quartile of the data. The median is shown as horizontal line within a box, whereas the whiskers span over $1.5 \times$ the interquartile range. Outliers are illustrated by circle markers.

First of all, the general influence of different k values (i. e. different partition sizes) is examined. As already stated in the introduction, preliminary experiments indicated a potentially increasing solution quality when the optimization algorithm is applied in a successive manner to sequential partitions of the planning horizon. For a more thorough analysis, we conducted 100 simulations for each $k \in \{2, 8, 24, 48, 96\}$. The results are visualized in Figure 3. The optimization error clearly decreases with more and thus smaller partitions (from right to left in the figure). Comparing the extreme points, the partitioning even

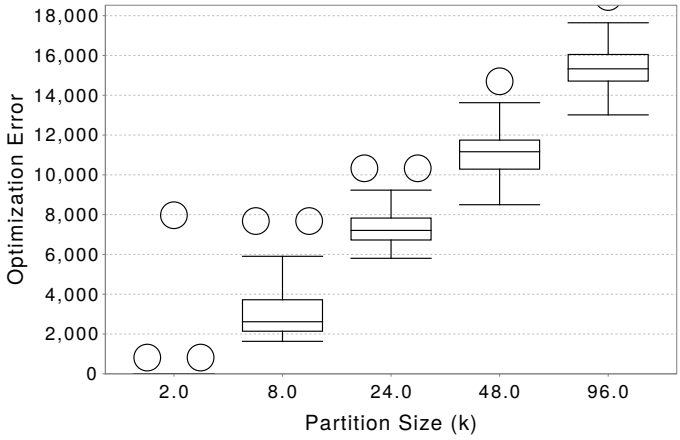


Fig. 3. Remaining optimization error for different partition sizes.

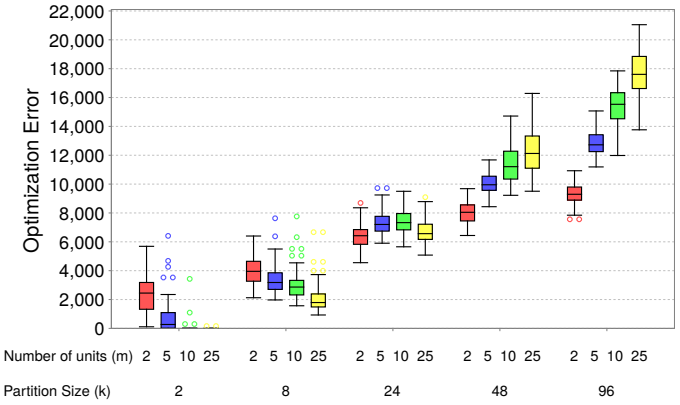


Fig. 4. Remaining optimization error for different partition sizes and varying amounts of energy units.

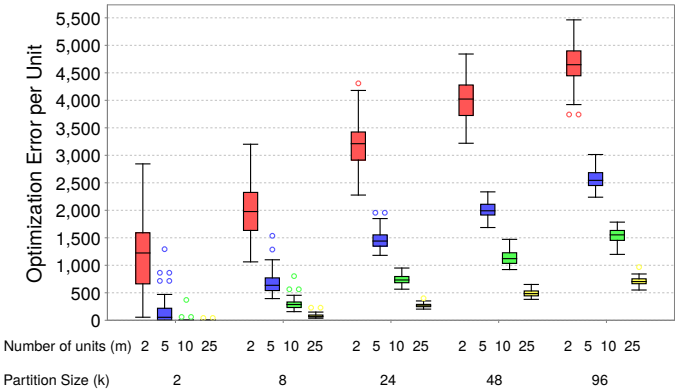


Fig. 5. Remaining optimization error per energy unit for different partition sizes and varying amounts of energy units.

allows approaching the theoretical optimum $\delta = 0$ when the partitions are generated as small as possible ($k = 2$: despite a few outliers, the box is squashed to a single line at $\delta = 0$), while the no-partitioning case yields the worst results most of the time ($k = 96$).

These results support our hypothesis strikingly, but they originate from a single experiment configuration only: On the one hand, a fixed number of energy units was involved, $m = 10$. On the other hand, the target profile ζ was generated by aggregating randomly chosen sample schedules (one for each energy unit) at the beginning of each experiment run. This way, ζ formed an "easy" target, because the energy units were able to approach it optimally in principle. In the following, we will vary this configuration in these two aspects, in order to gain more insights into the involved effects.

1) Interaction with the Number of Energy Units: In the considered application use-case of predictive scheduling for active power planning in day-ahead scenarios, virtual power plants may comprise different amounts of energy units, depending on e. g. regional conditions. From the optimization point of view, this corresponds to the problem size along the m axis. In a partitioned setting (i. e. $k < d$), each subproblem is of size $m \times k$. Hence, m affects the problem complexity for each partition likewise. To reveal possible interactions with the magnitude of k, the previous experiment with $k \in \{2, 8, 24, 48, 96\}$ was repeated for $m \in \{2, 5, 10, 25\}$. Figure 4 visualizes the results. Similar to Figure 3, the optimization error generally decreases with smaller partitions. Within each block, however, different effects with respect to the magnitude of m are visible: For the case of small partitions, the optimization error is lower with larger values of m, while this trend reverses for large partitions. As this is based on the absolute error, which is naturally different for varying magnitudes of m, Figure 5 complementarily shows the same results against the normed optimization error with respect to the number of units, i. e. the remaining error per energy unit. Here, the trend towards a lower error for small partitions is again clearly visible, whereas the magnitude of m results in a change of the slope for this trend. Concluding, m seems to affect the problem complexity as a whole only, and does not seem to interact with the partition size k.

2) Interaction with the Target Profile: In active power planning, usually an application-specific target profile is given. For instance, in day-ahead energy market scenarios, a target profile would be chosen such that the economic outcome of the VPP is maximized. In contrast, in supply-demand-matching scenarios, the target profile might be e. g. a constant zero value, such that the considered set of energy units (flexible producers and consumers) can be treated as autonomous energy-wise. While it is advisable to configure VPP and target profile in a matching way, so that the latter is actually a feasible target for the former, not all target profiles are equally easy to realize.

In our study, we abstract from application-specific scenarios as follows. As a first step, a feasible target can be formed by aggregating randomly chosen sample schedules (one for each energy unit). This way, the existence of the theoretical optimum ($\delta = 0$) is guaranteed. We denote this type of target with ζ_0. To generate more difficult target profiles in an easy but structured way, ζ_0 can simply be shifted in magnitude:

$$\zeta_i = \zeta_0 + i \tag{14}$$

Please note that ζ is a vector, and the summation is performed element-wise, of course. Matching the size of the considered

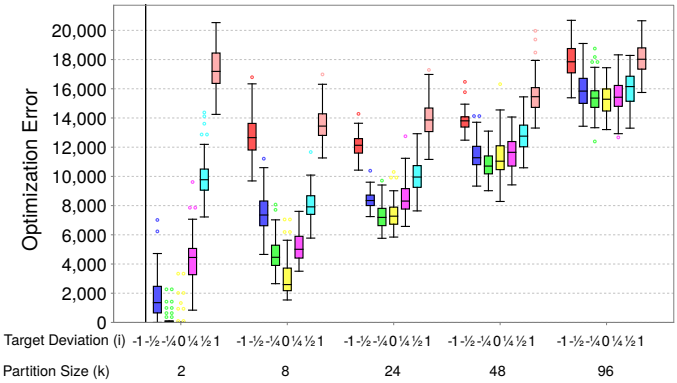


Fig. 6. Remaining optimization error for different partition sizes and varying target profile deviations.

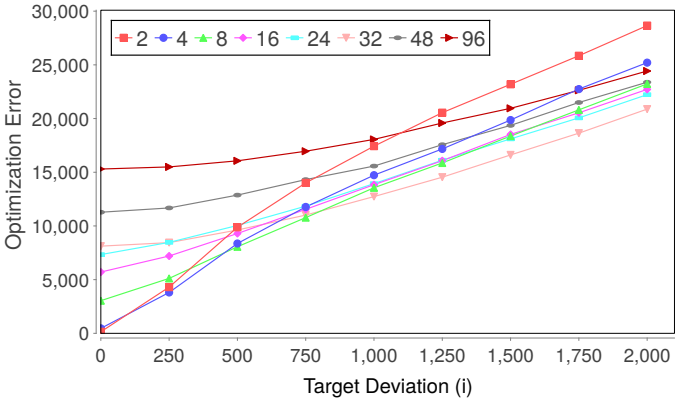


Fig. 7. Remaining optimization error for varying target profile deviations (horizontal axis) and different partition sizes (individual data series).

VPP in the present study, we choose values for i between $0\,\mathrm{kW}$ and $\pm 1\,\mathrm{kW}$ in the following experiment, in order to deviate the target profile from "easy to solve optimally" towards "hard to solve optimally". Thus, as in the previous section, the original experiment with $k \in \{2, 8, 24, 48, 96\}$ was repeated for all ζ_i with $i \in \{-1, -0.5, -0.25, 0, 0.25, 0.5, 1\}$ in kW. The results are presented in Figure 6. Similar to the results from Figure 4, the general trend of better optimization results with smaller partitions is visible. The case $k = 2,\ i = -1$ is an exception. Here, the optimization was not able to find a feasible schedule at all in the available time. This is due to the very low values in the target profile in combination with a large number of partitions: Due to the independent optimization of individual partitions, the simulated CHP units stay off at the beginning of the planning horizon until the thermal buffer stores are exhausted. At that point in time, however, thermal demand exceeds the available power from the CHPs, so that no feasible schedule cannot be found anymore. With larger partitions, the effect is not present, as the optimization can act anticipatory towards feasibility (i. e. by choosing schedules that lead to a poor optimization error, but in turn form a feasible solution). This effect indicates that a strong partitioning can yield better optimization results if enough flexibility is present, but might also lead to infeasible solutions in extreme cases.

In addition to the general trend regarding the value of k, a u-shaped course can be seen within each configuration of the same partition size. In other words, solution quality seems to deteriorate with larger deviations from ζ_0, which is not surprising at all. In order to focus on the interaction between these two effects, Figure 7 visualizes the results in a transposed way, i. e. the deviation i is visualized along the horizontal axis, while the partition sizes k are presented as line charts. For visualization purposes, the shown data comprises mean values only. Furthermore, in this experiment a larger amount of configurations was examined: $k \in \{2, 4, 8, 16, 24, 32, 48, 96\}$ and $|i| \in \{0, 0.25, \ldots, 2\}$. The results reveal an interesting relationship: For smaller target deviations, configurations with smaller partitions yield superior optimization results. In contrast, for larger target deviations, larger partition sizes yield better results.

In summary, the last experiment supports our previous hypothesis: With enough flexibility in a given problem configuration (in terms of feasible solution combinations with respect to the fitness function), the solver significantly benefits from a partitioning. On the other hand, in more difficult problem formulations (i. e. with less flexibility in terms of feasible solutions), the solver cannot cope with a large number of independent partitions.

V. Conclusion

The objective of this paper was to explore the potential benefit of partitioning the search space of the given combinatorial problem in the data domain using the example of predictive scheduling in the smart grid domain. We combined the partitioning approach with a sequential optimization solving each partition successively. Simulation models of different energy units have been integrated directly in the process for handling individual search spaces and operational constraints.

Several methods to cope with the challenge of high dimensionality in optimization problems have been proposed in the past. A good overview on methods for computationally expensive black-box functions (as might be the case when using simulation models for computing objectives) is e. g. given in [26]. Our approach is a mixture of design space reduction and decomposition into sub-problems. To achieve this we have to introduce simulation models as black-boxes into the optimization process for sequentialization. Introducing this sequence of independently solvable sub-problems reduces the overall computationally effort and at the same time reduces the design space so that modeling is more accurate and optimization effort is reduced [26]. At the same time this reduction leads to a limited choice especially for later sub-problems. Sub-space parts of the design space may be missed [26]. On the other hand, with our method, we may focus on the whole sub-space at once without a need for subsequent refinement like in other methods [26].

Our results support the hypothesis of an increasing solution quality when applying the optimization algorithm in a

successive manner to sequential partitions of the planning horizon. For our experiments we mainly used a simulated annealing approach as solver although our results can be generalized to other solvers. In general, any solver benefits for partitioned data domains in predictive scheduling if a problem configuration contains enough flexibility in terms of feasible solution combinations. With decreasing flexibility, additional complexity induced by a growing number of partitions prevails.

So far all simulations have been done with scenarios regarding predictive scheduling. Additional use cases like load balancing can be easily adapted by exchanging the objective functions, as the problem structure is similar to predictive scheduling. Future work will concentrate on methods to classify the situation at hand in order to automatically decide on appropriate partition of the combinatorial problem.

REFERENCES

[1] P. Palensky and D. Dietrich, "Demand side management: Demand response, intelligent energy systems, and smart loads," *Industrial Informatics, IEEE Transactions on*, vol. 7, no. 3, pp. 381–388, Aug 2011. doi: 10.1109/TII.2011.2158841. [Online]. Available: http://dx.doi.org/10.1109/TII.2011.2158841

[2] N. . P. F. Arteconi, A. ; Hewitt, "Domestic demand-side management (dsm): Role of heat pumps and thermal energy storage (tes) systems," *Applied Thermal Engineering, 2013, Vol.51(1-2), pp.155-165*, vol. 51, no. 1-2, p. 155. doi: 10.1016/j.applthermaleng.2012.09.023. [Online]. Available: http://dx.doi.org/10.1016/j.applthermaleng.2012.09.023

[3] J. Bremer and M. Sonnenschein, "Model-based integration of constrained search spaces into distributed planning of active power provision," *Comput. Sci. Inf. Syst.*, vol. 10, no. 4, pp. 1823–1854, 2013. doi: 10.2298/CSIS130304073B. [Online]. Available: http://dx.doi.org/10.2298/CSIS130304073B

[4] ——, "Constraint-handling for optimization with support vector surrogate models – A novel decoder approach," in *ICAART 2013 – Proceedings of the 5th International Conference on Agents and Artificial Intelligence*, J. Filipe and A. Fred, Eds., vol. 2. Barcelona, Spain: SciTePress, 2013. doi: 10.5220/0004241100910100. ISBN 978-989-8565-38-9 pp. 91–100. [Online]. Available: http://dx.doi.org/10.5220/0004241100910100

[5] J. Bremer, B. Rapp, and M. Sonnenschein, "Encoding distributed search spaces for virtual power plants," in *IEEE Symposium Series on Computational Intelligence 2011 (SSCI 2011)*, Paris, France, 4 2011. doi: 10.1109/CIASG.2011.5953329. [Online]. Available: http://dx.doi.org/10.1109/CIASG.2011.5953329

[6] S. McArthur, E. Davidson, V. Catterson, A. Dimeas, N. Hatziargyriou, F. Ponci, and T. Funabashi, "Multi-agent systems for power engineering applications—Part I: Concepts, approaches, and technical challenges," *IEEE Transactions on Power Systems*, vol. 22, no. 4, pp. 1743–1752, 2007. doi: 10.1109/TPWRS.2007.908471. [Online]. Available: http://dx.doi.org/10.1109/TPWRS.2007.908471

[7] M. Sonnenschein, C. Hinrichs, A. Nieße, and U. Vogel, "Supporting renewable power supply through distributed coordination of energy resources," in *ICT Innovations for Sustainability*, ser. Advances in Intelligent Systems and Computing, L. M. Hilty and B. Aebischer, Eds. Springer International Publishing, 2015, vol. 310, pp. 387–404. ISBN 978-3-319-09227-0. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-09228-7_23

[8] A. Nieße, S. Beer, J. Bremer, C. Hinrichs, O. Lünsdorf, and M. Sonnenschein, "Conjoint dynamic aggregation and scheduling for dynamic virtual power plants," in *Federated Conference on Computer Science and Information Systems - FedCSIS 2014, Warsaw, Poland*, M. Ganzha, L. A. Maciaszek, and M. Paprzycki, Eds., 9 2014. doi: 10.15439/2014F76. [Online]. Available: http://dx.doi.org/10.15439/2014F76

[9] J. Bremer and M. Sonnenschein, "A distributed greedy algorithm for constraint-based scheduling of energy resources," in *FedCSIS*, M. Ganzha, L. A. Maciaszek, and M. Paprzycki, Eds., 2012. ISBN 978-83-60810-51-4 pp. 1285–1292.

[10] C. A. Coello Coello, "Theoretical and numerical constraint-handling techniques used with evolutionary algorithms: a survey of the state of the art," *Computer Methods in Applied Mechanics and Engineering*, vol. 191, no. 11-12, pp. 1245–1287, Jan. 2002. doi: 10.1016/S0045-7825(01)00323-1. [Online]. Available: http://dx.doi.org/10.1016/S0045-7825(01)00323-1

[11] S. Koziel and Z. Michalewicz, "Evolutionary algorithms, homomorphous mappings, and constrained parameter optimization," *Evol. Comput.*, vol. 7, pp. 19–44, 03 1999. doi: 10.1162/evco.1999.7.1.19. [Online]. Available: http://dx.doi.org/10.1162/evco.1999.7.1.19

[12] J. Bremer and M. Sonnenschein, "Sampling the search space of energy resources for self-organized, agent-based planning of active power provision," in *EnviroInfo*, ser. Berichte aus der Umweltinformatik. Shaker, 2013, pp. 214–222.

[13] D. M. J. Tax and R. P. W. Duin, "Support vector data description," *Mach. Learn.*, vol. 54, no. 1, pp. 45–66, 2004. doi: 10.1023/B:MACH.0000008084.60811.49. [Online]. Available: http://dx.doi.org/10.1023/B:MACH.0000008084.60811.49

[14] V. K. Vassilev, T. C. Fogarty, and J. F. Miller, "Information characteristics and the structure of landscapes," *Evol. Comput.*, vol. 8, no. 1, pp. 31–60, Mar. 2000. doi: 10.1162/106365600568095. [Online]. Available: http://dx.doi.org/10.1162/106365600568095

[15] J. Bremer and M. Sonnenschein, "Parallel tempering for constrained many criteria optimization in dynamic virtual power plants," in *2014 IEEE Symposium on Computational Intelligence Applications in Smart Grid, CIASG 2014, Orlando, FL, USA, December 9-12, 2014*. IEEE, 2014. doi: 10.1109/CIASG.2014.7011551 pp. 51–58. [Online]. Available: http://dx.doi.org/10.1109/CIASG.2014.7011551

[16] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, no. 4598, pp. 671–680, 1983. doi: 10.1126/science.220.4598.671. [Online]. Available: http://dx.doi.org/10.1126/science.220.4598.671

[17] Y. Li, V. A. Protopopescu, N. Arnold, X. Zhang, and A. Gorin, "Hybrid parallel tempering and simulated annealing method," *Applied Mathematics and Computation*, vol. 212, no. 1, pp. 216–228, 2009. doi: 10.1016/j.amc.2009.02.023. [Online]. Available: http://dx.doi.org/10.1016/j.amc.2009.02.023

[18] A. Müller, J. J. Schneider, and E. Schömer, "Packing a multidisperse system of hard disks in a circular environment," *Phys. Rev. E*, vol. 79, p. 021102, Feb 2009. doi: 10.1103/PhysRevE.79.021102. [Online]. Available: http://dx.doi.org/10.1103/PhysRevE.79.021102

[19] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equation of state calculations by fast computing machines," *The Journal of Chemical Physics*, vol. 21, no. 6, pp. 1087–1092, 1953. doi: 10.1063/1.1699114. [Online]. Available: http://dx.doi.org/10.1063/1.1699114

[20] W. K. Hastings, "Monte carlo sampling methods using markov chains and their applications," *Biometrika*, vol. 57, no. 1, pp. 97–109, 1970. doi: 10.1093/biomet/57.1.97. [Online]. Available: http://dx.doi.org/10.1093/biomet/57.1.97

[21] W. H. Wong and F. Liang, "Dynamic weighting in Monte Carlo and optimization," *Applied Mathematics. Proceedings of the National Academic of Science*, vol. 94, pp. 14 220–14 224, Dec. 1997.

[22] E. Marinari and G. Parisi, "Simulated tempering: a new Monte Carlo scheme," *Europhys. Lett.*, vol. 19, no. 6, 1992.

[23] S. Brown and T. Head-Gordon, "Cool walking: A new markov chain monte carlo sampling method." *Journal of Computational Chemistry*, vol. 24, no. 1, pp. 68–76, 2003. doi: 10.1002/jcc.10181. [Online]. Available: http://dx.doi.org/10.1002/jcc.10181

[24] D. L. Donoho, "High-dimensional data analysis: The curses and blessings of dimensionality. aide-memoire of a lecture at," in *AMS Conference on Math Challenges of the 21st Century*, 2000.

[25] M. Verleysen and D. François, "The curse of dimensionality in data mining and time series prediction," in *Computational Intelligence and Bioinspired Systems, Lecture Notes in Computer Science 3512*. Springer, 2005. doi: 10.1007/11494669_93 pp. 758–770. [Online]. Available: http://dx.doi.org/10.1007/11494669_93

[26] S. Shan and G. G. Wang, "Survey of modeling and optimization strategies to solve high-dimensional design problems with computationally-expensive black-box functions," *Structural and Multidisciplinary Optimization*, vol. 41, no. 2, pp. 219–241, 2010. doi: 10.1007/s00158-009-0420-2. [Online]. Available: http://dx.doi.org/10.1007/s00158-009-0420-2

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 561–570
DOI: 10.15439/2016F201
ACSIS, Vol. 8. ISSN 2300-5963

Heuristics for Job Scheduling Reoptimization

Elad Iwanir, Tami Tamir
School of Computer Science
The Interdisciplinary Center, Herzliya, Israel

***Abstract*—Many real-life applications involve systems that change dynamically over time. Thus, throughout the continuous operation of such a system, it is required to compute solutions for new problem instances, derived from previous instances. Since the transition from one solution to another incurs some cost, a natural goal is to have the solution for the new instance close to the original one (under a certain distance measure). We study reoptimization problems arising in scheduling systems. Formally, due to changes in the environment (out-of-order or new machines, modified jobs' processing requirements, etc.), the schedule needs to be modified. That is, jobs might be migrated from their current machine to a different one. Migrations are associated with a cost – due to relocation overhead and machine set-up times. In some systems, a migration is also associated with job extension. The goal is to find a good modified schedule, with a low transition cost from the initial one.**

We consider reoptimization with respect to the classical objectives of minimum makespan and minimum total flow-time. We first prove that the reoptimization variants of both problems are NP-hard, already for very restricted classes. We then develop and present several heuristics for each objective, implement these heuristics, compare their performance on various classes of instances and analyze the results.

I. Introduction

REOPTIMIZATION problems arise naturally in dynamic scheduling environments, such as manufacturing systems and virtual machine managers. Due to changes in the environment (out-of-order or new resources, modified jobs' processing requirements, etc.), the schedule needs to be modified. That is, jobs may be migrated from their current machine to a different one. Migrations are associated with a cost due to relocation overhead and machine set-up times. In some systems, a migration is also associated with job extension. The goal is to find a good modified schedule, with a low transition cost from the initial one.

This work studies the reoptimization variant of two classical scheduling problems in a system with identical parallel machines: (i) minimizing the total flow-time (denoted in standard scheduling notation by $P||\Sigma C_j$ [13]), and (ii) minimum makespan (denoted by $P||C_{max}$).

The minimum total flow-time problem for identical machines can be solved efficiently by the simple greedy Shortest Processing Time algorithm (SPT) that assigns the jobs in non decreasing order by their length. The minimum makespan problem is NP-hard, and has several efficient approximation algorithms, as well as a polynomial-time approximation scheme [15]. These algorithms, as many other algorithms for combinatorial optimization problems, solve the problem from scratch, for a single arbitrary instance without having any constraints or preferences regarding the required solution - as long as they achieve the optimal objective value. However, many of the real-life scenarios motivating these problems involve systems that change dynamically over time. Thus, throughout the continuous operation of such a system, it is required to compute solutions for new problem instances, derived from previous instances. Moreover, since the transition from one solution to another consumes energy (used for the physical migration of the job, for warm-up or set-up of the machines, or for activation of the new machines), a natural goal is to have the solution for the new instance close to the original one (under certain distance measure). Solving a reoptimization problem involves two challenges:

1) Computing an optimal (or close to the optimal) solution for the new instance.
2) Efficiently converting the current solution to the new one.

Each of these challenges, even when considered alone, gives rise to many theoretical and practical questions. Obviously, combining the two challenges is an important goal, which shows up in many applications.

A. Problem description

An instance of our problem consists of a set J of n jobs and a set M_0 of m_0 identical machines. Denote by p_j the processing time of job j. An initial schedule S_0 of the jobs is given. That is, for every machine, it is specified what are the jobs it processes. At any time, a machine can process at most one job and a job can be processed by at most one machine. We consider the scenario in which a change in the system occurs. Possible changes include addition or removal of machines and/or jobs, as well as modification of processing times of jobs in J. We denote by M the set of machines in the modified instance and let $m = |M|$.

Our goal is to suggest a new schedule, S, for the modified instance, with good objective value and small transition cost form S_0. Assignment of a job to a different machine in S_0 and S is denoted *migration* and is associated with a cost. Formally, we are given a *price list* $\theta_{i,i',j}$, such that it costs $\theta_{i,i',j}$ to migrate job j from machine i to machine i'. Moreover, in some systems job migrations are also associated with an extension of the job's processing

978-83-60810-90-3/$25.00

time. Formally, in addition to the transition costs, we are given a job-extension penalty list $\delta_{i,i',j} \geq 0$, such that the processing time of job j is extended to $p_j + \delta_{i,i',j}$ when it is migrated from machine i to machine i'.

For a given schedule, let C_j denote the *flow-time* (also known as 'completion time') of job j, that is, the time when the process of j completes. The *makespan* of a schedule is defined as the maximal completion time of a job, that is, $C_{max} = \max_j C_j$.

While the schedule does not specify the internal order of jobs assigned to a machine, we assume throughout this work that jobs assigned to a specific machine are always processed in SPT-order (Shortest Processing Time), that is, $p_1 \leq p_2 \leq \ldots$. For a given set of jobs, SPT-order is known to achieve the minimal possible total flow-time, $\min \sum_j C_j$ [22], [8]. Clearly, the internal order has no effect on the makespan.

Given S_0, J and M, the goal is to find a good schedule for J that is close to the initial schedule S_0. We consider two problems:

1) Rescheduling to an optimal schedule using the minimal possible transition cost.
2) Given a budget B, find the best possible modified schedule that can be achieved without exceeding the budget B.

If the modification includes machines' removal, and the budget is limited, then we assume that a feasible solution exists. That is, the budget is at least the cost of migrating the jobs on the removed machines to some machine. Formally, $B \geq \sum_{j \ on \ i \in M_0 \setminus M} \min_{i' \in M} \theta_{i,i',j}$.

Applications: Our reoptimization problems arise naturally in manufacturing systems, where jobs may be migrated among production lines. Due to unexpected changes in the environment (out-of-order or new machines, timetables of task processing, etc.), the production schedule needs to be modified. Rescheduling tasks involves energy-loss due to relocation overhead and machine set-up times. In fact, our work is relevant to any dynamic scheduling environment, in which migrations of jobs are allowed though associated with an overhead caused due to the need to handle the modification and to absorb the migrating jobs in their new assignment.

With the proliferation of cloud computing, more and more applications are deployed in the data centers. Live migration is a common process in which a running virtual machine (VM) or application moves between different physical machines without disconnecting the client or application [7]. Memory, storage, and network connectivity of the virtual machine are transferred from the original host machine to the destination. Such migrations involve a warm-up phase, and a memory-copy phase. In pre-copy memory migration, the Hypervisor typically copies all the memory pages from source to destination while the VM is still running on the source. Alternatively, in post-copy memory migration the VM is suspended, a minimal subset of the execution state of the VM (CPU state, registers and, optionally, non-pageable memory) is transferred to the target, and the VM is then resumed at the target. Live migration is performed in several VM managers such as Parallels Virtuozzo [18] and Xen [24]. Sequential processing of jobs that might be migrated among several processors is performed also in several implementations of MapReduce (e.g., [3]), and in RPC (Remote Procedure Call) services, in which virtual servers can be temporarily rented [4].

B. Related Work

The work on reoptimization problems started with the analysis of dynamic graph problems (see e.g. [10], [23]). These works focus on developing data structures supporting update and query operations on graphs. A different line of research, deals with the computation of a good solution for an NP-hard problem, given an optimal solution for a close instance. Among the problems studied in this setting are TSP [1], [5] and Steiner Tree on weighted graphs [11].

The paper [21] suggests the framework we adopt for this work, in which the solution for the modified instance is evaluated also with respect to its difference from the initial solution. This framework is in use also in [20], to analyze algorithms for data placement in storage area network. Job scheduling reoptimization problems with respect to the total flow-time objective was studied in [2], were algorithms for finding an optimal scheduled using the minimal possible transition cost and algorithms for optimal utilization of a limited number of migration were described.

Lot of attention, in both the industry and the academia is given recently to the problem of minimizing the overhead associated with migrations (see e.g., [7], [14]). Using our notations, this refers to minimizing the transition costs and the job-extension penalties associated with rescheduling a job. Our work focuses in determining the best possible schedule given these costs.

C. Our Contribution

Our study includes both theoretical and comprehensive experimental results. We consider reoptimization with respect to the two classical objectives of minimum makespan and minimum total flow-time, and distinguish between instances with unlimited and limited budget. Our results are presented in Table I. For completeness, we include in the table previous results regarding the minimum total flow-time with unlimited budget [2].

We first analyze the computational complexity status of these problems. While the hardness result for the minimum makespan problem is straightforward, the hardness for the minimum total flow-time problem with limited budget is complex and a bit surprising - given that the minimum flow-time problem is solvable even on unrelated machines

[6], [16], and that the dual variant of achieving an optimal reschedule using minimum budget is also solvable [2]. Our hardness results are valid already for very restricted classes, with a single added machine and no job-extension penalties.

In order to be able to evaluate our heuristics against an optimal solution, we develop and implement an optimal solver based on *branch and bound* technique. Naturally, the solver could not handle very large instances, but we were able to run it against small but diverse instances to compare the different heuristics to the optimum. For example, problems instances with 4 machines and 20 jobs were easily solved.

We then present several heuristics for each objective function. Some of the heuristics distinguish between modifications that involve addition or removal of machines. For both objectives we also developed and applied a genetic algorithm [9], [19]. All the heuristics were implemented, their performance on various classes of instances have been compared, and the results were analyzed. Our experimental study concludes the paper.

II. Computational Complexity

In this section we analyze the computational complexity of reoptimization scheduling problems. We distinguish between the minimum makespan and the minimum total flow problems, as well as between the problem of finding an optimal solution using minimum budget and finding the best solution that can be obtained using limited budget. Due to space constraints, some of the proofs in this section are omitted.

We use the following notations: For a multiset $A = \{a_1, a_2, \ldots, a_{|A|}\}$ of integers, let $MAX(A) = \max_{j=1}^{|A|} a_j$ and $SUM(A) = \sum_{j=1}^{|A|} a_j$. Also, let $\vec{A}$ denote the vector consisting of the elements of A in non-decreasing order and define $SPT(A) = \vec{A} \cdot (|A|, \ldots, 2, 1)$. For example, for $A = \{5, 3, 1, 5, 8\}$ it holds that $SUM(A) = 22$, $\vec{A} = (1, 3, 5, 5, 8)$ and $SPT(A) = (1, 3, 5, 5, 8) \cdot (5, 4, 3, 2, 1) = 5 + 12 + 15 + 10 + 8 = 50$. Note that $SPT(A)$ is the value of an optimal solution for the minimum total-flow problem on a single machine for an instance consisting of $|A|$ jobs with lengths in A.

For the minimum makespan reoptimization problem, our result is not surprising - the classical load-balancing problem $P||C_{max}$ is known to be NP-complete even with no transition costs or extensions. For completeness we show that this hardness carry over to the simplest class of the reoptimization variant.

Theorem 2.1: The minimum makespan reoptimization problem is NP-complete even with a single added machine, unlimited budget, and no job-extension penalty.

The analysis of the minimum total flow problem is more involved. The corresponding classical optimization problem $P||\sum C_j$ is known to be solvable in polynomial time by the simple SPT algorithm. For the reoptimization problem, an efficient optimal algorithm for finding an optimal solution using minimum budget is presented in [2]. The algorithm is based on a reduction to a minimum weighted perfect matching in a bipartite graph. This reduction cannot be generalized to consider instances with limited budget, and the complexity status of the problem of finding the best solution that can be obtained using limited budget remains open in [2]. We show that this problem is NP-complete, even with no job-extension penalties and a single added machine.

Our proof refers to the compact representation of the problem. In a compact representation, the jobs assigned on each machine are given by a set of pairs $\langle p_j, n_j \rangle$, where n_j is the number of jobs of length p_j assigned on the machine. We first prove two simple observations:

Observation 2.2: Let A' be a subset of A then $SPT(A') + SPT(A \setminus A') < SPT(A)$.

Given a multiset A, and an integer Z, let $x > MAX(A)$. Extend A to a multiset A^* by adding to it Z elements of value x.

Observation 2.3: $SPT(A^*) = \frac{Z(Z+1)}{2}x + Z \cdot SUM(A) + SPT(A)$.

Using the above observations, we are now ready to prove the hardness result.

Theorem 2.4: The minimum total flow-time reoptimization problem with limited budget is NP-complete even with a single added machine, and no job-extension penalty.

Proof: The reduction is from the *Partition* problem: given a set $A = \{a_1, a_2, \ldots, a_n\}$ of integers, whose total sum is $2B$, the goal is to decide whether A has a subset $A' \subset A$ such that $SUM(A') = SUM(A \setminus A') = B$. The Partition problem is known to be NP-complete [12].

Given an instance of Partition $A = \{a_1, a_2, \ldots, a_n\}$, whose total sum is $2B$, let $Z = SPT(A)$. We construct the following instance for the reoptimzation problem: The initial schedule, S_0, consists of a single machine and $n+Z$ jobs. The first n jobs correspond to the Partition elements, that is, for $1 \leq j \leq n$ let $p_j = a_j$. Each of the additional Z *dummy jobs* have length x for some $x > B$. Assume that one machine is added, and that the transition cost of migrating job j from the initial machine to the new machine is p_j. Assume also that the budget is B.

Since the budget is B and each dummy jobs have length more than B, none of these jobs can be migrated to the new machine. Thus, a modified schedule S is characterized by a subset $A' \subset A$ of jobs corresponding to the partition elements, whose total length is at most B. These jobs are migrated to the new machine and assigned in SPT order on it. The remaining jobs are be assigned in SPT order on the initial machine. since x is larger than any a_i, the Z jobs of length x will be assigned after the jobs corresponding to $A \setminus A'$.

Finally, we note that the reduction is polynomial. Calculating $SPT(A)$ requires sorting and is performed in

TABLE I
SUMMARY OF OUR RESULTS.

Objective Function	Optimal Reschedule Using Minimum Budget			Best Reschedule Using Given Limited Budget		
	NP-hard	Optimal Solution	Heuristics	NP-Hard	Optimal Solution	Heuristics
$\min C_{max}$	Yes	Branch and Bound	• LPT-based greedy • Loads-based greedy • Genetic algorithm	Yes	Branch and Bound	• LPT-based greedy • Loads-based greedy • Genetic algorithm
$\min \sum_j C_j$	No [2]	Reduction to min-weight perfect matching [2]	–	Yes	Branch and Bound	• Greedy reversion • Cyclic reversion • SPT-like

time $O(n \log n)$. The instance constructed in the reduction consists of $n + Z$ jobs where $Z = SPT(A)$. The compact representation of this instance includes at most $n + 1$ different pairs, where all $\langle p_j, n_j \rangle$ values are polynomial in n.

Claim 2.5: The minimum total flow-time in an optimal modified schedule is less than $\frac{Z(Z+1)}{2}x + (B+1)Z$ if and only if a partition of A exists.

The above claim completes the hardness proof. ∎

Remark: It is possible that two multisets A, B will have the same cardinality, and that $SUM(A) > SUM(B)$ while $SPT(A) < SPT(B)$. For example, for $A = \{1, 2, 10\}$ and $B = \{3, 4, 5\}$, we have $|A| = |B| = 3$, $SUM(A) = 13 > 12 = SUM(B)$ while $SPT(A) = 15 < 24 = SPT(B)$. This emphasis an additional challenge of the reoptimization problem: an optimal solution may not use the whole budget. Such anomalies also explains why our hardness proof cannot be a simple reduction from the subset-sum problem, and the dummy jobs are required.

III. Optimal Algorithms

A. A Brute-force Solver based on Branch and Bound

Our brute-force solver was designed to utilize high performance multi-core machines in order to find optimal solutions for the problems that were shown to be NP-complete. The solution space for a scheduling problem can be described by a tree of depth n, where depth k corresponds to the assignment of job k, for $1 \leq k \leq n$. Specifically, the root (depth 0) corresponds to an empty schedule - none of the jobs were assigned; at level 1 there are m nodes, representing job 1 being assigned to each of the m machines. At level k there are m^k nodes, corresponding to all possible assignment of the first k jobs. This implies that the brute-force solver may need to consider m^n assignments find the optimal one.

We note that, as detailed in Section I-A, once the partition of jobs among the machines is determined, the internal job order on each machine either has no effect on the solution (in the $\min C_{max}$ problem), or is the unique SPT-order (in the $\min \sum_j C_j$ problem). Thus, the solution space need not distinguish between assignments with different internal order of the same set of jobs on every machine.

Obviously, even without considering different internal orders, iterating over all of the m^n configurations is not feasible when dealing with large instances. Our solver uses a *branch and bound* technique combined with other optimizations to effectively trim tree-branches that are guaranteed not to yield an optimal solution.

In particular, the solver keeps in memory the best solution it found so far (its objective value and its transition cost). When processing a tree node if the already accumulated transition cost is larger than the budget or if the objective-function value is larger than the current best, then the solver can safely discard this tree branch as it is guaranteed not to yield a feasible optimal solution. For partial assignments, the objective-value is calculated by combining the value (makespan or total flow-time) of the already assigned jobs, and a lower bound on the yet-to-be-assigned jobs. For the minimum makespan problem, the lower bound is calculated by assuming perfect load-balancing ($\sum_j p_j/m$), and for total flow-time the lower bound is calculated by assuming SPT-order with no job-extension penalties.

In addition, we find out that considering the jobs from longest to shortest, that is, depth 1 corresponds to the longest job in the initial assignment, etc.) drastically helps in trimming branches earlier in the process.

The solver was designed to use multi-core machines in order to shorten the run time, by doing the work in parallel on the different cores. In the heart of the design stands a concurrent queue to which tasks are enqueued. Different threads concurrently dequeue these tasks, in a consumer-producer like mechanism. A 'task' for that matter is a request to process a tree node. That is, when the solver starts the queue is empty and a task to process the root node is added, which ignites the process. The solver is done when the queue is empty.

The solver's ability to solve problems instances of different sizes is determined by the given machine, to be more

specific, by the CPU's clock speed, the number of available cores and sufficient memory (as the entire process is in memory). For example, the solver was able to handle a problem with 9 machines and 20 jobs in about 100 minutes, when ran on a machine with 8 cores and 32GB of RAM memory.

B. An Optimal Algorithm for $\Sigma_j C_j$

An algorithm for finding an optimal reschedule with respect to the minimum total flow-time objective is presented in [2]. The algorithm returns an optimal modified schedule using the minimal possible budget. The algorithm is based on reducing the problem to a minimum-weight complete-matching problem in a bipartite graph. The algorithm fits the most general case - arbitrary modifications, arbitrary transition costs and arbitrary job-extension penalties.

We have implemented this algorithm, and use its results as a benchmark so we can evaluate how well our heuristics perform. The algorithm is based on matching the jobs with possible slots on the machines. For completeness, we give here the technical details that are relevant to its implementation. Recall that n and m represent, respectively, the number of jobs and machines in the modified instance. Let $G = (V, E)$, where $V = J \cup U$. The vertices J correspond to the set of n jobs (a single node per job). The set U consists of mn nodes, q_{ik}, for $i = 1, \ldots, m$ and $k = 1, \ldots, n$, where node q_{ik} represents the k^{th} from last position on machine i. The edge set E includes an edge (v_j, q_{ik}) for every node in J and every node in U (a complete bipartite graph). The edge weights consist of two components: a dominant component corresponding to the contribution of a job assigned in a specific position to the total flow-time, and a minor component corresponding to the associated transition cost. Both components are combined to form a single weight. Formally, for a large constant Z,

- For every job that is assigned to i in S_0, let $w(v_j, q_{ik}) = Zkp_j$.
- For every $i' \neq i$, let $w(v_j, q_{i'k}) = Zk(p_j + \delta_{i,i',j}) + \theta_{i,i',j}$.

These weights are based on the observation that a job assigned to the k-th from last position, contributes k times its processing-time to the total flow-time (see details in [2]). We implemented the algorithm by using the Hungarian method [17], a combinatorial optimization algorithm that solves the assignment problem in polynomial time. The solver's run time is $O(|V|^3)$, where $|V| = n(m + 1)$. In practice, this optimal solver can easily handle instances with 30 machines and 300 jobs.

IV. Our Heuristics

In this section we describe the heuristics we have designed and implemented. Some heuristics were designed for specific modification (e.g. machines removal, limited budget), or for specific objective function, while some are general and fit all our reoptimization variants.

A. Heuristics for Minimum Makespan

We suggest two greedy heuristics, both intended to solve the *minimal makespan* problem ($\min C_{max}$). In the first, we select the next migration to be performed according to the job's processing times, while in the second, we select the next migration according to the loads on the machines. Both algorithms begin with S_0 as the initial configuration. If the modification involves machines' removal, we first perform migrations of jobs assigned to the removed machines and migrate each such job j assigned to a removed machine i, to a machine $i' \in M$ for which $\theta_{i,i',j}$ is minimal. Ties are broken in favor of short extension penalty. As mentioned in the introduction, we assume that the budget is sufficient for this reschedule, as otherwise no feasible solution exists. Following the above preprocessing, we perform the following:

1. LPT-Based: In every iteration we consider the jobs in non-increasing processing-time order. When considering job j, we check whether migrating it to one of the two least loaded machines increases the load-balancing, formally, assume j is assigned to machine i, and we consider moving it to machine i', we check whether $p_j + \theta_{i,i',j} + L_{i'} < L_i$. If the answer is positive and the remaining budget allows, the migration is performed. We repeat the iterations until a complete pass over the jobs yields no migration.

2. Loads-Based: In every iteration we try to migrate some job out of the most loaded machine. We first consider the pair of most-loaded and least-loaded machines. Denote these machines by i and i'. We consider jobs on machine i according to order $\theta_{i,i',1} \leq \theta_{i,i',2} \leq \ldots$. When considering job j we check whether migrating it to machine i' increases the load-balancing, that is, $p_j + \theta_{i,i',j} + L_{i'} < L_i$. If the answer is positive and the remaining budget allows, the migration is performed, and a new iteration begins (maybe with a different pair of most- and least-loaded machines). If the answer is negative for all the jobs on i, we move to the next candidate for target machine i' - the second least-loaded machine, etc. The iteration ends when some beneficial migration from the most-loaded machine is detected. If none such migration exists, the algorithm terminates.

B. Heuristics for Minimum Total Flow-time

The minimum total flow-time reoptimization problem can be solved optimally assuming unlimited budged. While the optimal algorithm presented in Section III-B solves optimally the $\Sigma_j C_j$ problem using the minimal possible budget, it cannot be modified to solve the problem when the budget is limited. In fact, as shown in Section II, this variant is NP-hard. We propose two heuristics that use the optimal algorithm as a first step and then each, in its own

way, change the assignment to reach a feasible solution which obeys the budget constraints. A third algorithm that we propose, tries to reach an SPT-like schedule.

1. Greedy Reversion: The optimal algorithm returns an assignment S minimizing the total flow-time, which might not conform to the budget limitation. The following steps are performed to reach a feasible solution. First, we sort all the jobs which migrated in the transition from S_0 to S in non-increasing order according to the transition cost their migration caused.

We then distinguish between two cases:

1) The modification consists of only machines' *addition*. We revert the transitions one by one until we reach an allowed budget.
2) The modification includes machines' *removal*. We revert the transitions of jobs which do not originate from a removed machine, one by one until we reach an allowed budget. If after all possible reverts were done, the budget is still not met, we continue to the next step: 'Handling jobs of removed machines'.

Handling jobs of removed machines: This step is performed only when removed machines are involved, and all the jobs assigned to remaining machines are back in their initial machines. We sort the jobs originated from removed machines in non-increasing order according to the transition cost their migration (determined by the optimal algorithm) caused. Job after job, we migrate a job j assigned in S_0 to a removed machine $i \in M_0 \setminus M$ to the machine $i' \in M$ for which $\theta_{i,i',j}$ is minimal, breaking ties in favor of better objective value. As explained in the introduction, we assume that the budget is sufficient to complete all these migrations.

2. Cyclic Reversion: The optimal algorithm returns an assignment S minimizing the total flow-time, which might not conform to the budget limitation. Similar to the previous heuristic, we choose the most expensive transition involved, denote by j the corresponding job. We migrate job j back to its origin machine, $M_{0,j}$. Since we wish to keep jobs distributed as evenly as possible, we now choose a job that migrated to $M_{0,j}$ and migrate it back to its initial machine. We choose the job whose migration to $M_{0,j}$ was most expensive. We keep these cyclic reverts until one of following conditions holds: a) We made a complete loop and reached back the machine from which we started. b) We have reached a machine to which no job was migrated. If the budget conforms to the limitation, we stop, Otherwise we choose a job with the most expensive transition cost and start a new revert cycle.

If the modification includes machines' *removal*, jobs originated from the removed machines cannot be selected to migrate back to their original machine. If the budget is not met after all the allowed reverts were performed, we continue to the step 'Handling jobs of removed machines', as described in the greedy heuristic.

3. SPT-like: It is known that SPT ordering is optimal for $P||\Sigma_j C_j$. We therefore try to reach a schedule that fulfils the following basic properties of an SPT schedule:

1) In any optimal schedule the number of jobs on any machine is either $\lfloor \frac{n}{m} \rfloor$ or $\lceil \frac{n}{m} \rceil$.
2) The jobs can be partitioned into $\lfloor \frac{n}{m} \rfloor$ rounds, such that the k-th round consists of all the jobs that are k from last on some machine. In an SPT schedule, each of the jobs in the k-th round is not shorter than each of the jobs in the $k+1$st round.

This heuristic consists of three stages. The first stage, applied when machines were removed, is to move to some feasible schedule - each of the jobs assigned to a removed machine, is migrated into a machine for which the transition cost is the lowest.

In the second stage, we try to make the machines as balanced as possible in terms of number of jobs. While the budget allows and while there exists a pair of machines, one with more than $\lceil \frac{n}{m'} \rceil$ jobs and the other with less than $\lfloor \frac{n}{m'} \rfloor$ jobs, we migrate the cheapest-to-move job on the first machine, to the second one.

In the third phase, we try to make our solution as close to the SPT ordering as possible, we compare the solution round after round to the desired SPT ordering, and switch jobs whenever required, as long as the budget allows.

C. Genetic algorithm

In the Genetic algorithm the idea is for the solution to be obtained in a way that resembles a biological evolution process [9], [19]. That is, we let the method's evolutionary process find the solution by itself. The idea is to define the *Genome* of a single solution and a *Ranking* method $Rank : Genome \rightarrow \mathbb{R}$. The genome of a single solution represents and gives all the needed information regarding the solution. In our case the Genome is simple, for a problem instance with m machines and n jobs, a solution genome id, g, defined as $g = (g_1, g_2, ..., g_n)$, where $g_i \in [1, m]$ and $g_i \notin \{x|\text{x is a removed machine id}\}$, in other words each cell with index i represents the $i-th$ jobs and the cell's value is the machine this job is assigned to in the modified schedule. The ranking method is used to define how good is a given solution. In our case the ranking method sorts the different genomes first by the objective method value (C_{max} or $\Sigma_j C_j$) and then by the transition cost. We create a $population$ (generation 1), which is a collection of $genomes$, we rank each member of the population and sort them from best to worst.

When solving a reoptimization problem with limited budget, to guarantee the algorithm end up with a feasible solution, we create at least one feasible genome in generation1. In the case of 'machines addition', this solution will be S_0 as it is both valid and requires no transition cost. In the case of 'machines removal', a job j assigned in S_0 to a removed machine $i \in M_0 \setminus M$ is assigned in the feasible

genome to the machine $i' \in M$ for which $\theta_{i,i',j}$ is minimal. We assume that the budget is sufficient for this reschedule, as otherwise no feasible solution exists.

The next step is the evolutionary-like step, in which we create the next generation according to the following methods:

1) Elitism mechanism: We take the best 5% genomes and move them 'as-is' to the next generation. This guarantee that the next generation will be at least as good as the current one. To deal with the case of limited budget, we also pass 'as-is' the best solution which meets the given budget. This must be done since the genetic process is pushing the genomes population for a better objective values as a primary goal and to minimize the transition cost as a secondary goal.
2) Cross over: From the entire genomes population, we choose randomly 2 elements, denoted g_x and g_y, we choose a pivot point from the range of $ind \in [1, n-1]$, and create two new items:

$$newItem1_i = \begin{cases} g_{x_i} & \text{if } i \leq ind \\ g_{y_i} & \text{if } i > ind \end{cases}$$

and

$$newItem2_i = \begin{cases} g_{y_i} & \text{if } i \leq ind \\ g_{x_i} & \text{if } i > ind \end{cases}$$

43% of the next generation is a result of this operator.
3) Mutate: We choose a random genome, we choose from its genome a random cell and change its value. 43% of the next generation is a result of this operator.
4) Fresh Items: We generate new genomes. These new elements have the potential of shifting the results in a new direction and to help avoid local optimum. 9% of the next generation is be a result of this operator.

Each genome in the newly created population is then re-evaluated, meaning, its score is computed. The process repeats itself until it fails to improve any further and the genome with the best ranking is selected as output from the most recent generation. Obviously if we examine the best solution from generation $i+1$ compared to that of generation i we will notice that the 'quality' of the best solution is non decreasing over the generations as we have the 'Elitism mechanism' which ensures us that the best individuals will survive to the next generation.

V. Experimental results

The datasets for our experiments were created using our own data generator which supports any parameters combination of number of machines and jobs, number of added or removed machines, number of added or removed jobs, as well as the distributions of job lengths, job-extension penalty, and transition costs.

Instances on which we run our heuristics could be very large (hundreds of jobs, and several dozens of machines). Instances on which we run our brute-force solver (introduced in Section III-A) had to be smaller. In particular, we ran the brute-force solver on instances with 20 jobs and 4 machines. We find out that even such small instances can provide a good comparison between different heuristics; therefore, the brute-force solver is helpful for concluding how far from the optimum our heuristics perform. The optimal algorithm for $\min \Sigma_j C_j$, based on a reduction to perfect matching (introduced in Section III-B), was able to handle relatively large instances of 15 machines and 300 jobs. In our basic template for job creation, the jobs' lengths were uniformly distributed in $[1, 20]$. The job-extension penalty of job j was uniformly distributed in $[1, \frac{p_j}{2}]$, and the transition costs were uniformly distributed in $[1, 5]$.

To generate problems instances for a specific experiment we took a template instance, decided on one parameter that will vary in the different runs, and set the rest of the parameters to basic values. For example, to understand how the number of added machines affects the heuristics performance, we fixed all the other parameters (jobs' lengths, transition costs, etc.), and run the heuristics on multiple instances which vary only in the number of added machines. To avoid anomalies, we have generated for each experiment 5 instances with the same parameters, based on 5 templates instances (same configuration, different instance) and considered the average performance as the result.

Another parameter that could affect the performance of our heuristics is the initial assignment - which may be random, SPT or LPT. We found out that in practice the initial assignment does not affect the results significantly and the results we present in the sequel were all generated with a randomize initial assignment - which is a bit more 'challenging' and therefore emphasizes the differences among the heuristics.

The results of our experiments are presented and analyzed below. In all the figures, the bars show the objective value ($\Sigma_j C_j$ or C_{max}), and the lines show the corresponding transition cost.

A. Results for the Minimal Total Flow-Time problem

1) Machines' Addition: The template for heuristics that analyze the $\min \Sigma_j C_j$ problem consists of 15 machines and 300 jobs. We start by showing how the different heuristics performs on instances with both transition costs and job-extension penalties, where the number of added machines was set to $m/2$. As shown in Fig. 1, with unlimited budget the genetic algorithm is the closest to the known optimum, calculated by the perfect matching algorithm result. As budget is limited, the performance of the genetic algorithm drops. As expected, the lower the budget, the higher the objective value. Also, the differences

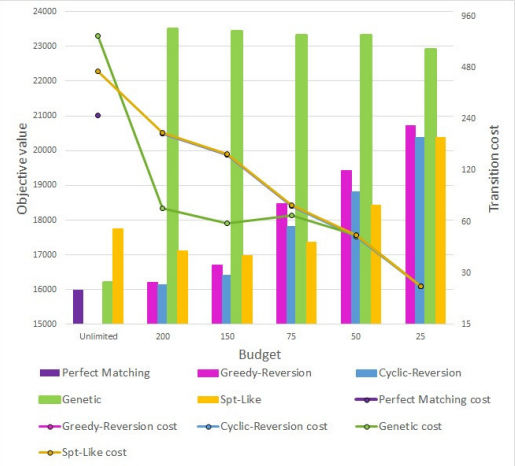


Fig. 1. Results for $\min \Sigma_j C_j$ with $m/2$ added machines and variable budget.

between the heuristics are less significant as the budget decreases, as less transitions are allowed. Interesting fact is that the genetic algorithm has a slight improvement as the limitation is getting stricter. We explain this by the fact that before starting the algorithm we include in the population items that obey the allowed budget. Later, these items are influencing the genetic process and are helping the algorithm to converge to a good solution, better than with a more relaxed budget limitation.

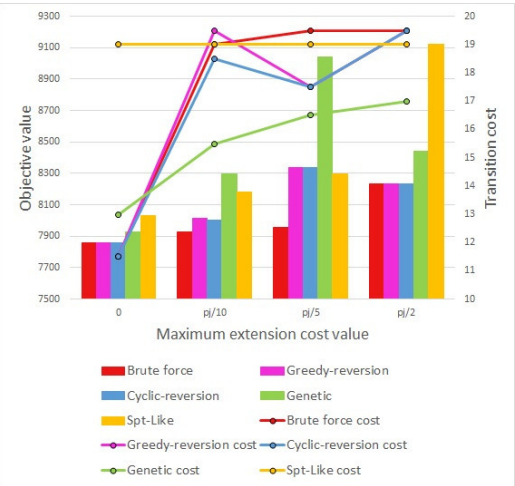


Fig. 2. Results for $\min \Sigma_j C_j$ with variable extension penalty.

The goal of our next experiment was to see how close the heuristics get to the actual optimum. For this test we have used our brute-force solver on relatively small problem instances (4 machines and 20 jobs), two machines were added and the budget was set to 20. The results for various extension penalties are shown in Fig. 2. We observe that both 'Greedy-reversion' and 'Cyclic-reversion' heuristics were very close to the optimum.

2) Machines' Removal: Fig. 3 presents the performance of the different heuristics when the modification is machines' removal and the budget is limited to 150. According to our parameters, this budget is expected to be sufficient for the migration of about 20% of the jobs. The initial assignment was of 300 jobs on 15 machines. Not surprisingly, we see an increase of the objective value as the number of removed machines increases. All of the heuristics performed more or less the same, both in terms of the achieved objective value and in term of the budget utilization, with an exception of the Genetic algorithm which manage to use a significantly lower budget.

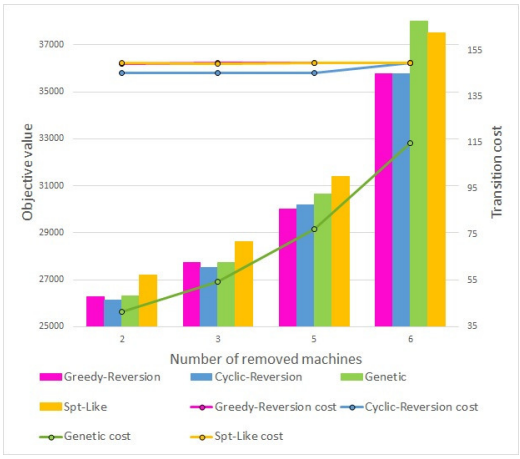


Fig. 3. Results for $\Sigma_j C_j$ for machines' removal and limited budget.

Fig. 4 presents the results for the same instance and the same modification only with unlimited budget. This problem is the one for which we have an efficient optimal algorithm (see Section III-B). The genetic algorithm perform very close to the optimum for every number of removed machines. In fact, for two removed machines its transition cost is lower than the optimum and only slightly higher in the total flow-time (recall that the optimal algorithm 'insists' on finding a reschedule that minimizes the total flow-time). On the other hand the SPT-Like heuristic is both very costly and gives poor results. This can be explained by the fact that insisting on a complete SPT order requires many transition, and involves many job-extension penalties.

B. Results for the Minimum Makespan problem

1) Machines' Addition: Our template for experiments analyzing the $\min C_{max}$ problem consists of 30 machines and 500 jobs. Fig. 5 presents results for adding 15 machines and variable budget. The 'Loads-based' heuristic is the best heuristic. The genetic algorithm perform poorly compared to the other heuristics, but on the other hand, it does not utilize the whole budget.

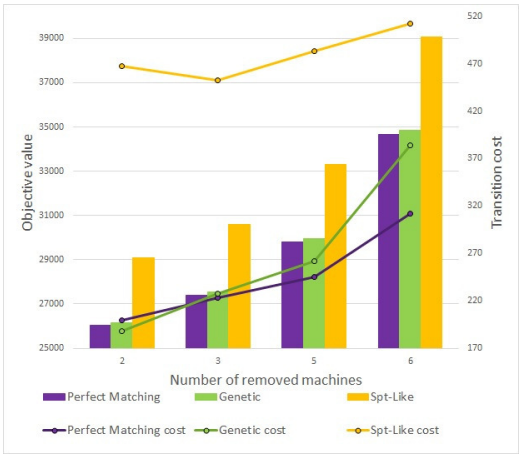


Fig. 4. Results for $\Sigma_j C_j$ for machines' removal and unlimited budget.

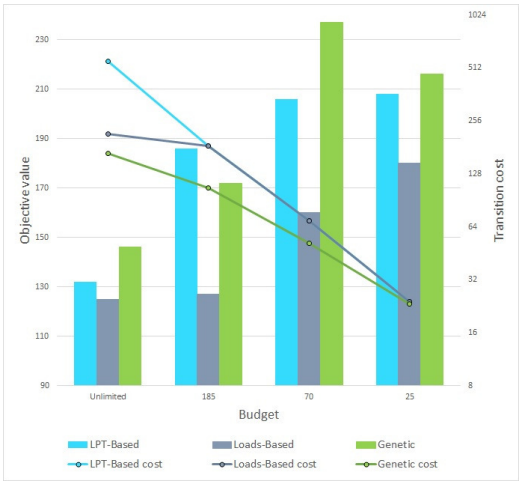


Fig. 5. Results for $\min C_{max}$ with $m/2$ added machines and variable budget.

In the our next experiment we measure how close the heuristics get to the optimum. We have used our brute-force solver on relatively small problem instances, consisting of 4 machines, 20 jobs, and a limiting budget of 20. The results for various extension penalties are shown in Fig. 6. Once again, the 'Loads-based' heuristic outperform the others, and is relatively close to the optimum. The 'LPT-based' heuristic seems to perform (relatively) better as the job-extension penalty increases.

2) Removing machines: Our next experiments compare the performance of the different heuristics when the modification is machines' removal. We performed two experiments - with budget limited to 250 and with unlimited budget. Due to space constraints, the figures are omitted. Our results show that with unlimited budget, all three heuristics (LPT-based, Loads-based and genetic) perform more or less the same. While a similar makespan is achieved, the

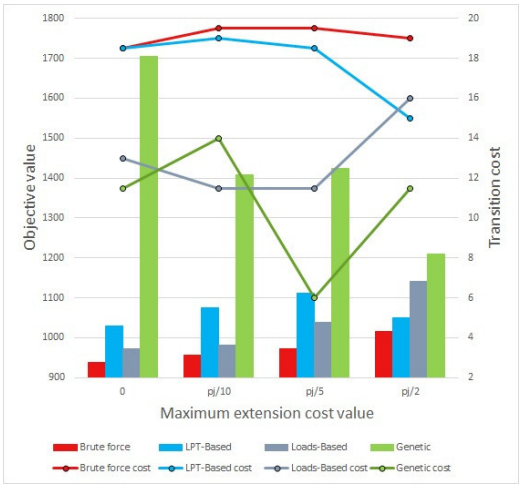


Fig. 6. Results for $\min C_{max}$ with variable extension penalty.

Loads-based heuristic requires the lower transition cost, then the LPT-based (that needs 10% higher cost) and the genetic ($15-20\%$ higher than Loads-based). With limited budget, the Loads-based and LPT-based heuristics perform significantly better than the genetic algorithm, but they also require a much higher budget.

VI. Conclusions and Future Work

We presented theoretical and experimental results for the reoptimization variant of job scheduling problems. We have shown that the problems of finding the minimum total flow-time with limited budget, finding the minimum makespan with limited budget and finding the minimum makespan with unlimited budget are NP-Complete. We have designed and implemented several heuristics for each of these problems, and performed a comprehensive empirical study to analyze their performance. To see how well these heuristics perform compared to the actual optimum, an efficient branch-and-bound brute-force solver was designed and implemented. An optimal algorithm for the minimum total-flow problem with unlimited budget was also implemented.

In general, our experiments reveal that while the problems are NP-hard, heuristics whose time complexity is polynomial in the number of jobs and machines, perform very well in practice. Simple algorithms, that are based on adjustment of known heuristics for the one-shot problem (with no modifications) are both simple to implement and provide results that are, on average, within $10\%-15\%$ from the optimum. More complex algorithms, that are based on a preprocessing in which a perfect matching algorithm is implemented, perform on average even better.

We have observed that while the Genetic algorithm does not perform well when given a limited budget, it performs relatively well with unlimited budget for the

minimum C_{max} problem, and close to optimum for $\Sigma_j C_j$. It also takes a considerable amount of time to run. In some scenarios, its objective value may not be competitive compared to other heuristics, however, its budget utilization is impressively good (see for example Fig. 3). A known issue of genetic algorithm is that the parameters must be carefully tuned in order for the algorithm to converge into a good solution. Future work on this algorithm might refactor the population size or operators we have used (Elitism, Crossover, Mutation) by adding new operators, modify the existing or change the possibilities of each to create a more optimized genetic algorithm. Another direction to explore is to create a dedicated score method for each variant of the problems. For real life applications where budget utilization is a big concern, we are sometimes allowed to use lot of budget but strive to use as less as possible. The genetic algorithm is a good choice for such scenarios.

Our greedy heuristics performed well, both on the $\Sigma_j C_j$ and the C_{max} problems. We have observed that the 'Loads-based' heuristic outperformed the 'LPT-based' both in terms of objective value and budget utilization. With unlimited budget, the budget utilization difference was more significant. For $\min \Sigma_j C_j$, the 'Cyclic-reversion' showed better performance compared to the 'Greedy-reversion', This result does not surprise us as a more balanced solution is expected to yield better results for the minimum total flow-time problem. In terms of budget utilization, it was expected that this greedy, budget oriented method will utilize as mush budget as possible.

An additional direction for future work is to develop algorithms for the minimum makespan problem with a guaranteed approximation ratio. While tuning existing approximation-algorithm for the classical one-shot problem seems to be a promising direction, the presence of transition-costs and job-extension penalties give rise to new challenges and considerations. Finally, it would be interesting to consider different objective functions, or different scheduling environments, for example, jobs with deadlines or precedence constraints, as well as unrelated or restricted machines.

REFERENCES

[1] G. Ausiello, B. Escoffier, J. Monnot, and V. Th. Paschos. Reoptimization of minimum and maximum traveling salesmans tours. *J. of Discrete Algorithms* 7(4):453–463, 2009.

[2] G. Baram and T. Tamir, Reoptimization of the Minimum Total Flow-Time Scheduling Problem. *Sustainable Computing, Informatics and Systems*. vol. 4(4):241–251, 2014.

[3] J. Berlinskaa and M. Drozdowskib. Scheduling divisible MapReduce computations. In *Journal of Parallel and Distributed Computing*. vol.71(3):450-459, 2011.

[4] A. D. Birrell and B. J. Nelson. Implementing remote procedure calls. *ACM Transactions on Computer Systems* 2:39–59 ,1984.

[5] H. J. Bockenhauer, L. Forlizzi, J. Hromkovic, J. Kneis, J. Kupke, G. Proietti, and P. Widmayer. On the approximability of TSP on local modifications of optimally solved instances. *Algorithmic Operations Research* 2(2), 2007.

[6] J. L. Bruno, E.G Coffman, and R. Sethi. Scheduling independent tasks to reduce mean finishing time. *Communications of the ACM*, 17:382–387, 1974.

[7] C. Clark, K. Fraser, S.Hand, J.G. Hansen, E. Jul, C. Limpach, I. Pratt, A. Warfield. Live migration of virtual machines. The 2nd *Symp. on Networked Systems Design and Implementation* (NSDI). 2005.

[8] R.W. Conway, W.L. Maxwell, and L.W. Miller. Theory of Scheduling. *AddisonWesley*, 1967.

[9] A. E. Eiben and J. E. Smith. Introduction to Evolutionary Computing. *Springer*, 2007

[10] D. Eppstein, Z. Galil, and G. F. Italiano. Dynamic graph algorithms, Chapter 8. In *CRC Handbook of Algorithms and Theory of Computation*, ed. M. J. Atallah, 1999.

[11] B. Escoffier, M. Milanic, and V. Th. Paschos. Simple and fast reoptimizations for the Steiner tree problem. *DIMACS Technical Report* 2007-01.

[12] M. R. Garey and D.S. Johnson. *Computers and Intractability: a guide to the theory of NP-completeness*. W. H. Freeman and Co., New York, 1979.

[13] R. L. Graham, E.L. Lawler, J. K. Lenstra, A. H. G. Rinnooy Kan, Optimization and approximation in deterministic sequencing and scheduling: a survey, *Ann. Discrete Math*. vol. 5:287–326, 1979.

[14] S. Hacking and B. Hudzia. Improving the live migration process of large enterprise applications. The 3rd *international workshop on Virtualization technologies in distributed computing* (VTDC), 2009.

[15] D. S. Hochbaum and D. B. Shmoys. Using dual approximation algorithms for scheduling problems: Practical and theoretical results. *Journal of the ACM*, 34(1):144–162, 1987.

[16] W. Horn. Minimizing average flow-time with parallel machines. *Operations Research*, 21:846–847, 1973.

[17] Harold W. Kuhn. The Hungarian Method for the assignment problem, *Naval Research Logistics Quarterly*, vol. 2: 83–97, 1955.

[18] Parallels Virtuozzo *http://www.parallels.com/products/pvc*.

[19] L. M. Schmitt. Theory of Genetic Algorithms, *Theoretical Computer Science* vol. 259:1-61, 2001.

[20] H. Shachnai, G. Tamir, and T. Tamir, Minimal cost reconfiguration of data placement in storage area network. *Theoretical Computer Science*. vol. 460:42–53, 2012.

[21] H. Shachnai, G. Tamir, and T. Tamir. A theory and algorithms for combinatorial reoptimization. *In Proc. of 10th LATIN*, 2012.

[22] W.E. Smith. Various optimizers for single-stage production. *Naval Research Logistics Quarterly*. vol. 3:59–66, 1956.

[23] M. Thorup, and D.R. Karger. Dynamic graph algorithms with applications. *In Proc. of 7th SWAT*, 2000.

[24] Xen Project, *http://www.xenproject.org/*.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 571–575
DOI: 10.15439/2016F206
ACSIS, Vol. 8. ISSN 2300-5963

Evaluation of selected fuzzy particle swarm optimization algorithms

Tomasz Krzeszowski, Krzysztof Wiktorowicz
Rzeszow University of Technology
Faculty of Electrical and Computer Engineering
al. Powstanców Warszawy 12, 35-959 Rzeszów
Email: {tkrzeszo, kwiktor}@prz.edu.pl

***Abstract*—This paper is devoted to an evaluation of selected fuzzy particle swarm optimization algorithms. Two non-fuzzy and four fuzzy algorithms are considered. The Takagi-Sugeno fuzzy system is utilized to change the parameters of these algorithms. A modified fuzzy particle swarm optimization method is proposed, in which each of the particles has its own inertia weight and coefficients of the cognitive and social components. The evaluation is based on the common nonlinear benchmark functions used for testing optimization methods. The ratings of the algorithms are assigned on the basis of the mean of the objective function and the relative success.**

I. Introduction

PARTICLE swarm optimization (PSO) is a stochastic optimization technique that was developed by Kennedy and Eberhart [1]. The PSO is mainly inspired by the social behavior of organisms that live and interact within large groups, for example, schools of fish, flocks of birds or swarms of bees. The usefulness of PSO in solving a wide range of optimization problems has been repeatedly confirmed. It has been applied to: the intelligent identification and control of a dynamic system [2]; solving an economic dispatch problem in power systems [3]; human motion tracking [4]; feature selection [5]; automatic incident detection [6]; fuzzy anomaly detection in networks [7]; the estimation of hurdles clearance parameters [8] and many more problems. Many variants of the PSO have been developed since it was introduced in 1995 [1]. The most common are algorithms with a constriction factor [9] and with a linear inertia weight [10]. Among the PSO modifications we can distinguish algorithms that utilize fuzzy systems [2], [3], [11]–[15]. For example, in papers [2], [11] a fuzzy system was used to dynamically modify the inertia weight. Another approach was presented in [3], where a fuzzy system is used to change the inertia weight and the coefficients of the cognitive and social components.

The goal of this study is to evaluate selected fuzzy PSO algorithms and to propose a modified fuzzy PSO algorithm. In our research, we use the Takagi-Sugeno system [16] instead of the Mamdani system [17] because it has shorter processing time. In this paper, we consider six different versions of PSO, including two non-fuzzy, and four fuzzy algorithms. The evaluation is based on nonlinear benchmarks in the form of Ackley, Griewank, Rastrigin and Rosenbrock functions. The calculations were conducted using Matlab software and the "PSO Research Toolbox" by Evers [18].

II. Particle swarm optimization

The particle swarm model consists of a group of particles that are randomly initialized in the d-dimensional search space. During an iterative process, particles explore this space effectively by exchanging information to find the optimal solution. Each i-th particle is described by its position $\mathbf{x}_i$, velocity $\mathbf{v}_i$, and best position $\mathbf{pbest}_i$. Moreover, the particles have access to the best global position $\mathbf{gbest}$ that has been found by any particle in the swarm.

In the basic PSO algorithm [1], the velocity and the position of each particle in k-th iteration are updated using the following equations:

$$\mathbf{v}_i^{k+1} = \mathbf{v}_i^k + c_1\mathbf{r}_1(\mathbf{pbest}_i^k - \mathbf{x}_i^k) + c_2\mathbf{r}_2(\mathbf{gbest}^k - \mathbf{x}_i^k) \quad (1)$$

$$\mathbf{x}_i^{k+1} = \mathbf{x}_i^k + \mathbf{v}_i^{k+1} \quad (2)$$

where $\mathbf{r}_1$, $\mathbf{r}_2$ are vectors with uniformly distributed random numbers in the interval $[0, 1]$, and c_1, c_2 are positive constants equal to 2.

The velocities of particles are determined by three components. The first component is the inertia that models the particle's tendency to continue moving in the same direction. The second component is cognitive and attracts particles towards the best position previously found by the particle. The last component is a social component that moves particles towards the best position found earlier by any particle. Selection of the best global position and the best position for i-th particle is based on the objective function (denoted later by $f(\cdot)$).

A. PSO1: Clerc, Kennedy algorithm [9]

Many approaches have been developed to improve the performance of the basic PSO algorithm. One way is to use the constriction factor χ that was proposed by Clerc and Kennedy [9]. The application of this factor controls the velocity magnitude.

The velocity equation has the form:

$$\mathbf{v}_i^{k+1} = \chi[\mathbf{v}_i^k + c_1\mathbf{r}_1(\mathbf{pbest}_i^k - \mathbf{x}_i^k) + c_2\mathbf{r}_2(\mathbf{gbest}^k - \mathbf{x}_i^k)] \quad (3)$$

where χ is calculated as $\chi = \frac{2}{|2-\varphi-\sqrt{\varphi^2-4\varphi}|}$ and $\varphi = c_1 + c_2$, $\varphi > 4$. In this paper, the following typical values are used: $c_1 = c_2 = 2.05$, $\varphi = 4.1$ and $\chi = 0.7298$.

978-83-60810-90-3/$25.00

B. PSO2: Eberhart, Shi algorithm [10]

Another way to improve the performance of PSO is to use the inertia weight ω. The inertia weight is significant for the performance of PSO, because it balances the global exploration and local exploitation abilities of the swarm. Exploration is facilitated when the inertia weight is high, but convergence is slower. On the other hand, when the inertia weight is low then convergence is faster, but it sometimes leads to local solutions. Hence, linearly decreasing inertia weight is proposed in [10].

The velocity equation has the form of

$$\mathbf{v}_i^{k+1} = \omega^k \mathbf{v}_i^k + c_1 \mathbf{r}_1 (\mathbf{pbest}_i^k - \mathbf{x}_i^k) + c_2 \mathbf{r}_2 (\mathbf{gbest}^k - \mathbf{x}_i^k) \quad (4)$$

The inertia weight ω is calculated by the formula

$$\omega^k = \omega_{max} - \frac{\omega_{max} - \omega_{min}}{iter_{max}} \cdot k \quad (5)$$

where ω_{max} is the initial weight, ω_{min} is the final weight and $iter_{max}$ is the maximum number of iterations. The limits for ω are set to $\omega_{max} = 0.9$ and $\omega_{min} = 0.4$.

III. Takagi-Sugeno system

Consider the Takagi-Sugeno (T-S) fuzzy system with two inputs y_1, y_2 and one output u. For the input y_1 we define m fuzzy sets A_i (Fig. 1), for which the vertices are placed in points p_i, where $i = 1, \ldots, m$. Similarly, for the input y_2, we define n fuzzy sets B_j with vertices in points q_j, where $j = 1, \ldots, n$. The coordinates p_i and q_j are written in the form of the vectors $\mathbf{p} = [p_i] = [p_1, \ldots, p_m]$ and $\mathbf{q} = [q_j] = [q_1, \ldots, q_n]$, respectively.

The output of the system is described by $m{\cdot}n$ fuzzy inference rules in the form of

$$R_{ij}: \text{ IF } y_1 \in A_i \text{ AND } y_2 \in B_j, \text{THEN } u = z_{ij} \quad (6)$$

where $z_{ij} \in \mathbb{R}$ is the consequent of the rule R_{ij}. The rules (6) are written in the following table:

$y_1 \backslash y_2$	B_1	$\ldots$	B_n
A_1	z_{11}	$\ldots$	z_{1n}
$\vdots$	$\vdots$	$\ddots$	$\vdots$
A_m	z_{m1}	$\ldots$	z_{mn}

(7)

The output u of the Takagi-Sugeno system is calculated as the weighted average of z_{ij} and determined by

$$u = TS(y_1, y_2) = \frac{\sum_{i=1}^{m} \sum_{j=1}^{n} w_{ij}(y_1, y_2) z_{ij}}{\sum_{i=1}^{m} \sum_{j=1}^{n} w_{ij}(y_1, y_2)} \quad (8)$$

where $w_{ij}(y_1, y_2) = A_i(y_1) \cdot B_j(y_2)$ denotes the degree of fulfillment of the rule R_{ij}.

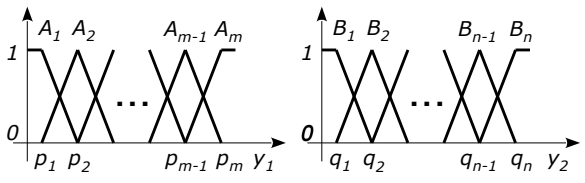


Fig. 1. Fuzzy sets for the inputs y_1 and y_2

IV. Fuzzy PSO

A. FPSO1: algorithm based on the work by Shi, Eberhart [11]

Better PSO performance can be obtained using the nonlinearly changing inertia weight that balances global and local search abilities. It is difficult to design a mathematical model to adapt the inertia weight dynamically. The solution to this problem may be obtained using a linguistic description of the search process. For example, we can use a fuzzy inference system for tuning the inertia weight [11].

In the FPSO1 algorithm, the inertia weight is described by the formula

$$\omega^{k+1} = \omega^k + \Delta\omega, \quad \Delta\omega = TS(nf^k, \omega^k) \quad (9)$$

where the T-S fuzzy system (8) is used to calculate the change of inertia weight $\Delta\omega$. The input nf^k is the normalized objective function value described by

$$nf^k = \frac{fg^k - f_{min}}{f_{max} - f_{min}} \quad (10)$$

where $fg^k = f(\mathbf{gbest}^k)$, f_{min} is the optimal solution (for the test functions considered in this paper, it is equal to 0), f_{max} is the worst solution (in our paper $f_{max} = f(\mathbf{gbest}^0)$). The fuzzy sets for the inputs nf^k and ω^k have vertices in points $\mathbf{p} = [0, 0.5, 1]$, $\mathbf{q} = [0.4, 0.7, 1]$, respectively, and the fuzzy rules have the form of

$nf^k \backslash \omega^k$	B_1	B_2	B_3
A_1	Z	N	N
A_2	P	Z	N
A_3	P	Z	N

(11)

where $N = -0.1$, $Z = 0$ and $P = 0.1$.

B. FPSO2: algorithm based on the work by Alfi, Fateh [2]

The improvement of the FPSO1 algorithm was proposed by Alfi and Fateh [2]. In their method, the inertia weight is calculated for each particle according to its current state. This is justified because each particle in the swarm is in a different place in a complex environment and may have a different balance between global and local search abilities.

In the FPSO2 algorithm, the change of inertia weight is determined by the T-S system (8) as

$$\Delta\omega_i = TS(nf_i^k, \omega_i^k) \quad (12)$$

where

$$nf_i^k = \frac{fp_i^k - f_{min}}{fp_i^0 - f_{min}} \quad (13)$$

and $fp_i^k = f(\mathbf{pbest}_i^k)$. The vertices of fuzzy sets for nf_i^k and ω_i^k are chosen as $\mathbf{p} = [0, 0.5, 1]$, $\mathbf{q} = [0.4, 0.6, 0.8]$ respectively, and the fuzzy rules are

$nf_i^k \backslash \omega_i^k$	B_1	B_2	B_3
A_1	P	N	N
A_2	P	Z	N
A_3	P	Z	N

(14)

where $N = -0.1$, $Z = 0$ and $P = 0.1$.

C. FPSO3: algorithm based on the work by Niknam [3]

In the FPSO3 algorithm, the fuzzy system proposed by Niknam [3] is used to change not only ω, but also the coefficients c_1 and c_2. These coefficients determine the influence of the personal best position $\mathbf{pbest}_i$ and the global best position $\mathbf{gbest}$ on the particle velocity. For example, if c_1 is larger than c_2, then the particle has the tendency to move to the personal best position, rather than to the global best position found by the swarm.

In the FPSO3 algorithm, three T-S systems are used to determine ω, c_1 and c_2:

$$\omega = TS(nf^k, nu^k) \tag{15}$$
$$c_1 = TS(nf^k, nu^k) \tag{16}$$
$$c_2 = TS(nf^k, nu^k) \tag{17}$$

The input nf^k is defined in (10) and nu^k is the normalized number of iterations without change of the best global position:

$$nu^k = \frac{u^k - u_{min}}{u_{max} - u_{min}} \tag{18}$$

where u^k is the number of iterations without change of the best global position, $u_{min} = 0$ and $u_{max} = iter_{max}$. The fuzzy sets are defined by the vectors $\mathbf{p} = [0.2, 0.4, 0.6, 0.8]$, $\mathbf{q} = [0.2, 0.4, 0.6, 0.8]$. The fuzzy rules for ω are defined as

$nf \setminus nu$	B_1	B_2	B_3	B_4
A_1	PS	PM	PB	PB
A_2	PM	PM	PB	PR
A_3	PB	PB	PB	PR
A_4	PB	PB	PR	PR

(19)

In table (19) we have $PS = 0.4$, $PM = 0.6$, $PB = 0.8$ and $PR = 1$. The fuzzy rules for c_1/c_2 are defined as

$nf \setminus nu$	B_1	B_2	B_3	B_4
A_1	PR/PR	PB/PB	PB/PM	PB/PM
A_2	PB/PB	PM/PM	PM/PS	PS/PS
A_3	PB/PM	PM/PM	PS/PS	PS/PS
A_4	PM/PM	PM/PS	PS/PS	PS/PS

(20)

In table (20) we have $PS = 1.2$, $PM = 1.4$, $PB = 1.6$ and $PR = 1.8$.

D. MFPSO: authors' proposition

The modified fuzzy PSO (MFPSO) algorithm proposed by the authors combines the previously described concepts developed by Alfi, Fateh [2] and Niknam [3]. In this algorithm, each of particles has its own coefficients ω, c_1 and c_2 changing according to the linguistic description represented by the fuzzy rules. In this way, each of the particles may be treated individually. For example, if a particle has found the new local best position $\mathbf{pbest}_i$, then the inertia weight ω should be decreased and the coefficients c_1 and c_2 should be increased. On the other hand, if $\mathbf{pbest}_i$ has not changed for a long time, then a better strategy would probably be to increase ω and decrease c_1 and c_2 to improve the ability of exploration.

In the MFPSO algorithm, the authors propose ω, and c_1, c_2 for each particle to be determined using three T-S systems:

$$\omega_i = TS(nf_i^k, nu_i^k) \tag{21}$$
$$(c_1)_i = TS(nf_i^k, nu_i^k) \tag{22}$$
$$(c_2)_i = TS(nf_i^k, nu_i^k) \tag{23}$$

where nf_i^k is defined in (13), nu_i^k has the form of

$$nu_i^k = \frac{u_i^k - u_{min}}{u_{max} - u_{min}} \tag{24}$$

and u_i^k is the number of iterations without change to the best personal position for the i-th particle. It should be noted that in equation (18), nu^k is calculated based on the global best position $\mathbf{gbest}$, whereas in equation (24) nu_i^k is calculated based on the personal best position $\mathbf{pbest}_i$. The vertices of fuzzy sets for nf_i^k and nu_i^k are defined as $\mathbf{p} = [0.2, 0.45, 0.65, 0.9]$, $\mathbf{q} = [0.2, 0.45, 0.65, 0.9]$.

The fuzzy rules for ω are the same as in (19). The fuzzy rules for c_1 and c_2 are given in tables (20). In (20) we have $PS = 1.4$, $PM = 1.7$, $PB = 1.9$ and $PR = 2.2$. For example, the rule R_{11} has the form

$$\begin{aligned} R_{11}: &\text{ IF } nf_i^k \in A_1 \text{ AND } nu_i^k \in B_1, \\ &\text{THEN } \omega = PS \text{ AND } c_1 = PR \text{ AND } c_2 = PR \end{aligned} \tag{25}$$

and the rule R_{44} has the form

$$\begin{aligned} R_{44}: &\text{ IF } nf_i^k \in A_4 \text{ AND } nu_i^k \in B_4, \\ &\text{THEN } \omega = PR \text{ AND } c_1 = PS \text{ AND } c_2 = PS \end{aligned} \tag{26}$$

Other rules can be interpreted similarly.

V. Results and discussion

In order to evaluate the algorithms, four common nonlinear benchmarks [11], [19] in the form of Ackley, Griewank, Rastrigin and Rosenbrock functions were used. For these functions, the asymmetric initialization method, similar to [11], was used. The velocities of particles were clamped to v_{max}, however, the positions of the particles were not limited. In Table I, the initialization ranges and v_{max} for the test functions are listed. In our experiments, two dimension sizes were chosen: $d = 10$ and $d = 30$. The number of iterations was set to 1000 and 2000 corresponding to the dimensions 10 and 30. The number of particles was equal to 30 and the number of trials was equal to 30 in all experiments. The parameters of algorithms were chosen on the basis of the papers [11] and [14]. The calculations were conducted using Matlab software and the "PSO Research Toolbox" by Evers [18]. The maximum average time of execution for one trial on a mobile workstation equipped with Intel(R) Core(TM) i7-2820QM did not exceeded 10 s.

The results for benchmark functions are presented in Table II. This table contains the basic statistics for the final value of the objective function and the iteration ($iter\_success$) in which the algorithm achieved the given value (th) of the

TABLE II
RESULTS FOR THE BENCHMARK FUNCTIONS

			fg			**iter_success**			
Algorithm	**d**	**iter**	**mean ± std**	**min**	**max**	**mean ± std**	**min**	**max**	**success rate** [%]
Ackley function: for $d = 10$, $th = 5\text{e}{-05}$; for $d = 30$, $th = 5$									
PSO1	10	1000	2.223e−05 ± 1.218e−04	3.553e−15	6.669e−04	244 ± 23	212	300	96.7
	30	2000	8.218e+00 ± 7.791e+00	1.421e−14	1.980e+01	191 ± 51	141	343	56.7
PSO2	10	1000	6.685e−01 ± 3.662e+00	8.882e−14	2.006e+01	735 ± 21	697	777	96.7
	30	2000	6.909e−01 ± 3.784e+00	6.994e−07	2.073e+01	1073 ± 46	991	1191	96.7
FPSO1	10	1000	1.332e+00 ± 5.068e+00	3.553e−15	2.006e+01	216 ± 23	189	280	93.3
	30	2000	9.953e−01 ± 3.781e+00	1.421e−14	2.079e+01	472 ± 129	352	1004	96.7
FPSO2	10	1000	3.790e−15 ± 9.013e−16	3.553e−15	7.105e−15	257 ± 23	224	338	100
	30	2000	4.146e+00 ± 8.032e+00	2.807e−13	2.003e+01	257 ± 63	198	444	80.0
FPSO3	10	1000	9.000e−01 ± 3.662e+00	3.553e−15	2.006e+01	175 ± 152	118	883	80.0
	30	2000	8.351e+00 ± 7.650e+00	1.344e+00	1.998e+01	192 ± 59	126	362	66.7
MFPSO	10	1000	2.392e−14 ± 4.870e−14	3.553e−15	2.558e−13	366 ± 37	296	474	100
	30	2000	4.125e+00 ± 8.384e+00	1.028e−04	2.087e+01	440 ± 115	330	765	80.0
Griewank function: for $d = 10$, $th = 0.1$; for $d = 30$, $th = 0.05$									
PSO1	10	1000	7.258e−02 ± 3.584e−02	3.197e−02	2.115e−01	188 ± 80	109	483	86.7
	30	2000	2.701e−02 ± 4.005e−02	0.000e+00	1.858e−01	356 ± 31	299	435	83.3
PSO2	10	1000	1.050e−01 ± 5.726e−02	7.396e−03	2.172e−01	724 ± 129	541	974	46.7
	30	2000	1.303e−02 ± 1.775e−02	2.092e−11	9.064e−02	1524 ± 52	1463	1702	96.7
FPSO1	10	1000	8.642e−02 ± 3.961e−02	1.969e−02	1.796e−01	204 ± 153	76	569	70.0
	30	2000	1.375e−02 ± 1.688e−02	0.000e+00	5.888e−02	329 ± 38	292	476	93.3
FPSO2	10	1000	7.701e−02 ± 3.531e−02	3.201e−02	1.847e−01	234 ± 138	108	534	76.7
	30	2000	1.492e−02 ± 2.037e−02	0.000e+00	9.562e−02	446 ± 37	368	539	93.3
FPSO3	10	1000	9.796e−02 ± 5.344e−02	1.970e−02	2.511e−01	118 ± 75	56	348	56.7
	30	2000	3.036e+00 ± 2.411e+00	1.049e+00	1.128e+01	–	–	–	0
MFPSO	10	1000	9.623e−02 ± 5.589e−02	2.955e−02	2.488e−01	372 ± 201	145	823	60.0
	30	2000	1.956e−02 ± 2.617e−02	1.718e−06	1.298e−01	1184 ± 172	901	1507	93.3
Rastrigin function: for $d = 10$, $th = 5$; for $d = 30$, $th = 50$									
PSO1	10	1000	7.097e+00 ± 3.969e+00	1.990e+00	1.890e+01	235 ± 118	119	555	40.0
	30	2000	1.053e+02 ± 2.743e+01	4.676e+01	1.512e+02	330 ± 0	330	330	3.33
PSO2	10	1000	3.715e+00 ± 1.865e+00	0.000e+00	7.960e+00	717 ± 118	533	939	83.3
	30	2000	3.819e+01 ± 9.564e+00	2.389e+01	6.766e+01	1454 ± 128	1179	1655	93.3
FPSO1	10	1000	5.804e+00 ± 2.575e+00	2.985e+00	1.293e+01	229 ± 85	100	406	56.7
	30	2000	4.580e+01 ± 8.148e+00	3.084e+01	6.368e+01	486 ± 124	266	745	60.0
FPSO2	10	1000	4.743e+00 ± 3.394e+00	0.000e+00	1.791e+01	276 ± 153	100	690	66.7
	30	2000	4.852e+01 ± 1.391e+01	2.388e+01	8.457e+01	505 ± 174	256	869	56.7
FPSO3	10	1000	1.270e+01 ± 5.817e+00	9.950e−01	2.388e+01	117 ± 31	93	163	13.3
	30	2000	9.155e+01 ± 2.314e+01	5.330e+01	1.353e+02	–	–	–	0
MFPSO	10	1000	3.689e+00 ± 2.101e+00	4.832e−03	8.955e+00	447 ± 208	178	976	80.0
	30	2000	3.317e+01 ± 9.586e+00	1.435e+01	5.132e+01	960 ± 402	374	1814	96.7
Rosenbrock function: for $d = 10$, $th = 30$; for $d = 30$, $th = 100$									
PSO1	10	1000	2.155e+01 ± 3.702e+01	1.381e−02	1.261e+02	136 ± 81	64	391	80.0
	30	2000	3.793e+01 ± 5.813e+01	6.057e−02	2.642e+02	558 ± 419	255	1597	90.0
PSO2	10	1000	3.602e+01 ± 1.308e+02	6.977e−01	7.244e+02	613 ± 136	438	966	86.7
	30	2000	8.484e+01 ± 7.490e+01	5.490e+00	3.359e+02	1657 ± 177	1361	1996	66.7
FPSO1	10	1000	1.761e+01 ± 3.553e+01	2.459e−03	1.371e+02	239 ± 249	47	791	90.0
	30	2000	6.247e+01 ± 7.706e+01	3.930e−01	3.082e+02	499 ± 220	275	1252	76.7
FPSO2	10	1000	2.529e+01 ± 5.990e+01	7.227e−02	2.577e+02	171 ± 157	56	798	83.3
	30	2000	5.779e+01 ± 4.336e+01	1.420e+00	1.683e+02	688 ± 395	348	1992	83.3
FPSO3	10	1000	7.058e+01 ± 2.101e+02	2.104e+00	1.152e+03	133 ± 152	42	635	73.3
	30	2000	8.847e+04 ± 1.825e+05	2.877e+02	8.705e+05	–	–	–	0
MFPSO	10	1000	1.499e+01 ± 4.056e+01	1.937e−02	2.239e+02	276 ± 255	89	936	93.3
	30	2000	2.248e+02 ± 2.320e+02	1.188e+01	9.200e+02	1681 ± 186	1353	1936	36.7

TABLE I
PARAMETERS OF BENCHMARK FUNCTIONS

Function	Init. ranges	$\mathbf{v_{max}}$
Ackley	$(15, 30)^d$	30
Griewank	$(300, 600)^d$	600
Rastrigin	$(2.56, 5.12)^d$	5.12
Rosenbrock	$(15, 30)^d$	30

TABLE III
RATINGS OF THE PSO ALGORITHMS

	$\mathbf{d = 10}$		$\mathbf{d = 30}$		$\sum$	
Algorithm	**mfg**	**rs**	**mfg**	**rs**	**mfg**	**rs**
PSO1	16	18	11	17	27	35
PSO2	11	5	20	9	31	14
FPSO1	13	18	18	20	31	38
FPSO2	18	15	15	18	33	33
FPSO3	6	18	5	6	11	24
MFPSO	20	10	15	11	35	21

objective function. The ratings of the algorithms for all benchmark functions are summarized in Tab. III. The following performance measures were used to evaluate the algorithms:

- mean of the objective function (mfg),
- relative success defined as $rs = \frac{mean\ \ of\ \ iter\_success}{success\_rate}$.

For these measures, the sums of the ratings are shown in Tab. III. These ratings were assigned in such a way that the best algorithm has six points and the worst has one point. For $success\_rate = 0$ (the algorithm has not succeeded) the number of points is equal to zero.

For the dimension $d = 10$ and the measure mfg the highest rating has the algorithm MFPSO proposed by the authors, while for the measure rs the highest rating have the PSO1, FPSO1 and FPSO3. For the dimension $d = 30$ and the measure mfg the highest rating has the algorithm PSO2, while for the measure rs the highest rating has the FPSO1. Analyzing the sum of ratings for mfg it can be seen that the best algorithm is the MFPSO. For rs the MFPSO is the one before last. However, it should be emphasized that in the evaluation of optimization algorithms the most important criterion is the obtained objective function value.

VI. CONCLUSION

In this paper, the evaluation of selected fuzzy particle swarm optimization algorithms was presented. Two non-fuzzy and four fuzzy algorithms were considered. The main contributions of this paper are as follows:

- the application of the Takagi-Sugeno system that is more computationally efficient than the Mamdani system,
- a proposal for the use of the MFPSO algorithm, in which each of the particles has its own inertia weight and the coefficients of the cognitive and social components,
- the evaluation of selected fuzzy PSO algorithms using common benchmark functions.

Further work will focus on improving the proposed algorithm, building models to support the training process in sport [20], and the analysis of athletes' technique [8].

ACKNOWLEDGMENT

This work has been supported by the Polish Ministry of Science and Higher Education under grant No. U-722/DS/M.

REFERENCES

[1] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proc. of IEEE Int. Conf. on Neural Networks*, vol. 4. IEEE Press, Piscataway, NJ, 1995, pp. 1942–1948.
[2] A. Alfi and M.-M. Fateh, "Intelligent identification and control using improved fuzzy particle swarm optimization," *Expert Systems with Applications*, vol. 38, no. 10, pp. 12 312–12 317, 2011.
[3] T. Niknam, "A new fuzzy adaptive hybrid particle swarm optimization algorithm for non-linear, non-smooth and non-convex economic dispatch problem," *Applied Energy*, vol. 87, no. 1, pp. 327–339, 2010.
[4] S. Saini, N. Zakaria, D. R. A. Rambli, and S. Sulaiman, "Markerless human motion tracking using hierarchical multi-swarm cooperative particle swarm optimization," *PLoS ONE*, vol. 10, no. 5, 2015.
[5] M. Adamczyk, "Parallel feature selection algorithm based on rough sets and particle swarm optimization," in *Computer Science and Information Systems (FedCSIS), 2014 Federated Conf. on*, Sept 2014, pp. 43–50.
[6] D. Srinivasan, W. H. Loo, and R. L. Cheu, "Traffic incident detection using particle swarm optimization," in *Swarm Intelligence Symposium. SIS '03. Proceedings of the IEEE*, April 2003, pp. 144–151.
[7] A. Karami and M. Guerrero-Zapata, "A fuzzy anomaly detection system based on hybrid PSO-Kmeans algorithm in content-centric networks," *Neurocomputing*, vol. 149, Part C, pp. 1253–1269, 2015.
[8] T. Krzeszowski, K. Przednowek, K. Wiktorowicz, and J. Iskra, "Estimation of hurdle clearance parameters using a monocular human motion tracking method," *Computer Methods in Biomechanics and Biomedical Engineering*, vol. 19, no. 12, pp. 1319–1329, 2016, PMID: 26838547.
[9] M. Clerc and J. Kennedy, "The particle swarm - explosion, stability, and convergence in a multidimensional complex space," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 1, pp. 58–73, 2002.
[10] R. C. Eberhart and Y. Shi, "Comparing inertia weights and constriction factors in particle swarm optimization," in *Evolutionary Computation, 2000. Proceedings of the 2000 Congress on*, vol. 1, 2000, pp. 84–88.
[11] Y. Shi and R. C. Eberhart, "Fuzzy adaptive particle swarm optimization," in *Proceedings of the Congress on Evolutionary Computation*, vol. 1, 2001, pp. 101–106.
[12] A. M. Abdelbar, S. Abdelshahid, and D. C. Wunsch, "Fuzzy PSO: a generalization of particle swarm optimization," in *Proceedings. IEEE International Joint Conference on Neural Networks*, vol. 2, July 2005, pp. 1086–1091.
[13] H. Liu, A. Abraham, and W. Zhang, "A fuzzy adaptive turbulent particle swarm optimisation," *Int. J. Innov. Comput. Appl.*, vol. 1, no. 1, pp. 39–47, 2007.
[14] Y.-T. Juang, S.-L. Tung, and H.-C. Chiu, "Adaptive fuzzy particle swarm optimization for global optimization of multimodal functions," *Information Sciences*, vol. 181, no. 20, pp. 4539–4549, 2011, Special Issue on Interpretable Fuzzy Systems.
[15] J. J. D. Nesamalar, P. Venkatesh, and S. C. Raja, "Managing multi-line power congestion by using Hybrid Nelder-Mead - Fuzzy Adaptive Particle Swarm Optimization (HNM-FAPSO)," *Applied Soft Computing*, vol. 43, pp. 222–234, 2016.
[16] T. Takagi and M. Sugeno, "Fuzzy identification of systems and its applications to modeling and control," *Systems, Man and Cybernetics, IEEE Transactions on*, no. 1, pp. 116–132, 1985.
[17] E. Mamdani and S. Assilian, "An experiment in linguistic synthesis with a fuzzy logic controller," *International Journal of Man-Machine Studies*, vol. 7, no. 1, pp. 1–13, 1975.
[18] G. Evers, "PSO Research Toolbox (Version 20110515), M.S. thesis code," 2016. [Online]. Available: http://www.georgeevers.org/pso_research_toolbox.htm
[19] J. J. Liang, P. N. Suganthan, and K. Deb, "Novel composition test functions for numerical global optimization," in *Proceedings. IEEE Swarm Intelligence Symposium. SIS 2005*, June 2005, pp. 68–75.
[20] K. Wiktorowicz, K. Przednowek, L. Lassota, and T. Krzeszowski, "Predictive modeling in race walking," *Computational Intelligence and Neuroscience*, vol. 2015, p. 9, 2015, Article ID 735060.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 577–584
DOI: 10.15439/2016F441
ACSIS, Vol. 8. ISSN 2300-5963

Minimizing the Number of Late Multi-Task Jobs on Identical Machines in Parallel

Lingxiang Li
Hunan University of Science and Engineering
Yongzhou
Hunan, P.R.China

Haibing Li
BNP Paribas,
787 Seventh Ave
New York City, NY 10019, USA

Hairong Zhao
Purdue University Northwest
2200 169th Street, IN 46323, USA
Email: hairong@pnw.edu

***Abstract*—We consider the problem of scheduling multi-task jobs on identical machines in parallel. Each multi-task job consists of one or more tasks. Each job has a release date and a due date. A task of a job can be processed by any one of the machines. Multiple machines can process the tasks of a job concurrently. The completion time of a job is the time at which all its individual tasks have been completed. A job is late if it is completed after its due date. We study the problem of minimizing the total number of late jobs. We show that while some special cases are solvable, the general problem is NP-hard and there exists no polynomial time ρ-approximation algorithm, for any $\rho > 1$. We present a general algorithm for the problem and derive from it six heuristics whose performance is evaluated by experimental results.**

I. Introduction

THE PROBLEM under consideration is scheduling multi-task jobs on identical machines in parallel. It can be stated as follows: Assume there are m identical machines and n jobs. Each job j $(j = 1, 2, \ldots, n)$, which is available at time r_j and has a due date d_j, consists of k_j $(1 \leq k_j \leq k)$ individual tasks (or operations), where k is the maximum number of tasks that a job may have. Each task $l = 1, 2, \ldots, k_j$ of job j, denoted by o_{lj}, can be processed by any one of the machines, and its processing time is denoted by p_{lj}. The individual tasks of a job can be assigned to multiple machines so that they can be processed concurrently. When a machine switches over from one task to another, no setup is required. The completion time of job j, denoted by C_j, is the time at which all individual tasks of job j have been completed. If we let C_{lj} denote the completion time of task o_{lj}, it is clear that $C_j = \max_{1 \leq l \leq k_j} \{C_{lj}\}$. For the ease of description, we also let $P_{i,j}$ and $C_{i,j}$ denote the total processing time and the completion time of job j on machine i, respectively. By definition, $C_j = \max_{1 \leq i \leq m} \{C_{i,j}\}$. A job is late if $C_j > d_j$, and by standard notation, $U_j = 1$ if $C_j > d_j$ and $U_j = 0$ otherwise. We are interested in minimizing the total number of late jobs $\sum U_j$. Let $jbs\ n$ denote n jobs and $tsk\ k$ denote the maximum number of tasks that a job may have. The problem is denoted by $Pm \mid jbs\ n, tsk\ k, r_j \mid \sum U_j$, where m, n, and k can be either fixed or arbitrary. If any of these is not fixed, it is removed from the notation. For example, $P \mid jbs\ 10, tsk \mid \sum U_j$ denotes that m and k are arbitrary but the number of jobs n is 10. If $r_j = 0$ for all jobs, r_j is removed from the notation as well.

The above problem is a more general description of the fully flexible case of customer order scheduling models described in [1], so it is not limited to any specific application contexts, e.g. manufacturing environments. In addition to the application examples surveyed in [1], we yet give another real-life application example in software project management, with the objective to minimize $\sum U_j$. It is not unusual that in a software development team, new projects with various due dates are requested from business lines. A development manager usually creates a parent task for each new project, and creates multiple child tasks (for example, independent modules or loosely coupled modules as a result of well-designed software architecture) associated with the parent task so that multiple software developers in the development team can work on the project simultaneously. A parent task (project) is completed if and only if all child tasks are completed. All software developers (assuming that they have the same skills at the same proficiency levels after certain cross-training) in the team can work on all child tasks. The challenge for the development manager is to find a good schedule for the team, to minimize the number of parent tasks (projects) that cannot be completed before their due dates, so that the relationship and partnership between the development team and the business teams can be positively built up.

Some past work has been done for this problem with the objective to minimize the total weighted completion time $\sum w_j C_j$ and its un-weighted version. Even when $w_j = 1$ for all j, the problem with an arbitrary k is ordinary NP-hard for any fixed $m \geq 2$ and strongly NP-hard when m is arbitrary (see Blocher and Chhajed [2]). On the other hand, when $k = 1$, the problem becomes the classical problem $P \mid\mid \sum w_j C_j$ which is strongly NP-hard, and for any fixed $m \geq 2$, the problem is equivalent to $Pm \mid\mid \sum w_j C_j$ which is ordinary NP-hard (see [3]). In the aspect of algorithms, when $w_j = 1$, Blocher and Chhajed [2] presented six heuristics with empirical analysis of the performance of the heuristics. One of the heuristics was also studied by Yang [4], [5] where it was shown a worst-case performance bound of $7/6$ for $m = 2$. For arbitrary m, two classes of nine heuristics with proven worst-case performance bounds of either $(2 - 1/m)$ or m were studied by Leung, Li and Pinedo [6].

To the best of our knowledge, no past work has ever been done for this problem with the objective to minimize $\sum U_j$.

978-83-60810-90-3/$25.00

In this paper, we are interested in both the complexity and the algorithm aspects of the problem. The remainder of the paper is organized as follows. In Section II, we present some preliminary results regarding some properties of optimal schedules. In Section III, we show that some cases are NP-hard, while some other cases are polynomially solvable. Then, after showing the non-approximability for the general case without release dates, we present in Section IV a general algorithm scheme for the problem and derive from it six heuristics, whose performance is evaluated by experimental results in Section V. Finally, we present conclusions in Section VI.

II. Preliminary Results

We first look into some properties of an optimal schedule for problem $P \mid jbs, tsk \mid \sum U_j$:

Lemma 2.1 (Optimal Property): For problem $P \mid jbs, tsk \mid \sum U_j$, there exists an optimal schedule in which:

a) the tasks (if more than two) of a job are assigned consecutively on each machine;
b) the early jobs are scheduled in nondecreasing order of their due dates on each machine.

Proof: **a)** Suppose that there exists an optimal schedule in which some tasks of job j assigned on machine i are not consecutive, we keep the last task of job j where it is, but make necessary interchange to shift all its other tasks backward so that they become consecutive, thus the completion time of job j remains unchanged. However, other jobs, whose tasks are shift forward due to the interchanges, would be completed earlier. Thus, no new late jobs are introduced, and the resulting schedule remains optimal.

b) Suppose that in an optimal schedule there exist two early jobs, j_1 and j_2, and a machine i, such that $C_{i,j_2} < C_{i,j_1}$ but $d_{j_1} < d_{j_2}$. We assume the tasks of both jobs are scheduled consecutively according to **a)**. We remove job j_2, and push forward all jobs before j_1 (and j_1 itself), to fill the hole produced by removing j_2, then place j_2 right after j_1. Clearly, all jobs except j_2 are completed earlier. As for job j_2, its new completion time $C'_{i,j_2} = C_{i,j_1} \leq d_{j_1} < d_{j_2}$. Thus, job j_2 is still completed on time, and the resulting schedule remains optimal. ■

Note that in an optimal schedule, the tasks of a job are not necessarily to be assigned across all machines. Some machines may be assigned with multiple tasks of the job, while some other machines may not be assigned with any tasks of the same job. To illustrate this, consider an example as follows: $m = 2, n = 2$, $p_{11} = p_{21} = 2, d_1 = 4$, $p_{12} = 4, d_2 = 5$. Clearly, in an optimal schedule for this instance, the only task of job 2 must be assigned to one machine, while the two tasks of job 1 must be assigned to another machine. This yields a schedule with no late jobs.

For any instance I of $P \mid jbs, tsk \mid \sum U_j$, to derive a lower bound for it, we can construct an instance I' of problem $1 \mid\mid \sum U_j$ which can be solved in $O(n \log n)$ by Moore-Hodgson's algorithm [7]: For each job j, construct a job j for I' with processing time $p'_j = \sum_l p_{lj}/m$ and due date $d'_j = d_j$. Let S_{OPT} and S'_{OPT} denote an optimal schedule for instance I and I', respectively. Then, we have the following lower bound which might be useful for the design of a branch-and-bound algorithm, or for evaluating the performance of heuristic algorithms by experimental analysis as what we will show later.

Lemma 2.2 (Lower Bound): For any instance I of problem $P \mid jbs, tsk \mid \sum U_j$, and its corresponding instance I' of problem $1 \mid\mid \sum U_j$ constructed in the way described above, the optimal objective value for I has the following lower bound:

$$\sum_j U_j(S_{OPT}) \geq \sum_j U_j(S'_{OPT}).$$

Proof: Consider an optimal schedule S_{OPT} for instance I, let S_e denote the sub-schedule of all early jobs in S_{OPT}. We construct a schedule S' for instance I' as follows: **a)** For the jobs in S_e, schedule the corresponding jobs of I' in nondecreasing order of d'_j which equals to d_j, let the sub-schedule be S'_e; **b)** Append the rest jobs to the end of S'_e in arbitrary order.

We shall show that all jobs in S'_e are on time as well. Without loss of generality, we assume that the early jobs in S_e are indexed by $1, 2, \ldots, |S_e|$. Consider any partial schedule for jobs $1, 2, \ldots, j^*$ in S_e where $1 \leq j^* \leq |S_e|$. Since job j^* is early, we have $C_{j^*} = \max_{1 \leq i \leq m} \left\{ \sum_{j=1}^{j^*} P_{i,j} \right\} \leq d_{j^*}$. Due to the fact that this partial schedule may not be aligned up at the end of each machine, we have $\max_{1 \leq i \leq m} \left\{ \sum_{j=1}^{j^*} P_{i,j} \right\} \geq \sum_{j=1}^{j^*} \sum_{l=1}^{k_j} p_{lj}/m = \sum_{j=1}^{j^*} p'_j = C'_{j^*}$, it follows that $C'_{j^*} \leq d_{j^*} = d'_{j^*}$, implying that job j^* is early in S'_e. Thus, $\sum_j U_j(S') \leq \sum_j U_j(S_{OPT})$. The lower bound follows due to the fact that $\sum_j U_j(S') \geq \sum_j U_j(S'_{OPT})$. ■

III. Complexity Results

In this section, we investigate the cases that are either NP-hard or polynomially solvable. The goal is to establish a borderline between the hard cases and the polynomially solvable ones.

A. NP-hard Cases

Before we proceed further, we first introduce the following NP-complete problems (see Garey and Johnson [8]) that will be used for reduction later:

Definition 1 (Partition Problem): Given a list $A = (a_1, a_2, \ldots, a_n)$ of n positive integers, can A be partitioned into two subsets A_1 and A_2 such that $A_1 \cup A_2 = A$ and $\sum_{a_j \in A_1} a_j = \sum_{a_j \in A_2} a_j = B = \frac{1}{2} \sum_{a_j \in A} a_j$?

Definition 2 (3-Partition Problem): Given a list $A = (a_1, a_2, \ldots, a_{3m})$ of $3m$ positive integers such that $\sum_j a_j = mB, B/4 < a_j < B/2$ for each $1 \leq j \leq 3m$, is there a partition A into m subsets $A_1, A_2, \ldots, A_m$ such that $\cup_i^m A_i = A$ and $\sum_{a_j \in A_i} a_j = B$ for each $1 \leq i \leq m$?

Note that even though these two problems are closely related, the **Partition** problem is NP-complete in the ordinary sense, while the **3-Partition** problem is strongly NP-complete.

To show the NP-hardness of several cases, we first start with two restricted cases.

Theorem 3.1: Problem $P \mid jbs, tsk\ 1, d_j = d \mid \sum U_j$ is NP-hard in the strong sense.

Proof: We shall show that the 3-Partition problem is reducible to problem $P \mid jbs, tsk\ 1, d_j = d \mid \sum U_j$. Given an instance of $A = (a_1, a_2, \ldots, a_{3m})$ of 3-Partition, we construct an instance of $P \mid jbs, tsk\ 1, d_j = d \mid \sum U_j$ as follows: There are m machines and $3m$ jobs such that $p_{1j} = a_j$ and $d_j = B$ for each $1 \leq j \leq 3m$. The transformation clearly takes polynomial time. The decision version of the scheduling problem asks if there exists a schedule such that $\sum U_j = 0$?

If the 3-Partition instance has a "Yes" solution, we let the partition be $A_1, A_2, \ldots, A_m$. For each A_i $(1 \leq i \leq m)$, we schedule on machine i the three jobs constructed from the three elements which are in A_i. Thus, we have a schedule such that the finish time on each machine is exactly B, implying that no job is late, i.e., $\sum U_j = 0$.

Conversely, if the scheduling problem instance has a schedule such that $\sum U_j = 0$, it implies that the finish time on each machine has to be exactly B. Due to $p_{1j} = a_j$ and $B/4 < a_j < B/2$, each machine must have 3 jobs exactly, otherwise, the finish time with less/more jobs on a machine would be strictly less/larger than B. Let A_i be the triplet corresponding to the 3 jobs scheduled on each machine $1 \leq i \leq m$, then $A_1, A_2, \ldots, A_m$ is a "Yes" solution to the 3-Partition instance. ■

Theorem 3.2: Problem $Pm \mid jbs, tsk\ 1, d_j = d \mid \sum U_j$ is NP-hard in the ordinary sense for every fixed $m \geq 2$.

Proof: It is sufficient to consider the special case for $m = 2$. We shall show that the Partition problem is reducible to $P2 \mid jbs, tsk\ 1, d_j = d \mid \sum U_j$. Given an instance of the Partition problem, we construct an instance of $P2 \mid jbs, tsk\ 1, d_j = d \mid \sum U_j$ as follows: Let there be n jobs, $p_{1j} = a_j, d_j = B$ for each job $1 \leq j \leq n$. The decision version of the scheduling problem asks if there exists a schedule such that $\sum U_j = 0$?

It is easy to see that the Partition problem instance has a "Yes" solution if and only if the $P2 \mid jbs, tsk\ 1, d_j = d \mid \sum U_j$ instance has a schedule such that $\sum U_j = 0$. ■

Theorem 3.1 and Theorem 3.2 immediately imply the NP-hardness of their general cases, respectively:

Theorem 3.3: As generalization of the cases with common due dates,

a) both problem $P \mid jbs, tsk\ 1 \mid \sum U_j$ and problem $P \mid jbs, tsk \mid \sum U_j$ are strongly NP-hard;
c) both problem $Pm \mid jbs, tsk\ 1 \mid \sum U_j$ and problem $Pm \mid jbs, tsk \mid \sum U_j$ are NP-hard in the ordinary sense for every fixed $m \geq 2$.

On the other hand, when it is restricted to only one job, which has arbitrary number of tasks, the special cases are still NP-hard, as shown below:

Theorem 3.4: Problem $P \mid jbs\ 1, tsk \mid \sum U_j$ is NP-hard in the strong sense.

Proof: We shall show that the 3-Partition problem is also reducible to $P \mid jbs\ 1, tsk \mid \sum U_j$. Given any 3-Partition instance, we construct an instance of the scheduling problem as follows: Let there be 1 job with $3m$ tasks such that $p_{l1} = a_l$ for each $l = 1, 2, \ldots, 3m$; and let $d_1 = B$. The decision version of the scheduling problem asks if there exists a schedule such that $U_1 = 0$?

Similar argument as described in Theorem 3.1 shows that the 3-Partition instance has a "Yes" solution if and only if the scheduling instance has a schedule such that $U_1 = 0$. ■

Theorem 3.5: Problem $Pm \mid jbs\ 1, tsk \mid \sum U_j$ is NP-hard in the ordinary sense for every fixed $m \geq 2$.

Proof: It is sufficient that we consider the special case $m = 2$. Again, a simple reduction from the Partition problem shows that the problem is NP-hard in the ordinary sense for $m = 2$. ■

With the presence of release dates, all the NP-hard cases presented above would be harder. Further, we show that the problem $1 \mid jbs, tsk, r_j \mid \sum U_j$ is NP-hard in the strong sense.

Theorem 3.6: Problem $1 \mid jbs, tsk, r_j \mid \sum U_j$ is NP-hard in the strong sense.

Proof: Since problem $1 \mid jbs, tsk\ 1, r_j \mid \sum U_j$ is equivalent to $1 \mid r_j \mid \sum U_j$ which is strongly NP-hard (due to that $1 \mid r_j \mid \sum L_{max}$ is strongly NP-hard [9] and L_{max} is reducible to U_j [10], [11], [12]), thus its general version $1 \mid jbs, tsk, r_j \mid \sum U_j$ is also NP-hard in the strong sense. ■

B. Polynomially Solvable Cases

We start with the single-machine cases:

Theorem 3.7: The following problems:

a) $1 \mid jbs, tsk\ k \mid \sum U_j$; and
b) $1 \mid jbs, tsk \mid \sum U_j$.

can be solved in polynomial time.

Proof: As a direct result of **a)** in Lemma 2.1, by aggregating the tasks of each job j into a single task with processing time $\sum_l p_{lj}$, both $1 \mid jbs, tsk\ k \mid \sum U_j$ and $1 \mid jbs, tsk \mid \sum U_j$ can be solved in polynomial time by Moore-Hodgson's algorithm [7]. ■

Now we consider a special case in which the tasks of all jobs have identical processing times:

Theorem 3.8: Problem $P \mid jbs, tsk, p_{lj} = p \mid \sum U_j$ can be solved in $O(n \log n + \sum k_j)$ time.

Proof: We can find the optimal schedule in two steps: (1) identify the early job set E; and (2) schedule the early jobs in E. To find the early jobs, we can do the following:

a) Sort and reindex the jobs such that $d_1 \leq d_2 \leq \ldots \leq d_n$.
b) $E = \emptyset$, $sumK = 0$
c) For each $1 \leq j \leq n$
   $E = E \cup \{j\}$, $sumK = sumK + k_j$
   If $\lceil \frac{sumK}{m} \rceil * p > d_j$
      Let $i = \text{argmax}_{x \in E}\, k_x$
      $E = E \setminus \{i\}$, $sumK = sumK - k_i$

To schedule the early jobs, we simply take the tasks of the early jobs from E in non-decreasing order of the due dates, and assign them one by one to the machines 1, 2, $\cdots$, m, then 1 2, $\cdots$ m again, and so on.

First we show that our schedule is optimal. Without loss of generality, we assume that for any job j, we have $\lceil \frac{k_j}{m} \rceil * p \leq d_j$. Otherwise, it must be late in any schedule.

Notice that in step (1) the jobs are processed in non-decreasing order of their due dates and in (c), if $\lceil \frac{sumK}{m} \rceil * p >$

d_j, we remove the job with the largest number of tasks (thus maximum processing time) from E. This guarantees that if job $j \in E$ after step (1) , then $\lceil \frac{\sum_{i \in E, i \leq j} k_i}{m} \rceil * p \leq d_j$. By the way we schedule the tasks in step (2), we have $C_j = \lceil \frac{\sum_{i \in E, i \leq j} k_i}{m} \rceil * p$, thus $C_j \leq d_j$. So all jobs in E are scheduled on time. Also notice that $|E|$ must be maximum due to the fact that the largest job is chosen to be tardy in (c).

For the time complexity, step (1) can be implemented in $O(n \lg n)$ time if we use priority queue to maintain the jobs in E, and step (2) can be implemented in $\sum k_j$. ■

With the presence of release dates, some special cases are still polynomially solvable. Before we proceed further, we first show that (as we are not aware of any proof in the literature), the classical problem $1 \mid r_j, d_j = d \mid \sum U_j$ can be solved in polynomial time, even though the general case $1 \mid r_j \mid \sum U_j$ is strongly NP-hard.

Theorem 3.9: Problem $1 \mid r_j, d_j = d \mid \sum U_j$ can be solved in $O(n \log n)$ time.

Proof: Consider the following algorithm:

a) For each job with $r_j + p_j > d$, simply mark it as late and exclude it from the next steps.
b) For the remaining jobs, define a new deadline $d'_j = d - r_j$.
c) Treat time d as time 0, and schedule the jobs with new deadlines and release times 0 backwards by applying Hodgson-Moore algorithm.

The correctness of the algorithm lies in the fact that the modified problem is equivalent to the original one and Hodgson-Moore algorithm is optimal for the modified problem. ■

Theorem 3.10: Problem $1 \mid jbs, tsk, r_j, d_j = d \mid \sum U_j$ can be solved in $O(n \log n + \sum k_j)$ time.

Proof: The key observation is that, due to the common due date, there exists an optimal schedule for any problem instance in which the tasks of each early job are scheduled consecutively. Otherwise, shifting the separated tasks (except the first one) of the job forward so that they are consecutive, and shifting the in-between tasks of other jobs backward would not violate their release dates and the common due date, and the schedule remains feasible and optimal. Thus, by aggregating all tasks of each job as a single task with processing time of $\sum_l p_{lj}$, problem $1 \mid jbs, tsk, r_j, d_j = d \mid \sum U_j$ can be polynomially solved by an equivalent $1 \mid r_j, d_j = d \mid \sum U_j$ problem according to Theorem 3.9. Since aggregating the tasks takes $O(\sum k_j)$ time, the algorithm runs in $O(n \log n + \sum k_j)$ time. ■

Theorem 3.11: Problem $1 \mid jbs, tsk, r_j, p_{lj} = 1 \mid \sum U_j$ can be solved in $O(n^5)$ time.

Proof: Consider the following algorithm:

a) For any instance I of problem $1 \mid jbs, tsk, r_j, p_{lj} = 1 \mid \sum U_j$, construct an instance I' of the classical preemptive scheduling problem $1 \mid r_j, pmtn \mid \sum U_j$ with $p'_j = \sum_l p_{lp}, r'_j = r_j, d'_j = d_j$.
b) Solve I' by Lawler's Dynamic Programming algorithm in $O(n^5)$ time [11], [13].
c) Construct the schedule for I exactly from the optimal schedule for I', by mapping the jobs one by one.

Apparently, the obtained schedule is optimal and the running time is dominated by Lawler's algorithm which runs in $O(n^5)$ time. ■

Since the cases with release dates are harder than the ones with no release dates, we focus on the design and analysis of algorithms for the latter cases in the next two sections.

IV. Heuristic Algorithms for Problem $P \mid jbs, tsk \mid \sum U_j$

Due to the strong NP-hardness of problem $P \mid jbs, tsk \mid \sum U_j$, it would be of interest to see if there exists any good approximation algorithm for it. Unfortunately, the following negative result shows that there exists no such approximation algorithm unless $P = NP$.

Theorem 4.1: Unless $P = NP$, there exists no polynomial time ρ-approximation algorithm ($1 < \forall \rho < \infty$) for both problem $P \mid jbs, tsk \mid \sum U_j$ and problem $Pm \mid jbs, tsk \mid \sum U_j$ (for any fixed $m \geq 2$).

Proof: It is sufficient to consider a special case of the problem, i.e., $P \mid jbs\ 1, tsk \mid \sum U_j$, for which we have only 1 job with arbitrary number of tasks to be scheduled on m machines where m is not fixed. Given any problem instance I, since it has only 1 job, being either late or on time, the objective value returned by $\mathcal{A}$ must be $\sum U_j \in \{0, 1\}$. We consider the following four cases:

Case1. The optimal objective value is 0 and $\mathcal{A}$ returns 0, which is optimal.
Case2. The optimal objective value is 1 and $\mathcal{A}$ returns 1, which is optimal.
Case3. The optimal objective value is 0 and $\mathcal{A}$ returns 1. The performance ratio is ∞.
Case4. The optimal objective value is 1 and $\mathcal{A}$ returns 0, which is impossible.

First of all, Case 4 can be excluded for $\mathcal{A}$, since it contradicts to the optimality of the optimal objective value. As for other cases, we claim that there exists at least one problem instance so that Case 3 is true for $\mathcal{A}$. Otherwise, assume that "only" Case 1 and Case 2 are true for $\mathcal{A}$, it simply implies that $\mathcal{A}$ is optimal. Since $\mathcal{A}$ is also polynomial by assumption, it would imply $P = NP$. Therefore, our claim must be true unless $P = NP$, implying that the algorithm $\mathcal{A}$ must be unbounded due to Case 3.

Similar arguments apply to $Pm \mid jbs, tsk \mid \sum U_j$ for its special case when $m = 2$. ■

An observation on the non-approximability of $P \mid jbs\ 1, tsk \mid \sum U_j$ and $Pm \mid jbs\ 1, tsk \mid \sum U_j$ is that, each of them intrinsically consists of a NP-hard subproblem to solve (i.e., the C_{max} problem on parallel machines) and yet its objective value is limited to only two numbers, i.e., either 0 or 1. Thus, there is no much freedom for any algorithm to approximate within.

We first present a general algorithm scheme to identify and schedule a set of early jobs, and then derive some specific

algorithms from it by customizing the task sorting criterion and the machine selection criterion.

General-Scheme GS **for** $P \mid jbs, tsk \mid \sum U_j$

Input. A set of n multi-task jobs; the number of machines m.
Output. A set of early jobs E and their schedule S_e.

Sort the n jobs such that $d_1 \leq d_2 \leq \ldots \leq d_n$.
for each job $j = 1, 2, \ldots, n$
  ≪***sort its tasks by certain criterion***≫
  assuming the sorted order is o_{1j}, o_{2j}, …, $o_{k_j j}$

Let $E = \emptyset$, and S_e be an empty schedule
$j = 1$, $firstTry = true$.
While $j \leq n$
  $late = false$
  $l = 1$.
  while ($l \leq k_j$ and $late = false$)
    ≪***select machine*** i^* ***by certain criterion***≫
    assign task o_{lj} to machine i^* in S_e.
    if o_{lj} is late
      late = true.
      remove all tasks o_{1j}, o_{2j}, …, o_{lj} from S_e.
      if $firstTry = true$
        let $j^* = \text{argmax}_{k \in E \cup \{j\}} \{\sum_l p_{lk}\}$
        if $j^* = j$, then $j = j + 1$.
        else
          remove all tasks of j^* from S_e.
          $firstTry = false$.
      else $j = j + 1$, $firstTry = true$.
    else
      if $l = k_j$
        if $firstTry = true$, then $E = E \cup \{j\}$.
        else $E = E \setminus \{j^*\} \cup \{j\}$.
        $j = j + 1$, $firstTry = true$.
      $l = l + 1$.
return S_e

It should be noted that, when $k = 1$ and $m = 1$, the above general algorithm schema would work in the same way as Moore-Hodgson's algorithm for $1 \mid\mid \sum U_j$. Thus, we can regard it as generalization of Moore-Hodgson's algorithm. To derive a specific algorithm from the above general scheme, we need to specify two criteria as marked within $\ll \ldots \gg$. The first criterion to be specified in Step 1 is for sorting the individual tasks of each job $j = 1, 2, \ldots, n$, while the second one in Step 2 is for choosing a machine to process the task under consideration.

Intuitively, we consider two **Task Sorting Criteria**:

- **Arbitrary Order.** No sorting, just keep the original ordering of the tasks as given in input. Thus, it takes no extra running time.
- **Non-increasing Order.** Sort the tasks of job j in non-increasing order of their processing times such that $p_{1j} \geq p_{2j} \geq \ldots p_{k_j j}$ for each $j = 1, 2, \ldots, n$. It takes $O(k \log k)$ time for each job.

To assign a task to a machine, we consider three **Machine Choosing Criteria**:

Smallest Load. Choose the machine with the smallest load. This is used in the well-known Longest Processing Time first rule (LPT) and List Scheduling algorithm (LS) for problem $P \mid\mid C_{max}$ [16].

First Fit. Choose the first machine which can process the task before the job's due date. The First-Fit algorithms were originally designed for the Bin Packing problem [17].

Best Fit. Choose the machine with the largest load but can still process the task before the job's due date.The Best-Fit algorithms were also originally designed for the Bin Packing problem.

TABLE I
SIX ALGORITHMS DERIVED FROM THE TWO TYPES OF CRITERIA

Algorithm	Machine Choosing	Sorting	Task Assignment
GS_{LS}	Smallest Load	Arbitrary	LS
GS_{LPT}	Smallest Load	Non-increasing	LPT
GS_{FF}	First Fit	Arbitrary	First Fit
GS_{FFD}	First Fit	Non-increasing	First Fit Decreasing
GS_{BF}	Best Fit	Arbitrary	Best Fit
GS_{BFD}	Best Fit	Non-increasing	Best Fit Decreasing

Naturally, combination of these two types of criteria produces six different concrete algorithms, which are enumerated in Table I.

In essence, each algorithm derived from the above general algorithm scheme combines the Earliest Due Date first (EDD) rule [18] (at job level), and either an algorithm for problem $P \mid\mid C_{max}$ [16] or an algorithm for the Bin Packing problem [17] (at task level).

Clearly, all algorithms derived from the general scheme run in polynomial time. It is not surprising that, due to Theorem 4.1, the performance ratio of these algorithms is not bounded. To illustrate this by a simple example, we consider the following instance: $n = 1, m = 4, k = 9, p_{11} = 7, p_{21} = 4, p_{31} = p_{41} = 5, p_{51} = p_{61} = 6, p_{71} = 7, p_{81} = p_{91} = 4, d_1 = 12$. The "trial" assignment of tasks by all these six algorithms, as illustrated by Table II, would result in schedules in which the job is late. Thus, all these heuristic algorithms return $\sum_j U_j = 1$. However, in an optimal schedule, as illustrated by the last column in the same table, the job is on time, and the objective value is 0. Thus, the performance ratio of all these heuristic algorithms is ∞ .

Even though the above worst-case example shows that the heuristic algorithms could perform arbitrarily bad, in practice, we expect that their average performance could be much better. To this end, we evaluate these algorithms by experimental results in the next section.

V. EXPERIMENTAL EVALUATION

To evaluate the above heuristic algorithms, we choose the number of jobs $n = 500$, and the number of machines $m = 20$. Problem instances of varying hardness are generated according

TABLE II
TRIAL ASSIGNMENT OF TASKS FOR THE WORST-CASE EXAMPLE

Machine	GS_{LS}	GS_{LPT}	GS_{FF}	GS_{FFD}	GS_{BF}	GS_{BFD}	OPT
1	7, 4, 4	7, 4, 4	7, 4	7, 5	7, 4	7, 5	7, 5
2	4, 6, 4	7, 4	5, 5,	7, 5	5, 5, 4	7, 5	7, 5
3	5, 6	6, 5	6, 6	6, 4	6, 6	6, 4, 4	6, 6
4	5, 7	6, 5	7, 4, 4	6, 4, 4	7, 4	6, 4	4, 4, 4

to different characteristics of the due dates, in a similar way described in Leung, Li and Pinedo [19].

First of all, for each job $j = 1, 2, \ldots, n$, the number of tasks k_j is randomly generated from the uniform distribution $[1, 10m]$. Then, for each task $l = 1, 2, \ldots, k_j$, p_{lj} is generated from the uniform distribution $[1, 100]$. Finally, after all jobs are generated, for each job $j = 1, 2, \ldots, n$, its due date d_j is generated from the following uniform distribution:

$$[P(1 - \delta_1/2 - \delta_2), P(1 + \delta_1/2 - \delta_2)],$$

where

$$P = \sum_{j=1}^{n} \sum_{l=1}^{k_j} p_{lj}/m,$$

and δ_1 and δ_2 determines the range in which the due dates lie and adjusts the tightness of the due dates, respectively. Also, in generating d_j, we ensure that

$$d_j \geq \max \left\{ \sum_{l} p_{lj}/m, \max_{l} \{p_{lj}\} \right\}.$$

Otherwise, job j would always be late.

We set $\delta_1 = 0.2, 0.4, 0.6, 0.8, 1.0$ and $\delta_2 = 0.2, 0.4, 0.6, 0.8, 1.0$. For each combination of δ_1 and δ_2, 100 instances are generated. Thus, there are 2500 instances in total. The algorithms are implemented in Java. The running environment is Windows 7 64-bit Operating System running on a dual core (2.50GHz + 2.50GHz) PC with 4GB RAM memory.

To compare the algorithms, for each generated instance I_i $(i = 1, 2, \ldots, 100)$, we also construct the corresponding single-machine instance I_i' as described in Lemma 2.2. The instance I_i' is solved optimally by Moore-Hodgson's algorithm, and then the result, denoted by $LB(I_i')$, is used as a reference objective value (lower bound) to evaluate the objective value produced by a heuristic algorithm A for I_i, denoted by $\sum_j U_j(A, I_i)$. Table III shows the collective results for all six algorithms. Each algorithm A has two columns, namely $\bar{\varepsilon}$ and $\bar{t}$, which are defined as follows. For each setting of δ_1 and δ_2, $\bar{\varepsilon}$ is defined for A as:

$$\bar{\varepsilon} = \frac{1}{100} \sum_{i=1}^{100} \left(\sum_{j} U_j(A, I_i) - LB(I_i') \right);$$

and let $t(A, I_i)$ denote the running time (in milliseconds) of algorithm A on solving instance I_i, $\bar{t}$ is defined for A as:

$$\bar{t} = \frac{1}{100} \sum_{i=1}^{100} t(A, I_i).$$

From the table, we have the following findings:

- The objective values produced by all six algorithms are actually very close to the lower bound values, the gaps are mostly less than 3, which means that the algorithms perform close to an optimal algorithm for these randomly generated instances.
- The frequencies that the six algorithms achieve the lowest $\bar{\varepsilon}$ are $(5, 9 \mid 11, 14 \mid 13, 22)$ corresponding to their order listed in the table. Thus, in terms of Machine Choosing Criterion, the algorithms based on Best-Fit criterion performs better than those based on First-Fit criterion, which in turn are better than those based on Smallest-Load criterion. In terms of Task Sorting Criterion, the algorithms based on non-increasing task sorting criterion perform better than those without task sorting.
- In terms of running time, the First-Fit based algorithms run faster than those based on Best-Fit criterion, which in turn run faster than algorithms based on Smallest-Load criterion.
- Interestingly, task sorting in initialization actually does not increase the running time, but helps reduce the running time. This could be due to that task sorting helps produce better results and hence results in less iterations for Step 2 and Step 3.
- Regarding the sensitivity of algorithms' performance to the hardness of problem instances, overall, $\bar{\varepsilon}$ increases when δ_2 increases. The explanation is that higher δ_2 results in tighter due dates generated for the instances. Hence, the number of late jobs is expected to be higher, and the gap between the heuristic result and lower bound result is expected to increase accordingly. On the other hand, $\bar{t}$ also increases when δ_2 increases. Indeed, when the number of late jobs increases with higher δ_2, more iterations are required by Step 2 and Step 3 of the algorithms.

The above findings are sufficient to give us an overview of the performance of the algorithms and provide guidelines for us to choose the best ones among them for practical use. Considering both solution quality and running time, we recommend that algorithm GS_{BFD} is the best choice.

VI. CONCLUSIONS

In this paper, we studied the problem of minimizing the total number of late multi-task jobs on identical and flexible machines in parallel. We first investigated the complexity aspect of the problem. As summarized in Table IV, complexity

TABLE III
COMPARISON OF THE SIX ALGORITHMS IN TERM OF $\overline{\varepsilon}$ AND $\overline{t}$

		GS_{LS}		GS_{LPT}		GS_{FF}		GS_{FFD}		GS_{BF}		GS_{BFD}	
δ_1	δ_2	$\overline{\varepsilon}$	$\overline{t}$	$\overline{\varepsilon}$	$\overline{t}$	$\overline{\varepsilon}$	$\overline{t}$	$\overline{\varepsilon}$	$\overline{t}$	$\overline{\varepsilon}$	$\overline{t}$	$\overline{\varepsilon}$	$\overline{t}$
0.2	0.2	0.56	60	0.47	55	0.46	40	0.45	40	0.44	41	0.45	40
0.2	0.4	0.57	170	0.46	155	0.46	121	0.47	115	0.47	123	0.47	118
0.2	0.6	0.65	283	0.57	258	0.56	210	0.56	199	0.56	215	0.56	204
0.2	0.8	0.65	360	0.51	331	0.51	277	0.51	264	0.51	285	0.51	271
0.2	1.0	3.03	379	2.39	352	1.77	295	1.6	284	1.53	304	1.51	292
0.4	0.2	0.83	389	0.83	365	0.83	302	0.83	296	0.83	312	0.83	304
0.4	0.4	0.67	479	0.53	446	0.52	365	0.52	356	0.52	377	0.52	366
0.4	0.6	0.53	583	0.41	542	0.41	446	0.41	436	0.41	461	0.41	447
0.4	0.8	0.82	657	0.51	612	0.51	508	0.47	501	0.48	526	0.47	513
0.4	1.0	3.03	680	2.54	637	1.83	530	1.65	526	1.53	550	1.5	538
0.6	0.2	0.0	687	0.0	648	0.0	536	0.0	536	0.0	556	0.0	549
0.6	0.4	0.57	745	0.5	703	0.5	576	0.5	577	0.5	598	0.5	591
0.6	0.6	0.58	835	0.49	785	0.49	643	0.48	647	0.48	667	0.48	662
0.6	0.8	2.05	888	1.51	836	1.21	687	1.05	695	1.06	713	1.01	710
0.6	1.0	3.14	911	2.71	862	2.06	709	1.88	719	1.72	736	1.69	735
0.8	0.2	0.0	919	0.0	873	0.0	715	0.0	729	0.0	743	0.0	746
0.8	0.4	0.62	931	0.55	888	0.54	724	0.53	742	0.53	753	0.53	759
0.8	0.6	0.73	987	0.5	942	0.5	766	0.5	789	0.5	797	0.49	807
0.8	0.8	2.22	1018	1.63	974	1.3	791	1.19	818	1.16	824	1.13	836
0.8	1.0	2.95	1036	2.42	994	1.88	807	1.71	837	1.62	841	1.61	856
1.0	0.2	0.0	1043	0.0	1005	0.0	813	0.0	847	0.0	848	0.0	867
1.0	0.4	0.0	1050	0.0	1016	0.0	819	0.0	857	0.0	855	0.0	878
1.0	0.6	1.37	1061	0.9	1031	0.68	829	0.55	870	0.62	865	0.57	892
1.0	0.8	2.24	1072	1.64	1045	1.18	838	1.08	883	1.05	876	1.02	906
1.0	1.0	2.89	1083	2.34	1058	1.79	847	1.67	896	1.53	886	1.49	919

results were established for some cases that are either NP-hard or polynomially solvable. Due to the NP-hardness of the general case, we then investigated its approximability. Unfortunately, the result was negative, as we showed that, unless $P = NP$, there exists no ρ-approximation algorithm ($1 < \forall\rho < \infty$) even for the case with no release dates. Thus, we designed a general algorithm scheme and derived from it six heuristic algorithms whose performance was evaluated by experimental results. The findings from the experimental results provided guidelines for choosing the best algorithm among them for practical use, and we recommended algorithm GS_{BFD} as the best choice.

We did not consider setup times, preemption and weights for the problem. It will be interesting to study the problem with these additional constraints. Even for release dates, we only considered the single machine cases. Hopefully, the heuristics presented in this paper can be extended to the parallel machine cases with release dates. We did not consider an exact algorithm in this paper either. It seemed that the design of an exact algorithm with intelligent search of an optimal solution is not trivial at all, even though we looked into some properties and derived a lower bound for optimal schedule. Indeed, although it has been shown that there exists an optimal schedule which complies with the EDD rule. However, the subproblem to assign the individual tasks to the parallel machines is NP-hard. This not only makes it hard for the design of an exact algorithm with intelligent search, but also makes it non-trivial for the design of effective local search heuristics or meta-heuristics. All of these are worthy of further research for the problem.

REFERENCES

[1] J.-T. Leung, H. Li, and M. Pinedo, "Order scheduling models: an overview," in *Multidisciplinary Scheduling: Theory and Applications*, G. Kendall, E. K. Burke, S. Petrovic, and M. Gendreau, Eds. Springer, 2005, pp. 37–53, http://dx.doi.org/10.1007/0-387-27744-7_3.

[2] J. Blocher and D. Chhajed, "The customer order lead-time problem on parallel machines," *Naval Research Logistics*, vol. 43, pp. 629–654, 1996, http://dx.doi.org/10.1002/(SICI)1520-6750(199608)43:5<629::AID-NAV3>3.0.CO;2-7.

[3] J. Bruno, E. Coffman, and R. Sethi, "Scheduling independent tasks to reduce mean finishing time," *Communications of the ACM*, vol. 17, no. 7, pp. 382–387, 1974, http://dx.doi.org/10.1145/361011.361064.

[4] J. Yang, "Scheduling with batch objectives," Ph.D. dissertation, Industrial and Systems Engineering Graduate Program, The Ohio State University, Columbus, Ohio, 1998.

[5] J. Yang and M. Posner, "Scheduling parallel machines for the customer order problem," *Journal of Scheduling*, vol. 8, no. 1, pp. 49–74, 2005, http://dx.doi.org/10.1007/s10951-005-5315-5.

[6] J.-T. Leung, H. Li, and M. Pinedo, "Approximation algorithms for minimizing total weighted completion time of orders on identical machines in parallel," *Naval Research Logistics*, vol. 53, no. 4, pp. 243–260, 2006, http://dx.doi.org/10.1002/nav.20138.

[7] J. Moore, "An n job, one machine sequencing algorithm for minimizing the number of late jobs," *Management Science*, vol. 15, pp. 102–109, 1968, http://dx.doi.org/10.1287/mnsc.15.1.102.

[8] M. Garey and D. Johnson, *Computers and Intractability: A Guide to the Theory of NP-completeness*. New York: W.H.Freeman, 1979.

[9] J. Lenstra, A. R. Kan, and P. Brucker, "Complexity of machine scheduling problems," *Annals of Discrete Mathematics*, vol. 1, pp. 343–362, 1977, http://dx.doi.org/10.1016/S0167-5060(08)70743-X.

[10] E. Lawler, J. Lenstra, A. R. Kan, and D. Shmoys, "Sequencing and scheduling: algorithms and complexity," in *Handbooks in Operations Research and Management Science*, 1993, pp. 445–522.

[11] P. Brucker, *Scheduling Algorithms, Fifth Edition*. Berlin: Springer, 2007.

[12] M. Pinedo, *Scheduling: Theory, Algorithms, and Systems*. Springer, 2008.

[13] E. Lawler, "A dynamic programming algorithm for preemptive scheduling of a single machine to minimize the number of late jobs," *An-*

TABLE IV
COMPLEXITY RESULTS

Problem	Complexity
$1 \mid jbs, tsk, r_j \mid \sum U_j$	NP-hard in the strong sense
$Pm \mid jbs, tsk \mid \sum U_j$	NP-hard in the ordinary sense for $m \geq 2$
$P \mid jbs, tsk \mid \sum U_j$	NP-hard in the strong sense
$1 \mid jbs, tsk \mid \sum U_j$	Solvable by Moore-Hodgson's algorithm
$1 \mid jbs, tsk, r_j, d_j = d \mid \sum U_j$	Solvable in $O(n \log n + \sum k_j)$ time
$1 \mid jbs, tsk, r_j, p_{lj} = 1 \mid \sum U_j$	Solvable by Lawler's algorithm in $O(n^5)$ time
$P \mid jbs, tsk, p_{lj} = p \mid \sum U_j$	Solvable in $O(n \log(n) + \sum k_j)$ time

nals of Operations Research, vol. 26, no. 1, pp. 125–133, 1990, http://dx.doi.org/10.1007/BF02248588.

[14] G. Ausiello, P. Crescenzi, G. Gambosi, V. Kann, A. Marchetti-Spaccamela, and M. Protasi, *Complexity and Approximation: Combinatorial Optimization Problems and Their Approximability Properties*. Springer, 1999.

[15] C. Papadimitriou and M. Yannakakis, "Optimization, approximation, and complexity classes," in *Journal of Computer and System Sciences*, Vol. 43(3), pp. 425–440, 1991, http://dx.doi.org/10.1016/0022-0000(91)90023-X.

[16] R. Graham, "Bounds on multiprocessing timing anomalies," *SIAM Journal of Applied Mathematics*, vol. 17, pp. 263–269, 1969, http://dx.doi.org/10.1137/0117039.

[17] E. Coffman, M. Garey, and D. Johnson, "Approximation algorithms for bin packing: a survey," in *Approximation Algorithms for NP-hard Problems*, D. Hochbaum, Ed. PWS Publishers, 1997, pp. 46–93.

[18] J. Jackson, "Scheduling a production line to minimize maximum tardiness," Management Science Research Project, UCLA, Tech. Rep. 43, 1955.

[19] J.-T. Leung, H. Li, and M. Pinedo, "Scheduling orders for multiple product types with due date related objectives," *European Journal of Operational Research*, vol. 168, no. 2, pp. 370–389, 2006, http://dx.doi.org/10.1016/j.ejor.2004.03.030.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 585–590
DOI: 10.15439/2016F273
ACSIS, Vol. 8. ISSN 2300-5963

A new polynomial class of cluster deletion problem

Sabrine Malek
University of Sfax - Tunis
Faculty of Economics and Management of Sfax
Email: sabrine.malek@gmail.com

Wady Naanaa
University of Monastir - Tunis
Faculty of Sciences of Monastir
Email: wady.naanaa@gmail.com

***Abstract*—Cluster Deletion (CD) problem asks to transform a given graph into a cluster graph by at most k edge deletions. CD is a combinatorial problem arising in the field of classification. In this paper, we introduce a graph transformation which enabled the identification of new polynomially solvable classes of CD problem. We show that if a graph is K_3-free or (*diamond*, *kite*, *house*, *xbanner*)-free then cluster deletion problem can be solved in polynomial time on that graph.**

***Index Terms*—graph clustering; cluster deletion; Line Graph; P_3 adjacency graph; forbidden patterns; diamond graph; claw graph; fork graph**

I. Introduction

CLASSIFICATION is the problem of identifying to which of a set of categories a new observation belongs [1]. In the terminology of machine learning, classification may be supervised or unsupervised. The corresponding unsupervised procedure is known as clustering which is considered the most important unsupervised learning problem [2]. On the other hand, many combinatorial problems are modelled using graphs, in particular, the partitioning of graph vertices into clusters is a classification task that may be used to better manage many real-world problems. From the theoretical point of view, the clustering task is closely related to partitioning problems. As every other problem of this kind, clustering aims to finding structures or patterns in a collection of unlabeled data. The goal is to partition these elements into subsets called clusters such that two meta-criteria are satisfied: homogeneity (elements in a same cluster should be highly similar to each other) and separation (elements from different clusters may have low similarity to each other). In the graph theoretic approach to clustering, data are often represented in the form of a graph G. Ideally, the resulting graph would be a cluster graph, that is, a graph in which every connected component is a clique, i.e., a complete subgraph. From the practical point of view, clustering algorithms can be applied in many fields, for instance in social networks, in Wireless sensor network (WSN) [3], in particular in optimizing energy distribution between access points [4], [5], or in designing electronic integrated circuits [6]...

In this paper, we deal with a specific version of the graph clustering problem, namely, cluster deletion (CD), which allows a graph partitioning, into a set of complete subgraphs, just by removing edges. This problem is known to be NP-hard [7] for general graphs. However, it may become easier and polynomial-time solvable in specific graphs, for instance split graphs, block graph, proper interval graph, cographs [8], [9]. Graph classes on which CD is polynomial-time solvable can also be specified by forbidding the occurrence of certain (small) subgraphs in the input graph. For instance, CD is polynomial-time solvable on a sub-class of P_4-sparse graphs that strictly includes P_4-reducible graphs (which are, in turn, a superclass of cographs) [10]. Those results were obtained for unweighted graphs. For weighted graphs, the cluster deletion problem can be solved in polynomial time on the class of K_3-free graphs for which the CD equivalent to maximum weighted matching [8], [11].

On the other hand, there are several works showing that CD problem is NP-hard on some subclasses of weighted graphs such, (C_5, P_5)-free graphs,($2K_2$, $3K_1$)-free graphs and (C_5, P_5, bull, 4-pan, fork, co-gem, co-4-pan)-free graphs [9].

Our aim is to derive polynomial subproblems of *CD* by resorting to graph transformation. In the literature, there are several graph transformation, among them the widely used line graph [12]. A graph H is a line graph of a graph G if the vertices of H are in a one-to-one correspondence with the edges of G, with two vertices being adjacent in H if and only if the corresponding edges of G have a vertex in common. Clearly, any graph matching in G corresponds to an independent set in its line graph, and therefore, the maximum independent set problem in the class of line graphs is equivalent to the maximum matching problem in general graphs. What is worth noting here is that finding a maximum matching in a graph is a polynomial problem [13], which implies that the maximum independent set problem in polynomially solvable in line graphs [14]. For this reason this graph transformation has been widely used both for reducing and solving the maximum independent set problem. Another transform, known as conic reduction [15], transforms a graph G into a graph G' with $\alpha(G') = \alpha(G) - 1$, where $\alpha(G)$ denotes the maximum cardinality of a maximum independent set in a graph G, that is, the stability number of G. Yet another interesting transformation is the one based on the removal of simplicial vertices [16]. Let v be a vertex and let $N(v)$ be the neighbours set of v. v is simplicial if $N(v)$ is a clique. For each simplicial vertex v, we have $\alpha(G) = \alpha(G \setminus N[v]) + 1$ holds such that $N[v] = N(v) \cup \{v\}$. Thus, given a simplicial vertex, it is easy to reduce the problem of determining $\alpha(G)$ to the same problem on a smaller graph. There are other graph transformation, such as clique reduction [17], C-reduction [18], graph reduction for QoS Prediction [19],

978-83-60810-90-3/$25.00

SWR reduction [20]. All these transformations may simplify combinatorial problems on graphs.

In the present paper, we introduce a new transformation called the P_3-*adjacency graph* and we use it to identify new polynomially-solvable classes of *CD*.

In the next section, we present a new proof of the tractability of the CD problem for unweighted K_3-free graphs which is much easier that the one proposed in [11]. Indeed, we prove that when the initial unweighted graph *G* is K_3-free, its $P_3(G)$ will correspond to its line graph and then finding a minimum deletion edge-set is equivalent to finding a maximum independent set in $P_3(G)$. In other words, solving the *CD* problem on a K_3-free graph amounts to finding a maximum matching in the line graph of *G*. Next, we show that if *G* is diamond-free then any maximum independent set of $P_3(G)$ provides a solution of *CD*. Secondly, we introduce a new collection of forbidden patterns, namely *kite*, *house*, *open-envelope* and *xbanner*, and prove that $P_3(G)$ is claw-free when *G* is (*kite*, *house*, *open-envelope*, *xbanner*)-free. This enables a polynomial computation of a maximum independent set of $P_3(G)$, and then, provides an optimal solution for *CD* on *G* in polynomial-time.

II. Definitions and notations

A graph is a mathematical structure consisting of a set of vertices and a set of edges connecting the vertices. There are several types of graphs, among which, one can distinguish simple graphs, which are defined by an ordered pair (V, E), where V is a finite set of vertices and $E \subseteq \mathcal{P}_2(V)$ is the set of edges, with $\mathcal{P}_2(V)$ being the set of all pairs of V. From a simple graph $G = (V, E)$, one can extract a partial graph $G_p = (V, E_p)$ obtained by removing some of the edges of *G*, we have therefore $E_p \subseteq E$.

Definition 1. *Let $G = (V, E)$ be a simple graph and let $U \subseteq V$. The simple graph $(U, E(U))$ is the sub-graph of G induced by U, where $E(U) = E \cap \mathcal{P}_2(U)$*

A sub-graph of a given simple graph is therefore defined as follows:

Definition 2. *A graph $G_s = (V_s, E_s)$ is the sub-graph of a graph $G = (V, E)$ if there exists $U \subseteq V$ such that G_s is the sub-graph of G induced by U i.e V_s = U and $E_s = E(U)$.*

A complete graph is a simple graph in which every pair of distinct vertices is connected by a unique edge. The complete graph with n vertices is denoted by K_n. A clique of a simple graph $G = (V, E)$ is a complete subgraph of *G*. The K_3 graph is the complete graph with three vertices and a P_3 graph is the path on three vertices as it is illustrated in Fig. 1.
Observe that a complete graph cannot contain any P_3 as an induced subgraph.

Let C be a collection of small graphs, that will be designated by patterns. *G* is said to be C-*free* if *G* contains no member of C as an induced subgraph.

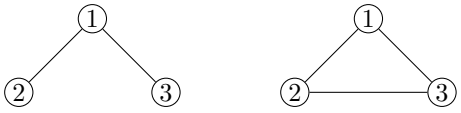


Fig. 1. A P_3 graph (left) and K_3 graph (right)

Besides, in the literature there exists a known class of graph called line graph which represents the adjacencies between the edges of a given graph.

Definition 3. *The line graph of a simple graph G= (V,E) is the graph $L(G)$= (V',E'), where:*

- *Each vertex of $L(G)$ represents an edge of G; and*
- *two vertices of $L(G)$ are adjacent if and only if their corresponding edges share a common endpoint in G.*

Example 1. *Fig 2. shows a simple graph and its line-graph.*

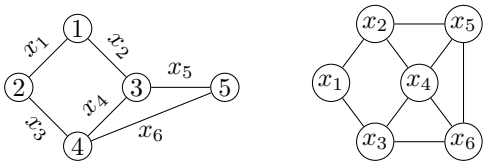


Fig. 2. Graph *G* (left) and its line graph L(G) (right)

The relevance of the line graph class is that many combinatorial problems, that are NP-hard on general graphs, are polynomially solvable on line graphs. The clustering problem is one of these combinatorial tasks. It consists in making the fewest changes to the set of edges of an input graph in order to obtain a set of cliques. There are three variations of graph clustering: cluster completion, cluster deletion and cluster editing. In the graph completion variations, edges can only be added. In cluster deletion, edges can only be deleted. In cluster editing, both edge additions and edge deletions are allowed. More formally, the cluster deletion (*CD*) problem consists in finding, for a given graph *G*, a P_3-free partial graph. An optimal solution for a *CD* instance given by a simple graph $G = (V, E)$ is a P_3-free partial graph $G_p = (V, E_p)$ of *G* that minimizes $|E - E_p|$.

III. A polynomial CD class

In this paper, we precisely consider the cluster deletion problem which allows a graph partitioning just by removing edges. This version partitions the graph into a set of complete subgraphs, i.e., cliques, and the goal is to remove the fewest edges from the input graph. We propose an algorithm that solves CD by resorting to a new graph transformation, which is defined as follows:

Definition 4. *The P_3 adjacency graph of a simple graph G= (V,E) is a simple graph, which we denote by $P_3(G)$, such that:*

- *Each vertex of $P_3(G)$ represents an edge of G; and*

- *two vertices of $P_3(G)$ are adjacent if and only if their corresponding edges form a P_3 sub-graph of G.*

Example 2.

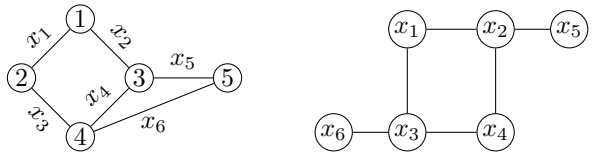


Fig. 3. Graph *G* (left) and its P_3(G) graph (right)

Lemma 1. *$P_3(G) = L(G)$ if and only if G is K_3-free*

Proof. Let $G = (V, E)$ be a simple graph with the associated P_3 adjacency graph $P_3(G) = (E, E_{P_3})$ and line graph $L(G) = (E, E_L)$.
($\Rightarrow$) Assume that $P_3(G) = L(G)$ while *G* contains a K_3 as an induced subgraph, and proceed to get a contradiction.

By Definition 3, a K_3 graph formed by edges x, y, z of *G* will be transformed into a K_3 in $L(G)$ whose edges are $\{x, y\}$,$\{x, z\}$,$\{y, z\}$. On the other hand, since any K_3 of *G* cannot contain a P_3, by Definition 4, x, y and z will not be connected in $P_3(G)$. We deduce that $E_L \neq E_{P_3}$, which contradicts our assumption. So, $P_3(G) = L(G)$ cannot hold true unless *G* is K_3-free.

($\Leftarrow$) Suppose that *G* is K_3-free and show that $P_3(G) = L(G)$, which amounts to establishing that $E_{P_3} = E_L$ since $P_3(G)$ and $L(G)$ have the same vertex-set by definition. So, let $\{x, y\}$ be any edge of E_{P_3}. According to Definition 4, the edges of *G* that correspond to vertices x and y in $P_3(G)$ must form an induced P_3 subgraph in *G*, and then, they must share a common endpoint. This implies that, $\{x, y\} \in E_L$. Thus, we have $E_{P_3} \subseteq E_L$. Conversely, let $\{x, y\}$ be in E_L, which implies that the edges x and y must share a common endpoint in *G*. Moreover, since *G* is K_3-free, it cannot contain a third edge which form a K_3 subgraph with x and y. This implies that x and y form a P_3 in *G*, which implies that $\{x, y\} \in E_{P_3}$. Thus, $E_L \subseteq E_{P_3}$. It follows, that $P_3(G) = L(G)$. □

Example 3. *Consider the graph G_1 depicted in Fig. 4. Its $L(G_1)$ and $P_3(G_1)$ are as shown in Fig. 5.*

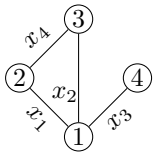


Fig. 4. A graph that contains both P_3 and K_3.

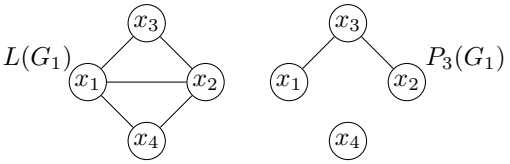


Fig. 5. The line graph and the P_3 adjacency graph of G_1.

We notice that $L(G_1)$ and $P_3(G_1)$ are different. This is due to the K_3 subgraph induced by vertices $1, 2$ and 3.

Example 4. *The figure below represents a graph G_2 composed of four vertices and containing two P_3 but no K_3:*

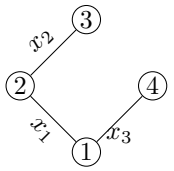


Fig. 6. A simple graph containing two P_3 but no K_3.

Fig. 7 represents the $L(G_2)$ associated with $P_3(G_2)$.

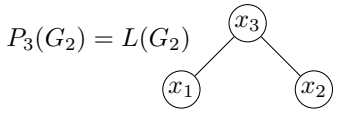


Fig. 7. The line graph and P_3 adjacency graph of G_2

Transforming G_2 into $L(G_2)$ and $P_3(G_2)$ gives the same simple graph because G_2 is K_3-free. This observation is in accordance with Proposition 1.

Lemma 2. *Let $G = (V, E)$ be a K_3-free graph. If $E' \subseteq E$ is a maximum independent set of $P_3(G)$ then (V, E') is a CD solution for G.*

Proof. ($\Leftarrow$) Assume that (V, E') is a *CD* solution for *G*. Since (V, E') is composed of a set of cliques and a clique cannot contain any P_3 as an induced subgraph, the vertices of $P_3(G)$ that correspond to the edges of E' will be pairwise not connected. This means that they form an independent set of $P_3(G)$.

It remains to prove that E' is a maximum independent set of $P_3(G)$. To this end, assume the converse is true, that is, there exists $E'' \subseteq E$ such E'' is an independent set of $P_3(G)$ and $|E''| > |E'|$. Since E'' is an independent set of $P_3(G)$, then, by Lemma 1, the partial graph of *G* defined by (V, E'') is P_3-free. Moreover, we have $|E - E''| < |E - E'|$ since $|E''| > |E'|$. This implies that (V, E') is not a CD solution for G which contradicts the hypothesis.

($\Rightarrow$) Let E' be a maximum independent set of $P_3(G)$. Assume that (V, E') is not a solution of *CD* for *G* and proceed to get a contradiction.

If (V, E') is not a solution of the *CD* instance defined by G then either (V, E') contains a P_3 or there exists a P_3-free partial graph (V, E'') such that $|E - E''| < |E - E'|$. By Lemma 1, the first case cannot hold true. In turn, the second case cannot hold because, otherwise, E' will not be a maximum independent set of $P_3(G)$ since $|E''| > |E'|$. □

Combining Lemma 1 and 2, we deduce that

Theorem 1. *The CD problem limited to the class of K_3-free graphs can be solved in polynomial time for simple graphs.*

Proof. According to Lemma 2, solving a *CD* instance defined by G amounts to finding a maximum independent set in $P_3(G)$. On the other hand, according to Lemma 1, if G is K_3-free, then $P_3(G)$ is identical to its line graph. In addition, it is well established that the maximum independent problem is polynomial on the class of line graphs. It follows that the *CD* problem for K_3-graphs can be solved by first constructing a P_3 adjacency graph , and then, by computing, in polynomial time, a maximum independent set in the latter graph. Finally, since the P_3 adjacency graph can be built in $O(|V|^4)$ step, the overall process is polynomial. □

In what follows, we identify a wider tractable class of *CD*, which is also defined via forbidding certain graph patterns.

Lemma 3. *Let $G = (V, E)$ be a diamond-free graph and let $E' \subseteq E$. (V, E') is a CD solution for G if and only if E' is a maximum independent set of $P_3(G)$.*

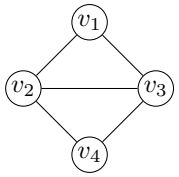


Fig. 8. Diamond graph

Proof. ($\Leftarrow$) Let E' be a maximum independent set of $P_3(G)$. Assume that (V,E') is not a *CD* solution for G. This implies that either (V, E') contains P_3 as an induced subgraph or there is a P_3-free partial graph (V, E'') such that $|E''| > |E'|$. Assure the former case, i.e., (V, E') contains a P_3, say u-v-w, and proceed to get a contradiction. Denote by e, the edge $\{u, v\}$ and by e', the edge $\{v, w\}$. Since e and e' are in E' and $E' \subseteq E$, which is a maximum independent set of $P_3(G)$, the latter two edges must not form a P_3 in (V, E). On the other hand, e and e' form a P_3 in (V, E'). This implies that $\{u, w\}$ is in E but not in E'. This cannot occur unless there is another edge $\{x, u\} \in E'$ that forms a P_3 with $\{u, w\}$ but neither with e nor with e'. Moreover, since $\{x, u\}$ and $\{u, w\}$ form a P_3, $\{x, w\}$ must not be in G. It follows that $(\{u, v, w, x\}, E(\{u, v, w, x\}))$ is an induced diamond subgraph of G, which contradicts the hypothesis. In the latter case, there is a P_3-free partial graph (V, E'') such that $|E''| > |E'|$. This implies that E' is not a maximum independent set of $P_3(G)$, which also contradicts the hypothesis.

($\Rightarrow$) Let (V,E') be a CD solution for G and assume that E' is not a maximum independent set of $P_3(G)$, then proceed to get a contradiction.

Since E' is not a maximum independent set of $P_3(G)$ then either E' is not an independent set of $P_3(G)$ or it is not maximum. From the definition of $P_3(G)$, the former case implies that there exist $e, e' \in E'$ that form a P_3 in G, which contradicts the fact (V, E') is a *CD* solution for G. The latter case implies that there exists a maximum independent set $E'' \subseteq E$ such that $|E''| > |E'|$. Using ($\Leftarrow$), we deduce that (V, E') is not a *CD* solution for G and this contradicts the hypothesis. □

The complete bipartite graph $K_{1,3}$ is known as the *claw* graph. It is illustrated by Fig. 9. In what follows, the goal

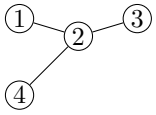


Fig. 9. A complete bipartite graph $K_{1,3}$ (Claw graph or Y graph)

is to obtain a $P_3(G)$ free from claw. Recall that this allows to polynomially solve the maximum independent set problem in $P_3(G)$. Thus, interested by determining the patterns which entail a claw in $P_3(G)$.

Lemma 4. *Let $G = (V, E)$ be a simple graph. $P_3(G)$ has a claw as an induced subgraph if and only if G contains one of the graphs in Fig. 10 as an induced subgraph.*

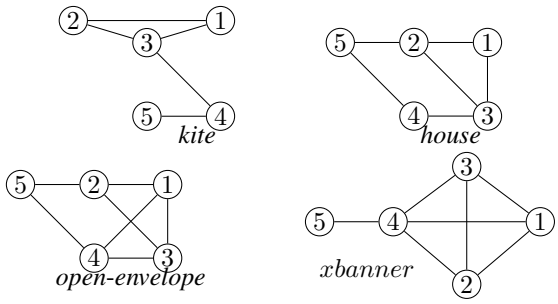


Fig. 10. The set of minimal graphs, which when they are present in G as subgraphs entail a claw in P_3(G) [21]

Proof. ($\Rightarrow$) Let G be a simple graph and let $P_3(G)$ be the P_3 adjacency graph of G. Assume that $K_{1,3}$ is a subgraph of $P_3(G)$ and, by the same time, G does not contain any *kite*, *house*, *open-envelope* and *xbanner* as an induced subgraph and proceed to get a contradiction. By referring to Definition 4 and since $K_{1,3}$ is composed by a vertex which has three not connected neighbors, if $P_3(G)$ contains a $K_{1,3}$ then G must contain three P_3 subgraphs sharing a common edge, say a. So, we should have in G one of the following two subgraphs:

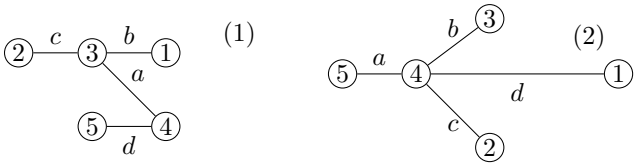


Fig. 11. Minimal patterns that entail a claw in $P_3(G)$

To ensure that a claw graph (Fig. 9) occurs in the P_3 adjacency graph of G, we add, in all possible manners, the edges that keep the claw graph as an induced subgraph of $P_3(G)$. This results in the following four patterns, which correspond exactly to the forbidden patterns of Fig 12 and contradicts the hypothesis.

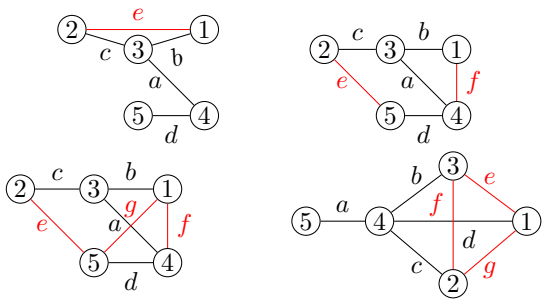


Fig. 12. After change: the set of minimal graphs, which when they are present in G as subgraphs, entail a claw in $P_3(G)$

($\Leftarrow$) If G contains a *kite*, *house*, *open-envelope* or *xbanner* as an induced subgraph, then $P_3(G)$ will respectively contain one of the following graphs as an induced subgraph (see Fig. 13, Fig. 14, Fig. 15).

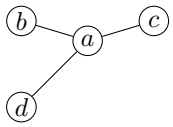


Fig. 13. P_3(kite) and P_3(xbanner)

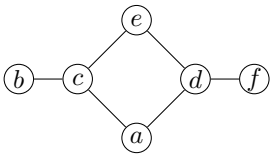


Fig. 14. P_3(house)

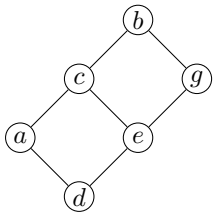


Fig. 15. P_3(open-envelope)

Observe that all these subgrahs contain a $P_{1,3}$ as an induced subgraph. □

Theorem 2. *The CD problem is polynomial in the class of (kite, house, xbanner, diamond)-free graphs.*

Proof. Let G be a (*kite*, *house*, *xbanner*, *diamond*)-free graph. Since *diamond* is an induced subgraph of *open-envelope* and by Lemma 4, $P_3(G)$ is claw-free. On the other hand, by Lemma 3, a maximum independent set of $P_3(G)$ corresponds to a *CD* solution for G. Since maximum independent set problem can be solved in polynomial time in claw-free graphs [12], *CD* can be solved in polynomial time in (*kite*, *house*, *xbanner*,*diamond*)-free graphs. □

IV. Conclusion

In this paper, we identified polynomial classes of the *CD* problem. We introduced a new graph transformation, namely, the P_3 adjacency graph. We used this transformation in order to show that whenever a simple graph G is diamond-free, any maximum independent set of $P_3(G)$ provides a solution of CD. Next, we showed that *CD* problem can be solve in polynomial time in K_3-free and (*kite*, *house*, *xbanner*,*diamond*)-free unweighted graphs.

References

[1] M. Azad and M. Moshkov, "Classification and optimization of decision trees for inconsistent decision tables represented as MVD tables," in *2015 Federated Conference on Computer Science and Information Systems, FedCSIS 2015, Lódz, Poland, September 13-16, 2015*, 2015, pp. 31–38. [Online]. Available: http://dx.doi.org/10.15439/2015F231

[2] R. Ziembinski, "Unsupervised extraction of graph-stream structure for purpose of knowledge retrieval and information fusion," in *Position Papers of the 2015 Federated Conference on Computer Science and Information Systems, FedCSIS 2015, Lódz, Poland, September 13-16, 2015.*, 2015, pp. 53–60. [Online]. Available: http://dx.doi.org/10.15439/2015F288

[3] C.-T. Cheng, C. K. Tse, and F. C. M. Lau, "A clustering algorithm for wireless sensor networks based on social insect colonies," *IEEE Sensors Journal*, pp. 711–721, 2010. [Online]. Available: http://dx.doi.org/10.1109/JSEN.2010.2063021

[4] C. S. Nam, Y. S. Han, and D. R. Shin, "Multi-hop routing-based optimization of the number of cluster-heads in wireless sensor networks," *Sensors Journal*, pp. 2875–2884, 2011. [Online]. Available: http://dx.doi.org/10.3390/s110302875

[5] N. Amini, A. Vahdatpour, W. Xu, M. Gerla, and M. Sarrafzadeh, "Cluster size optimization in sensor networks with decentralized cluster-based protocols." *Computer Communications*, pp. 207–220, 2012. [Online]. Available: http://dx.doi.org/10.1016/j.comcom.2011.09.009

[6] J. Cong and S. K. Lim, "Edge separability-based circuit clustering with application to multilevel circuit partitioning," *IEEE Trans. on CAD of Integrated Circuits and Systems*, pp. 346–357, 2004. [Online]. Available: http://dx.doi.org/10.1109/TCAD.2004.823353

[7] R. Shamir, R. Sharan, and D. Tsur, "Cluster graph modification problems," *Discrete Applied Mathematics*, pp. 173–182, 2004. [Online]. Available: http://dx.doi.org/10.1016/j.dam.2004.01.007

[8] F. Bonomo, G. Durán, and M. Valencia-Pabon, "Complexity of the cluster deletion problem on subclasses of chordal graphs," *Theor. Comput. Sci.*, pp. 59–69, 2015. [Online]. Available: http://dx.doi.org/10.1016/j.tcs.2015.07.001

[9] Y. Gao, D. R. Hare, and J. Nastos, "The cluster deletion problem for cographs," *Discrete Mathematics*, pp. 2763–2771, 2013. [Online]. Available: http://dx.doi.org/10.1016/j.disc.2013.08.017

[10] F. Bonomo, G. Durán, A. Napoli, and M. Valencia-Pabon, "A one-to-one correspondence between potential solutions of the cluster deletion problem and the minimum sum coloring problem, and its application to P_4-sparse graphs," *Inf. Process. Lett.*, pp. 600–603, 2015. [Online]. Available: http://dx.doi.org/10.1016/j.ipl.2015.02.007

[11] J. Edmonds, "Maximum matching and a polyhedron with 0, 1-vertices," *Journal of Research of the National Bureau of Standards B*, pp. 125–130, 1965. [Online]. Available: http://archive.org/details/jresv69Bn1-2p125

[12] V. V. Lozin and M. Milanic, "A polynomial algorithm to find an independent set of maximum weight in a fork-free graph," *J. Discrete Algorithms*, pp. 595–604, 2008. [Online]. Available: http://dx.doi.org/10.1016/j.jda.2008.04.001

[13] J. Edmonds, "Paths, trees, and flowers," *Canad. J. Math.*, pp. 449–467, 1965. [Online]. Available: http://dx.doi.org/10.4153/CJM-1965-045-4

[14] C. Berge, "Two theorems in graph theory," *Proceedings of the National Academy of Sciences of the United States of America*, pp. 842–844, 1957. [Online]. Available: http://www.jstor.org/stable/89875

[15] V. V. Lozin, "Conic reduction of graphs for the stable set problem," *Discrete Mathematics*, pp. 199–211, 2000. [Online]. Available: http://dx.doi.org/10.1016/S0012-365X(99)00408-2

[16] A. Brandstädt and P. L. Hammer, "On the stability number of claw-free P_5-free and more general graphs," *Discrete Applied Mathematics*, pp. 163–167, 1999. [Online]. Available: http://dx.doi.org/10.1016/S0166-218X(99)00072-4

[17] L. Lovàsz and M. D. Plummer, *Matching theory*. Amsterdam, New York: North-Holland, 1986. [Online]. Available: http://opac.inria.fr/record=b1086098

[18] P. A. Catlin, "A reduction method for graphs," in *Proc. 19th Southeastern Conf. (Baton Rouge), Congressus Numerantium 65*, 1988, pp. 159–170.

[19] A. Goldman and Y. Ngoko, "On graph reduction for qos prediction of very large web service compositions," in *IEEE Ninth International Conference on Services Computing, Honolulu, HI, USA*, 2012, pp. 258–265. [Online]. Available: http://dx.doi.org/10.1109/SCC.2012.21

[20] J. Cardoso, A. P. Sheth, J. A. Miller, J. Arnold, and K. Kochut, "Quality of service for workflows and web service processes," *J. Web Sem.*, pp. 281–308, 2004. [Online]. Available: http://dx.doi.org/10.1016/j.websem.2004.03.001

[21] H. N. de Ridder *et al.*, "Information System on Graph Classes and their Inclusions (ISGCI)," http://www.graphclasses.org.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 591–599
DOI: 10.15439/2016F213
ACSIS, Vol. 8. ISSN 2300-5963

A New Approach to the Discretization of Multidimensional Scaling

W. Gramacho*†, A. Mucherino†, J-H. Lin‡, C. Lavor§

*Federal University of Tocantins, Palmas-TO, Brazil.
wgramacho@uft.edu.br

†IRISA, University of Rennes 1, Rennes, France.
antonio.mucherino@irisa.fr, warley.gramacho-da-silva@irisa.fr

‡Research Center for Applied Sciences, Academia Sinica, Taiwan.
jhlin@gate.sinica.edu.tw

§IMECC-UNICAMP, Campinas-SP, Brazil.
clavor@ime.unicamp.br

***Abstract*—Given a set of points in a Euclidean space having dimension $K > 0$, we are interested in the problem of finding a realization of the same set in a Euclidean space having a lower dimension. In most situations, it is not possible to preserve all available interpoint distances in the new space, so that the best possible realization, which gives the minimal error on the distances, needs to be searched. This problem is known in the scientific literature as the Multidimensional Scaling (MDS). We propose a new methodology to discretize the search space of MDS instances, with the aim of performing an efficient enumeration of their solution sets. Some preliminary computational experiments on a set of artificially generated instances are presented. We conclude our paper with some future research directions.**

I. Introduction

GIVEN a set of points X in a Euclidean space $\mathbb{R}^K$, with $K > 0$, Multidimensional Scaling (MDS) consists in finding a realization of X in $\mathbb{R}^k$, with $0 < k < K$, such that the interpoint distances in $\mathbb{R}^K$ are preserved as much as possible [3]. The *initial* dimension K is generally very large, while the new dimension k is generally a priori unknown. However, for a fixed *destination* dimension k, the MDS can be seen as a particular class of the Distance Geometry Problem (DGP) [16]. In fact, in the DGP, suitable embeddings of a given simple weighted undirected graph $G = (V, E, d)$ are searched, in a way that the distances between embedded vertices u and $v \in V$ are as close as possible to the weights on the edges $(u, v) \in E$, when available. In the MDS, the graph G can be simply deduced from the original set of points in $\mathbb{R}^K$. A *valid embedding* is an embedding of G satisfying all distance constraints, with a given tolerance.

In recent works, it was shown that the DGP can be discretized when some particular assumptions are satisfied [17]. The discretization makes it possible to work with a finite search space, which is otherwise continuous. This search space has the structure of a tree, which is binary when all available distances can be considered as exact. The DGP is NP-hard [27], and even if its discretization does not reduce its complexity [14], it allows for employing a Branch & Prune (BP) framework for an ad-hoc exploration of the discretized search space [14], [24]. The present work is a preliminary step for the *discretization* of the MDS. In this work, in order to mainly focus on the problem discretization, we will suppose that the topology of the embeddings in both dimensions K and k can be represented well by considering all real inter-point distances.

We warn the reader that a previous work, devoted to the discretization of the MDS, was already published in [1]. However, differently from that work, we will consider, in our first analysis, the basic MDS without any additional constraints. This decision was taken with the aim of developing, first of all, an efficient procedure for the discretization of the MDS, to be extended thereafter for tackling more complex problems. In fact, differently from [1], the algorithms for discretized MDS that we propose in this paper are particularly tailored to this special class of problems.

The rest of the paper is organized as follows. In Section II, we will focus on previous works for the discretization of the DGP. Then, in Section III, we will make a parallel between the DGP and the MDS, while trying to extend and adapt the methodologies, already developed for the DGP, for the discretization of the MDS. Our computational experiments will be presented in Section IV. Finally, Section V will conclude the paper.

II. Distance Geometry Problem

The Distance Geometry Problem (DGP) consists in embedding a simple weighted undirected graph $G = (V, E, d)$ in a k-dimensional space so that all weights d_{uv} on the edges of G are realized as distances between the positions assigned to its vertices [16]. More formally, the DGP asks whether

978-83-60810-90-3/$25.00

it is possible to find an embedding $x : V \to \mathbb{R}^k$ satisfying constraints based on the available edge weights:

$$\forall (u,v) \in E, \quad ||x_u - x_v|| = d_{uv}.$$

Notice that, in these distance constraints, we can obtain the equation of a hyper-sphere in $\mathbb{R}^k$ by fixing one of the two vertices in a certain position. The discretization of the DGP is based on the idea to intersect as many hyper-spheres as necessary to obtain a discrete set of possible positions for a given vertex $v \in V$ (see Section II-A).

A classical approach to the DGP is to reformulate it as an unconstrained global optimization problem [23]. The satisfaction of the constraints based on the distances can be measured by computing the difference between the actual value $||x_u - x_v||$ of the distance, and its expected value d_{uv}. In order to verify the overall satisfaction of the available constraints, a penalty function can be introduced, whose general term is related to the generic constraint. Various penalty functions can be defined for the DGP, and one of the most used penalty functions is the so-called Medium Distance Error (MDE):

$$MDE(x) = \frac{1}{|E|} \sum_{(u,v)\in E} \frac{|\,||x_u - x_v|| - d_{uv}\,|}{d_{uv}}. \qquad (1)$$

Finding the global minimum of this penalty function allows to obtain solutions to the DGP. If all distances are compatible to each other and they are not affected by numerical errors, the MDE value for a valid embedding x needs to be equal to zero.

There are two main applications of the DGP that can be commonly found in recent publications. One application arises in biology, and it is concerned with the identification of protein conformations by exploiting some known interatomic distances that can be obtained by experimental techniques [19]. Another common application is given by the so-called Sensor Network Localization Problem (SNLP) [2], [5], where the positions of the sensors forming a network need to be identified.

As already mentioned above, when some particular assumptions are satisfied, the DGP can be discretized, so that its search space can be reduced from a continuous (and infinite) domain to a discrete (and finite) domain. We will give some details about the discretization of the DGP in Section II-A, and we will focus on discretization orders (vertex orders for G which make the discretization assumptions satisfied) in Section II-B.

A. The Discretization

Let $G = (V, E, d)$ be a simple weighted undirected graph representing a DGP instance. Let $G[\cdot]$ be the subgraph of G induced by a subset of vertices in V; let $\mathcal{V}_S(\cdot)$ be the volume of the simplex defined by the vertices given as arguments. The discretization in dimension $k > 0$ of a DGP instance can be performed when there exists an order defining sequence $r : i \in \mathbb{N} \longrightarrow v \in V \cup \{\diamondsuit\}$ of length $|r|$ (for which $r_i = \diamondsuit$ if $i > |r|$) such that the following two assumptions are satisfied:

Algorithm 1 The BP algorithm.

```
1: BP(i, k, G, r, ε)
2: let v = r_i;
3: compute x'_v in dimension k;
4: if (x'_v is feasible with tolerance ε) then
5:    if (i = |r|) then
6:       print new solution;
7:    else
8:       BP(i + 1, k, G, r, ε);
9:    end if
10: end if
11: compute x''_v in dimension k;
12: if (x''_v is feasible with tolerance ε) then
13:    if (i = |r|) then
14:       print new solution;
15:    else
16:       BP(i + 1, k, G, r, ε);
17:    end if
18: end if
```

(a) $G[\{r_1, r_2, \ldots, r_k\}]$ is a clique;
(b) $\forall i \in \{k+1, \ldots, |r|\}$, there exist k "reference" vertices, i.e. there exist $j_1, j_2, \ldots, j_k$ such that
   1) $j_1 < i$, $j_2 < i$, …, $j_k < i$;
   2) $\{(r_{j_1}, r_i), (r_{j_2}, r_i), \ldots, (r_{j_k}, r_i)\} \subset E$;
   for which
$$\mathcal{V}_S(r_{j_1}, r_{j_2}, \ldots, r_{j_k}) > 0.$$

Vertex orders satisfying assumptions (a) and (b) are named *discretization orders*: more details about these special orders are given in Section II-B. The class of DGP instances for which at least one discretization order exists for the corresponding graph is named Discretizable DGP (DDGP) [24].

The search space of DDGP instances is finite. It has the structure of a tree, where nodes contain candidate vertex positions, organized layer by layer. In fact, assumption (a) allows us to place the first k vertices given by the vertex order r in k fixed positions. In this way, we can avoid to consider congruent solutions that can be obtained by rotating and/or translating other solutions. Assumption (b) ensures the existence of at least k *reference vertices* for every vertex having a rank $i > k$: a vertex u can play the role of "reference" for a given vertex v if u precedes v in the vertex ordering r, and $(u, v) \in E$. We say that the corresponding distances d_{uv} are the *reference distances* for the vertex v. We exploit the k reference vertices for defining k spheres centered in the reference vertices and having as radius the corresponding reference distances: in dimension k, the intersection of these k spheres provides us with a subset of vertex positions having cardinality 2 [8], [14]. The condition on the volume of the simplex in assumption (b) ensures that this sphere intersection does not provide a full circle (which is obviously not discrete).

We employ a Branch & Prune (BP) algorithm [25] for the solution of DDGP instances (see Alg. 1 for a sketch). The algorithm recursively calls itself for the exploration of the

search tree. In the algorithm call, i is the current rank in the given vertex order r: it is supposed that the coordinates of all vertex positions on the current branch are stored in the global memory. The value of k corresponds to the dimension in which we wish to embed the graph G. ε is our tolerance for the errors that can affect the generated vertex coordinates: as soon as new vertex positions are computed by performing the sphere intersection, in fact, their feasibility is verified by exploiting additional distances that were not used in the discretization process. In order to do so, we consider the corresponding terms of equation (1), and we declare as "infeasible" a new generated position if at least one of such terms is greater than the predefined tolerance ε. Once a new solution is found by the BP algorithm, the quality of such a solution can be verified by applying equation (1): since all its terms are smaller than our tolerance ε, we expect to obtain good-quality solutions.

The complexity of one single BP call corresponds to the complexity of computing twice a new vertex position, and of verifying whether the exploration of the tree should continue in those two directions or not. If, for example, x'_v is feasible (see Alg. 1), then this position could be part of a solution, and therefore the branch of the binary tree rooted at x'_v needs to be explored. In this case, the algorithm invokes itself for computing the possible positions for the next rank. Instead, if for example x''_v is not feasible with the given tolerance, then the current branch does not contain any solution. It is therefore pruned: the algorithm does not invoke itself in this case. This algorithm phase is fundamental and named *pruning phase*.

There are two main approaches to deal with the uncertainty on the values of the available distances. In [15], for example, intervals are used for representing such an uncertainty, and the terms of equation (1) are adapted for providing a positive measure only when the actual distance is not contained in the given interval (for both feasibility verification during the execution of BP, and computation of the MDE function once a new solution is found). However, in some applications (such as the MDS), the uncertainty on the distance values is not known a priori, and therefore our approach will consist in considering instances with only exact distances, which we will need to tackle by admitting larger values for the tolerance ε. In other words, since the range of the intervals to be assigned to uncertain distances cannot be predefined, we will try to define them during the execution of the BP algorithm: the tolerance ε represents in this case the maximal allowed interval range.

B. Vertex Orders

A *discretization order* is a vertex order associated to the graph G such that the discretization assumptions (a) and (b) are satisfied [9], [13]. Given a graph G representing a DGP instance, one question that can immediately arise is whether this instance belongs to the DDGP class or not. When instances are stored in text files, an order may be implicitly associated to the vertices of G, but such an order may not satisfy the discretization assumptions. In order to find suitable discretization orders, a greedy algorithm was proposed in [13] and extended subsequently for dealing with interval distances [20]. A heuristic, that has as worst-case complexity the one of the greedy algorithm, was also proposed in [10], for dealing more efficiently with very large instances.

In some particular applications (the reader can refer for example to [15]), additional assumptions on the discretization orders are required, and the problem of finding a discretization order may become NP-hard [4]. Naturally, the above-mentioned greedy algorithm cannot guarantee the generation of a vertex order that, apart from the discretization assumptions, could as well satisfy the additional ones.

For example, we say that a discretization order satisfies the *consecutivity assumption* if the reference vertices for a given vertex v, and v itself, are consecutive in the order. Orders satisfying the consecutivity assumption can be seen as sequences of overlapping cliques, which can be searched on the pseudo de Bruijn graphs introduced in [21]. Even if it implies the execution of an exponential search, the use of such pseudo de Bruijn graphs aided the identification of discretization orders satisfying the consecutivity assumption for the backbones of protein instances.

The complexity of this ordering problem can also increase when *optimal* discretization orders are required. We say that an order is optimal when the rank assigned to every vertex maximizes some predefined objective functions [22]. For example, when low-rank vertices have the maximal number of reference vertices, a direct impact on the width of the corresponding search tree can be observed: it becomes smaller. Several objectives, together with their priority orders, can be defined and used for finding optimal discretization orders. If the objectives are given by a set of simple functions, then the problem of finding an optimal discretization order (without any other additional assumption) still has a polynomial complexity, and the greedy algorithm can be extended for dealing with this class of vertex orders [9].

III. Multidimensional Scaling

Since one of the first pioneer papers on this topic [29] in 1952, the MDS received an increasing interest. Surveys on the MDS can be found in the scientific literature: a recent example is [11], published in 2013. In the cited survey, the MDS is defined as the set of statistical techniques that are employed for reducing the dimensionality of a given set of data, with the aim of improving the visual appreciation of the underlying relational structures contained therein. The main idea is to attempt mapping the data into a (generally Euclidean) space, where *similar* items are represented by near points in the mapping, and *dissimilar* items are represented by points that are located proportionally further apart. By doing so, the complexity in the data is reduced, and the primary dimensions, along with the items differ, are identified. The MDS has a wide range of applications, and the interested reader can refer to the citations in [11] for additional information about these applications.

One of the best-known methods for dimensionality reduction, especially in the field of mathematical statistics, is Principal Component Analysis (PCA) [7], [12]. PCA was

proposed by Karl Pearson in 1901 [26], and it consists of an orthogonal transformation that can project an ensemble of the high-dimensional data into a new set of coordinate systems (principal axes) in the order of the variances. One of the most notable advantages of PCA is the preservation of the distance metric during the linear transformation. However, this essential feature of PCA, i.e., the use of linear transformation, also prevents PCA from discovering intrinsic non-linear degrees of freedom underlying complex natural phenomena.

One of the most prominent approaches in non-linear dimensionality reduction is the *Isomap* method proposed by Tenenbaum et al [28]. Unlike PCA, Isomap method has a better ability for successfully capturing the intrinsic global geometric features of the given set of points. It was noted recently, however, that the application of the Isomap method is able to provide good results only when transferring the data between two similar geometries [6].

Let X be a set consisting of points of the Euclidean space $\mathbb{R}^K$. Since the coordinates for each point in X are known, it is possible to compute their relative distances. With this information, we can define a simple weighted undirected graph $G = (V, E, d)$. In the graph, every vertex $v \in V$ represents one single point in X, and an edge $(u, v) \in E$, together with the weight that is associated to it, represents the relative distance between the two corresponding points in X. All graphs G constructed in this way have the following two properties: they are complete, and all the weights associated to the edges (the distances) are exact. These graphs represent instances of the DGP.

A very simple but nice example of MDS is given in [11] and is concerned with the problem of drawing a small geographic map. All relative distances between the cities of Los Angeles, New York, Chicago and Dallas are given, and the aim is to find their correct locations on a two-dimensional map. When the information on the distances is precise, a very accurate map can be generated, i.e. a map for which the distances between points are proportional to the true distances between city pairs (modulo translations and rotations). In general, however, solutions where the overall distance information is precisely verified may not exist, so that approximated mappings need to be searched.

Let G be the simple weighted undirected graph defined as described above from a known set in dimension K. In the example of the 4 US cities, the distances are measured by approximating a surface on Earth with a plane containing the 4 cities: a small error is introduced in the distances because of the plane approximation of the Earth region. In general, every graph is embeddable in dimensions $k \geq |V|$ without introducing any error in the distances. However, when the number of vertices in V is large, the dimensionality is huge. The main interest therefore is to reduce the data dimensionality, and to converge to small dimensions, such as 3, 2, or even 1, where the visualization of the data is possible. In the example of the 4 US cities, the embedding of the corresponding graph G needs to be searched in dimension $k = 2$. Recall that K is the initial dimension of our MDS instances, while k is the destination dimension.

In this work, we will extend the concept of discretizability to MDS instances (see Section III-A), and we will provide two variants of the BP algorithm that are particularly tailored to the MDS (in both Sections III-A and III-B).

A. The Discretization

The methodology that we propose for discretizing the MDS is strictly related to the previous works on the discretization of the DGP (see Section II-A). The entire theory of the discretization can be in fact inherited almost unchanged: the novelties in this paper are mostly *operational*: we propose some ad-hoc strategies to be integrated in our new algorithms.

Since all relative distances are generally available in MDS instances, the main task that is required for performing the discretization is the selection, for every vertex $v \in V$ that does not belong to the initial clique, of a k-plet of reference vertices. We point out that this task has a fundamental importance. In fact, once a k-plet of reference vertices has been selected for the vertex v, the corresponding reference distances remain fixed in the search tree, i.e. we cannot allow the introduction of any error in the corresponding distances.

The main idea for our first variant on the BP algorithm for the MDS is to try to consider all possible k-plets of reference vertices, and to choose, at each recursive call, the k-plet for which the minimal error is observed during the pruning phase. If this minimal error is larger than a predetermined tolerance $\varepsilon > 0$, then the current branch of the tree is pruned and the search is backtracked.

In order to measure the quality of partial embeddings (up to the current rank of the discretization order), we need to consider the following function, that we name *partial* MDE (pMDE, see equation (1)):

$$pMDE(x, r, i) = \frac{1}{|E(r, i)|} \sum_{(u,v) \in E(r,i)} \frac{|\,||x_u - x_v|| - d_{uv}\,|}{d_{uv}}, \tag{2}$$

where

$$E(r, i) = \{(u, v) \in E \,|\, \exists j < i \,:\, u = r_j,\, v = r_i\}.$$

Notice that, while the vertex ordering associated to G is irrelevant for the computation of the MDE, it becomes important for the pMDE. The value of $pMDE(x, r, i)$ is in fact the average error introduced in the partial embedding x of G up to the rank i, in the vertex ordering r.

The sketch of our BP variant for the MDS is given in Alg. 2. The algorithm keeps the general structure of Alg. 1, but, every time it invokes itself recursively, it verifies all the possible k-plets of reference vertices for the current vertex v, and chooses the best k-plet p_{best} in terms of introduced error pMDE. We are aware that the choice of the k-plet is greedy, and that it might have, in theory, a negative impact on the computations. However, our computational experiments (see Section IV) show that our methodology works quite well in conjunction with the algorithm pruning phase, which does not allow the overall introduced error to grow more than desired.

Algorithm 2 A variant of the BP algorithm for the MDS.

BP$(i, k, G, r, \varepsilon)$
let $v = r_i$;
for (every k-plet p of reference vertices for v) **do**
  compute x'_v and pMDE(x'_v) in dimension k;
  compute x''_v and pMDE(x''_v) in dimension k;
end for
let p_{best} be the k-plet leading to the lowest pMDE;
let $\hat{x}'_v$ and $\hat{x}''_v$ be the two positions for v given by p_{best};
if $(pMDE(\hat{x}'_v, r, i) < \varepsilon)$ **then**
  if $(i = |r|)$ **then**
    print new solution;
  else
    BP$(i + 1, k, G, r, \varepsilon)$;
  end if
end if
if $(pMDE(\hat{x}''_v, r, i) < \varepsilon)$ **then**
  if $(i = |r|)$ **then**
    print new solution;
  else
    BP$(i + 1, k, G, r, \varepsilon)$;
  end if
end if

In comparison with Alg. 1, Alg. 2 has a higher complexity. The worst-case complexity is achieved (as in Alg. 1) when the pruning phase is never able to prune away infeasible tree branches. In this case, the algorithm needs to invoke itself

$$\sum_{i=k+1}^{|r|} 2^{i-k}$$

times. At each call, moreover, it is necessary to select the best k-plet of reference vertices in a set of $i - 1$ vertices, where i is the rank in r of the current vertex. The number of combinations of k objects among $i - 1$ objects, without repetitions and without assigning a specific order to them, are

$$\frac{(i-1)!}{k!(i-k-1)!}.$$

Therefore, the complexity of finding all these possible combinations depends upon the current layer of the tree i. As a consequence, the worst-case complexity of Alg. 2 is

$$\sum_{i=k+1}^{|r|} \frac{2^{i-k}(i-1)!}{k!(i-k-1)!}.$$

B. Vertex Orders

The reader might have remarked that, while vertex orders have been presented as a fundamental concept for the discretization of DGPs, there is only a quick mention to vertex orders in the previous section, devoted instead to the discretization of the MDS. In fact, the completeness of graphs G representing MDS instances ensures that all vertex orders allow for the discretization (the discretization assumptions in Section II-A are always satisfied). However, as already done for the DGP (see Section II-B), we may want to select optimal vertex orders for a given instance of the MDS, which may help us in improving the performances of the BP algorithm.

Algorithm 3 Another variant of BP for the MDS.

BP$(i, k, G, r, \varepsilon, \bar{V})$
for (every new candidate vertex $v \in \bar{V}$) **do**
  for (every k-plet p of reference vertices for v) **do**
    compute x'_v and pMDE(x'_v) in dimension k;
    compute x''_v and pMDE(x''_v) in dimension k;
  end for
end for
let $\bar{v} \in \bar{V}$ be the vertex leading to the lowest pMDE;
let $r_i = \bar{v}$;
let p_{best} be the k-plet leading to the lowest pMDE;
let $\hat{x}'_{\bar{v}}$ and $\hat{x}''_{\bar{v}}$ be the two positions for $\bar{v}$ given by p_{best};
if $(pMDE(\hat{x}'_{\bar{v}}, r, i) < \varepsilon)$ **then**
  let $\bar{V} = \bar{V} \setminus \{\bar{v}\}$;
  if $(\bar{V} = \emptyset)$ **then**
    print new solution;
  else
    BP$(i + 1, k, G, r, \varepsilon, \bar{V})$;
  end if
end if
if $(pMDE(\hat{x}''_{\bar{v}}, r, i) < \varepsilon)$ **then**
  let $\bar{V} = \bar{V} \setminus \{\bar{v}\}$;
  if $(\bar{V} = \emptyset)$ **then**
    print new solution;
  else
    BP$(i + 1, k, G, r, \varepsilon, \bar{V})$;
  end if
end if

In terms of partial orders, a vertex order that allows for the discretization of the MDS can be any order with an initial k-clique at the first rank, and all other vertices at rank two. In other words, once the initial clique has been embedded, as well as some other vertices until the vertex v, any of the remaining vertices can be embedded as next. This observation leaded us to develop another variant of the BP algorithm for the MDS.

Consider we have already embedded m vertices of our graph G, with $m > k$ and $m < |V|$. Instead of considering a predefined vertex ordering, we can assume here that every non-embedded vertex can be the candidate to be embedded as next. Moreover, for each of such candidate vertices, different k-plets can be selected to play the role of reference vertices in the discretization (see Section II-A). For every candidate vertex, and for every k-plet, an error on the pruning distances can be computed, and the best vertex and k-plet can be selected together. If the minimal obtained error is greater than our tolerance ε, then the current tree branch needs to be pruned.

The sketch of this second variant of the BP algorithm for the MDS is given in Alg. 3. In the algorithm call, the discretization order r is initially empty, and it is constructed

step by step during the several recursive calls. In this situation, backtracking also means changing vertex ordering, because every tree branch may admit a different optimal order. In practice, at each recursive call, the best vertex $\bar{v}$ and the best k-plet, which lead to the minimal pMDE value, are selected at the same time. This double choice is performed, as in Alg. 2, in a greedy manner. In the algorithm call, we also added the set $\bar{V}$, which is supposed to contain the vertices of the graph that have not been yet embedded.

The complexity of this algorithm is evidently higher, in comparison with Algs. 1 and 2. With respect to Alg. 2, an additional task needs to be performed at every recursive call of the algorithm: it is necessary to select the next vertex to be considered. This task requires a search over the remaining vertices to be embedded (with complexity $n - i$, where $n = |V|$) of the optimal one, i.e of the one for which it is possible to identify the best k-plet of reference vertices. Therefore, the overall worst-case complexity of Alg. 3 is

$$\sum_{i=k+1}^{|r|} \frac{2^{i-k}(|V|-1)(i-1)!}{k!(i-k-1)!},$$

which is much higher of the original complexity of Alg. 1.

IV. Computational experiments

A. Generation of MDS instances

We consider artificially generated instances, with $K > 3$ and $k = 3$. Let n be the number of points in the original set $X \subset \mathbb{R}^K$; let $e \in [0, 1]$. The procedure we employ for the instance generation consists in the following steps:

- we generate n random points in the cube of $\mathbb{R}^3$ with sides equal to 1;
- we add $K - 3$ coordinates to each point, having random values extracted from the interval $[0, e]$;
- we compute all relative distances between point pairs in $\mathbb{R}^K$;
- we define a simple weighted undirected graph G containing this distance information.

We point out that the error e introduced in these instances is distributed over the set of distances, as well as over the added coordinates, in a quite uniform manner. This property is unlikely to be satisfied by real instances in general. We also remark that the overall introduced error per instance naturally depends upon the value of e, but also on the distance size n.

B. Preliminary experiments

All the experiments presented in this section have been performed on a laptop computer equipped with an Intel Core i5 with 2.4 GHz and 8GB/1600 MHz RAM, running Mac OS X 10.11.3.

Table I shows some experiments where a set of instances that are embeddable in dimension 3 are considered (our introduced error e, in fact, is here set to 0). We fix the destination dimension to $k = 3$, while various initial dimensions K are taken into consideration. The ε value is set to 0.001 in all the experiments, even if $e = 0$, in order to deal with some errors introduced by imprecise computer arithmetics. In all generated instances, the total number of solutions is always 2 (when G is embeddable in dimension k, then the total number of solutions is 2, because all distances are available [18]). We point out however that, in general when an approximated embedding is searched, the introduced errors in the distances may lead to the identification of multiple solutions. The experiments lasting more than 2 hours were aborted. When executing Algs. 1 and 2, a vertex order was predefined. In Alg. 1, moreover, the selected reference vertices are always the k immediate preceding ones in the predefined ordering r. The computational time is measured in seconds.

The table shows that Alg. 2 is able to find better quality solutions w.r.t. the ones obtained by the standard Alg. 1, in terms of MDE. Such an improvement does not appear so evident instead when comparing Alg. 2 and Alg. 3: the MDE values remain mostly unchanged. It seems therefore that it is not important in which place of the ordering a vertex is located, but mostly which its reference vertices are. The computational times that we report reflect the theoretical worse-case complexities provided in Sections III-A and III-B. For instances with $n = 200$, Alg. 3 would have taken more than 2 hours to run to termination.

Table II shows some computational experiments where valid embeddings cannot be identified in the destination dimension k without introducing some errors in the distances. All considered instances have size $n = 20$, and part of the experiments already reported in Table I are here provided again for completeness. As the error e increases, the MDE values in the found solutions show that the quality of the solutions decreases as well. In some cases, Alg. 1 is not able to find any solution, because of the propagation of such errors on our search tree. Alg. 2 is instead always able to get the solutions, and the MDE reflects the initial introduced error e. Alg. 3 is always able to find solutions as well (the increase in the complexity is here not so important because all instances are small), and it finds better quality results in some cases.

Table III shows some experiments with our artificially generated instances having larger size ($n = 50$, 100 and 200), for larger errors (we consider $e \in \{0.01, 0.02\}$) and where we compare Alg. 2 to Alg. 3 only. The table shows that the quality of the solutions found by the two algorithms is comparable (in this set of experiments, only one time Alg. 3 was able to do better than Alg. 2), but the computational time grows too much when executing Alg. 3. In fact, the instances with $n = 200$ would have taken more than 2 hours, and these experiments were therefore aborted.

Finally, only Alg. 2 is considered in our last table of experiments. These experiments are aimed at stressing our algorithm with high initial dimensions; the destination dimension is still fixed to 3. As already pointed out, the errors introduced in our instances are quite uniformly distributed over the set of distances, and over the coordinates. This is the reason why we consider small errors e, but repeated uniformly over the n vertices forming the instance, and its $K - 3$ additional dimensions. In fact, as Table IV shows, the MDE values

TABLE I
COMPUTATIONAL EXPERIMENTS ON A SET OF INSTANCES THAT ARE EMBEDDABLE IN DIMENSION 3.

instance		Alg. 1		Alg. 2		Alg. 3	
n	K	MDE	time	MDE	time	MDE	time
20	4	10^{-15}	0.00	10^{-17}	0.02	10^{-17}	0.09
20	5	10^{-15}	0.00	10^{-16}	0.03	10^{-16}	0.14
20	8	10^{-14}	0.00	10^{-17}	0.03	10^{-17}	0.14
20	10	10^{-14}	0.00	10^{-17}	0.03	10^{-17}	0.14
50	4	10^{-14}	0.00	10^{-17}	2.94	10^{-17}	28.01
50	5	10^{-15}	0.00	10^{-16}	2.53	10^{-16}	25.77
50	8	10^{-15}	0.00	10^{-17}	2.84	10^{-17}	28.02
50	10	10^{-14}	0.00	10^{-17}	2.81	10^{-16}	28.34
100	4	10^{-13}	0.00	10^{-16}	71.42	10^{-17}	1444.88
100	5	10^{-14}	0.00	10^{-17}	75.23	10^{-17}	1388.16
100	8	10^{-15}	0.00	10^{-17}	68.90	10^{-17}	1035.28
100	10	10^{-13}	0.00	10^{-17}	74.68	10^{-17}	1041.27
200	4	10^{-13}	0.00	10^{-17}	4289.86	-	-
200	5	10^{-14}	0.00	10^{-17}	4272.32	-	-
200	8	10^{-13}	0.00	10^{-17}	2525.45	-	-
200	10	10^{-14}	0.01	10^{-17}	2880.03	-	-

TABLE II
COMPUTATIONAL EXPERIMENTS ON A SET OF SMALL INSTANCES FOR WHICH A VALID EMBEDDING CANNOT BE FOUND IN DIMENSION 3 WITHOUT INTRODUCING AN ERROR ON THE DISTANCES.

instance				Alg. 1		Alg. 2		Alg. 3	
n	K	e	ε	MDE	time	MDE	time	MDE	time
20	4	0	0.001	10^{-15}	0.00	10^{-17}	0.02	10^{-17}	0.09
20	5	0	0.001	10^{-15}	0.00	10^{-16}	0.03	10^{-16}	0.14
20	8	0	0.001	10^{-14}	0.00	10^{-17}	0.03	10^{-17}	0.14
20	10	0	0.001	10^{-14}	0.00	10^{-17}	0.03	10^{-17}	0.14
20	4	0.01	0.01	10^{-3}	0.00	10^{-5}	0.03	10^{-5}	0.14
20	5	0.01	0.01	10^{-3}	0.00	10^{-5}	0.03	10^{-5}	0.14
20	8	0.01	0.01	-	-	10^{-5}	0.03	10^{-5}	0.14
20	10	0.01	0.01	-	-	10^{-3}	0.03	10^{-3}	0.15
20	4	0.02	0.02	10^{-3}	0.00	10^{-3}	0.03	10^{-5}	0.16
20	5	0.02	0.02	10^{-3}	0.00	10^{-3}	0.05	10^{-3}	0.22
20	8	0.02	0.02	10^{-2}	0.00	10^{-3}	0.03	10^{-3}	0.14
20	10	0.02	0.02	-	-	10^{-2}	0.03	10^{-3}	0.15
20	4	0.03	0.03	10^{-2}	0.00	10^{-3}	0.05	10^{-3}	0.21
20	5	0.03	0.03	10^{-3}	0.00	10^{-3}	0.05	10^{-3}	1.09
20	8	0.03	0.03	10^{-2}	0.00	10^{-2}	0.05	10^{-2}	0.22
20	10	0.03	0.03	-	-	10^{-2}	0.03	10^{-2}	0.17

TABLE III
COMPUTATIONAL EXPERIMENTS ON A SET OF LARGER INSTANCES, WHICH ARE AFFECTED BY A LARGER OVERALL ERROR.

instance				Alg. 2		Alg. 3		instance				Alg. 2		Alg. 3	
n	K	e	ε	MDE	time	MDE	time	n	K	e	ε	MDE	time	MDE	time
50	4	0.01	0.01	10^{-5}	2.88	10^{-5}	27.75	50	4	0.02	0.02	10^{-3}	4.80	10^{-3}	44.49
50	5	0.01	0.01	10^{-3}	2.86	10^{-3}	28.10	50	5	0.02	0.02	10^{-3}	4.86	10^{-3}	44.89
50	8	0.01	0.01	10^{-3}	3.09	10^{-3}	29.67	50	8	0.02	0.02	10^{-2}	4.61	10^{-3}	44.25
50	10	0.01	0.01	10^{-3}	3.93	10^{-3}	36.48	50	10	0.02	0.02	10^{-3}	3.31	10^{-3}	32.31
100	4	0.01	0.01	10^{-5}	87.12	10^{-5}	1622.05	100	4	0.02	0.02	10^{-3}	98.19	10^{-3}	1786.72
100	5	0.01	0.01	10^{-3}	120.72	10^{-3}	2188.86	100	5	0.02	0.02	10^{-3}	97.27	10^{-3}	1797.52
100	8	0.01	0.01	10^{-3}	118.10	10^{-3}	2136.80	100	8	0.02	0.02	10^{-3}	130.74	10^{-3}	2360.76
100	10	0.01	0.01	10^{-3}	90.67	10^{-3}	1649.14	100	10	0.02	0.02	10^{-2}	97.31	10^{-2}	1870.78
200	4	0.01	0.01	10^{-5}	5521.94	-	-	200	4	0.02	0.02	10^{-3}	1988.79	-	-
200	5	0.01	0.01	10^{-3}	4407.33	-	-	200	5	0.02	0.02	10^{-3}	4415.64	-	-
200	8	0.01	0.01	10^{-3}	2589.42	-	-	200	8	0.02	0.02	10^{-3}	4049.64	-	-
200	10	0.01	0.01	10^{-3}	3017.70	-	-	200	10	0.02	0.02	10^{-2}	4846.58	-	-

TABLE IV
COMPUTATIONAL EXPERIMENTS ON A SET OF INSTANCES HAVING A HIGH INITIAL DIMENSION.

instance				Alg. 2	
n	K	e	ε	MDE	time
100	10	0.001	0.002	10^{-6}	145.77
100	20	0.001	0.002	10^{-6}	178.47
100	50	0.001	0.002	10^{-5}	142.68
100	100	0.001	0.002	10^{-5}	157.11
100	200	0.001	0.002	10^{-5}	157.61
100	300	0.001	0.002	10^{-3}	154.71
100	400	0.001	0.002	10^{-3}	132.22
100	500	0.001	0.002	10^{-3}	124.23
100	600	0.001	0.002	10^{-3}	120.62

in the found solutions strongly depend upon the difference between destination and initial dimensions. However, with a rather constant computational time, all the experiments were able to provide a good approximation of a valid embedding of G.

V. CONCLUSIONS

When the destination dimension k is fixed, the Multidimensional Scaling (MDS) can be seen as a DGP where an embedding of the graph G is available in dimension K, and it is necessary to find a valid embedding of the same graph in the given destination dimension. Actually, the graph G, which is the input of the DGP with the dimension k, can be generally obtained from the known embedding in dimension K. However, not all distances between vertices in dimension K can be realized in dimension k, and therefore approximated valid embeddings need to be searched.

This work represents the first step for the discretization of the MDS. The main ideas and the main methodology are inherited from previous works on the discretization of the DGP. We proposed two algorithms, that are variants on the BP algorithm, which was previously proposed for solving discretizable DGP instances. Our Alg. 2 seems to achieve the best trade-off between solution quality and increase in complexity w.r.t. the original BP algorithm.

The computational experiments that we have reported in this paper take into consideration a set of artificially generated instances, where the errors on the distances are introduced in a quite uniform manner. In order to deal with more realistic instances, there are several points to be improved in our approach. For example, the simplex inequalities (i.e. the triangular inequalities in dimension 3) need to be verified before executing our algorithms, and, when necessary, some distances may need to be corrected. Because of the nature of the problem, every corrected distance may imply to modify other distances accordingly, in order to avoid invalidating the geometry of the obtained embeddings.

From an algorithmic point of view, moreover, the branching phase of the algorithm can be improved by verifying, every time a given branch is pruned, whether other k-plets of reference vertices can be chosen for the vertices over the same branch. On the one hand, in fact, pruning allows us to focus the searches on the feasible parts of the tree, but with the risk, in presence of errors, to prune too much and obtain no solutions. On the other hand, however, branching over all the possible k-plets of reference vertices can be very expensive. It is necessary therefore to identify the best trade-off.

Finally, as a future work, we may also consider the possibility to let one of the k reference vertices to be related to an interval distance, as it was already done in works related to the discretization of the DGP. The decision not to consider yet interval distances during the discretization process is motivated by the fact that work is still in progress for an efficient management of interval distances during the generation of DGP discrete search spaces.

ACKNOWLEDGMENTS

The present work was entirely performed during WG's 1-year stay at IRISA, Rennes (France), which was funded by the Brazilian program "Ciencias sem Fronteiras". CL also wishes to thank FAPESP and CNPq for financial support.

REFERENCES

[1] J. Alencar, C. Lavor, T.O. Bonates, *A Combinatorial Approach to Multidimensional Scaling*, IEEE conference proceedings, 2014 IEEE International Congress on Big Data, 562–569, 2014.

[2] P. Biswas, T. Lian, T. Wang, Y. Ye, *Semidefinite Programming based Algorithms for Sensor Network Localization*, ACM Transactions in Sensor Networks **2**, 188–220, 2006.

[3] I. Borg, P.J.F. Groenen, *Modern Multidimensional Scaling: Theory and Applications*, Springer Series in Statistics, Second Edition, 355 pages, 2005.

[4] A. Cassioli, O. Günlük, C. Lavor, L. Liberti, *Discretization Vertex Orders in Distance Geometry*, Discrete Applied Mathematics **197**, 27–41, 2015.

[5] Y. Ding, N. Krislock, J. Qian, H. Wolkowicz, *Sensor Network Localization, Euclidean Distance Matrix Completions, and Graph Realization*, Optimization and Engineering **11**(1), 45–66, 2010.

[6] M.J. Duan, M.H. Li, L. Han, S.H. Huo, *Euclidean Sections of Protein Conformation Space and their Implications in Dimensionality Reduction*, Proteins **82**, 2585–2596, 2014.

[7] A.E. Garcia, *Large-Amplitude Nonlinear Motions in Proteins*, Physical Review Letters **68**, 2696–2699, 1992.

[8] D.S. Gonçalves, A. Mucherino, *Discretization Orders and Efficient Computation of Cartesian Coordinates for Distance Geometry*, Optimization Letters **8**(7), 2111–2125, 2014.

[9] D.S. Gonçalves, A. Mucherino, *Optimal Partial Discretization Orders for Discretizable Distance Geometry*, International Transactions in Operational Research **23**(5), 947–967, 2016.

[10] W. Gramacho, D.S. Gonçalves, A. Mucherino, N. Maculan, *A new Algorithm to Finding Discretizable Orderings for Distance Geometry*, Proceedings of Distance Geometry and Applications (DGA13), Manaus, Amazonas, Brazil, 149–152, 2013.

[11] M.C. Hout, M.H. Papesh, S.D. Goldinger, *Multidimensional Scaling*, Wiley Interdisciplinary Reviews: Cognitive Science **4**(1), 93–103, 2013.

[12] A. Kitao, F. Hirata, N. Go, *The Effects of Solvent on the Conformation and the Collective Motions of Protein − Normal Mode Analysis and Molecular-Dynamics Simulations of Melittin in Water and in Vacuum*, Journal of Chemical Physics **158**, 447–472, 1991.

[13] C. Lavor, J. Lee, A. Lee-St.John, L. Liberti, A. Mucherino, M. Sviridenko, *Discretization Orders for Distance Geometry Problems*, Optimization Letters **6**(4), 783–796, 2012.

[14] C. Lavor, L. Liberti, N. Maculan, A. Mucherino, *The Discretizable Molecular Distance Geometry Problem*, Computational Optimization and Applications **52**, 115–146, 2012.

[15] C. Lavor, L. Liberti, A. Mucherino, *The interval Branch-and-Prune Algorithm for the Discretizable Molecular Distance Geometry Problem with Inexact Distances*, Journal of Global Optimization **56**(3), 855–871, 2013.

[16] L. Liberti, C. Lavor, N. Maculan, A. Mucherino, *Euclidean Distance Geometry and Applications*, SIAM Review **56**(1), 3–69, 2014.

[17] L. Liberti, C. Lavor, A. Mucherino, N. Maculan, *Molecular Distance Geometry Methods: from Continuous to Discrete*, International Transactions in Operational Research **18**(1), 33–51, 2011.

[18] L. Liberti, B. Masson, J. Lee, C. Lavor, A. Mucherino, *On the Number of Realizations of Certain Henneberg Graphs arising in Protein Conformation*, Discrete Applied Mathematics **165**, 213–232, 2014.

[19] T.E. Malliavin, A. Mucherino, M. Nilges, *Distance Geometry in Structural Biology: New Perspectives*. In: "Distance Geometry: Theory, Methods and Applications", A. Mucherino, C. Lavor, L. Liberti, N. Maculan (Eds.), Springer, 329–350, 2013.

[20] A. Mucherino, *On the Identification of Discretization Orders for Distance Geometry with Intervals*, Lecture Notes in Computer Science **8085**, F. Nielsen and F. Barbaresco (Eds.), Proceedings of Geometric Science of Information (GSI13), Paris, France, 231–238, 2013.

[21] A. Mucherino, *A Pseudo de Bruijn Graph Representation for Discretization Orders for Distance Geometry*, Lecture Notes in Computer Science **9043**, Lecture Notes in Bioinformatics series, F. Ortuño and I. Rojas (Eds.), Proceedings of the 3^{rd} International Work-Conference on Bioinformatics and Biomedical Engineering (IWBBIO15), Granada, Spain, 514–523, 2015.

[22] A. Mucherino, *Optimal Discretization Orders for Distance Geometry: a Theoretical Standpoint*, Lecture Notes in Computer Science **9374**, Proceedings of the 10^{th} International Conference on Large-Scale Scientific Computations (LSSC15), Sozopol, Bulgaria, 234–242, 2015.

[23] J. Moré, Z. Wu, *Distance Geometry Optimization for Protein Structures*, Journal on Global Optimization **15**, 219–234, 1999.

[24] A. Mucherino, C. Lavor, L. Liberti, *The Discretizable Distance Geometry Problem*, Optimization Letters **6**(8), 1671–1686, 2012.

[25] A. Mucherino, L. Liberti, C. Lavor, `MD-jeep`*: an Implementation of a Branch & Prune Algorithm for Distance Geometry Problems*, Lectures Notes in Computer Science **6327**, K. Fukuda et al. (Eds.), Proceedings of the Third International Congress on Mathematical Software (ICMS10), Kobe, Japan, 186–197, 2010.

[26] K. Pearson, *On Lines and Planes of Closest Fit to Systems of Points in Space*, Philosophical Magazine **2**, 559–572, 1901.

[27] J. Saxe, *Embeddability of Weighted Graphs in k-Space is Strongly NP-hard*, Proceedings of 17^{th} Allerton Conference in Communications, Control and Computing, 480–489, 1979.

[28] J.B. Tenenbaum, V. de Silva, J.C. Langford, *A Global Geometric Framework for Nonlinear Dimensionality Reduction*, Science **290**, 290(5500), 2319-23, 2000.

[29] W.S. Torgerson, *Multidimensional Scaling: I. Theory and Method*, Psychometrika **17**(4), 401–419, 1952.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 601–604
DOI: 10.15439/2016F258
ACSIS, Vol. 8. ISSN 2300-5963

A Concept of Automatic Tuning of Longwall Scraper Conveyor Model

Piotr Przystałka, Andrzej Katunin
Silesian University of Technology,
Institute of Fundamentals of Machinery Design,
18a Konarskiego Street, 44-100, Gliwice, Poland,
Telephone: +48 32 237 10 69
Email: {piotr.przystalka, andrzej.katunin}@polsl.pl

***Abstract*—The modeling of machines and their operation modes is a key approach for optimization of their performance as well as for avoiding unwanted operational states which may lead to the occurrence of faults, and finally, to the breakdown. The developed model of a machine should be always parametrized, i.e. the certain number of parameters should be selected in the certain ranges. The most of the parameters can be selected based on engineering documentation and experts' knowledge, however, for some of them this knowledge cannot be directly acquired which leads to the parameter uncertainty. One of the approaches allowing selection of these uncertain parameters is a tuning procedure of a model. The paper deals with a concept of heuristic optimization method for automatic tuning of key parameters of longwall scraper conveyor model. In the first part of the paper, the evolutionary algorithm for tuning this model is proposed. In the case study, the merits and limitations of the evolutionary approach are analysed. The obtained results prove that the proposed approach of tuning of the considered model has high practical potential and it may be applied in real mining conditions.**

I. Introduction

THE longwall scraper conveyors are the machines used in the mechanized underground coal mines for transporting a coal. Since these machines usually work in extremely difficult operational conditions it is essential to monitor their performance in order to prevent unwanted operational modes and machinery downtime as well as to implement the knowledge obtained from the monitoring process into the re-designing processes which allows increasing their reliability, effectiveness and safety. However, considering the operational conditions in the underground coal mines as well as difficulties in physical access to sources of various signals and difficulties with their measurement, the monitoring of physical working parameters is often limited to few main quantities, which causes that the measurement data is incomplete, and makes the diagnosis and prognosis of these machines difficult. In order to predict an inappropriate behavior of a conveyor and prevent its unwanted operational modes, it is essential to develop a simulator (or mathematical/numerical model) which allows testing various operational scenarios, including the occurrence of various types of faults.

The model-based approach in diagnostics and condition monitoring of underground mines machinery is widely applicable in numerous industrial solutions. To date, many of such models were developed for scraper and belt conveyors. For the mentioned purposes, Cenacewicz, Przystałka and Katunin [1], [2] developed the models of belt and scraper conveyors for evaluation of their dynamical behavior under certain operational scenarios.

In the most cases, modeling of dynamic behavior of such machinery is difficult, since many operational parameters are not available or even not measurable or impossible to acquire. This leads to the incompleteness and uncertainty of a developed model. Thus, in order to achieve such parameters one can perform simplifications of the model, assume them basing on literature data and experts' knowledge, or use knowledge discovery and optimization techniques. Currently, the most common approach in this task is the manual adjustment of values of behavioural parameters of the model taking into account the data included in technical documentation or in domain literature as well as domain expert's knowledge. Obviously, such an approach is ineffective and leads to the increasing error with an increase of number of uncertain parameters and overall level of model uncertainty. On the other hand, in the recent years, heuristic methods based on the natural phenomena of evolution, such as simulated annealing algorithm, genetic algorithms, differential evolution, harmony or tabu search ideas, swarm-inspired methods, etc. have been developed and applied to model and solve real-life global optimization problems [3]. Furthermore, this kind of optimization algorithms has long been applied for tuning values of unknown parameters of different types of models [4], [5].

The problem of unknown parameter estimation is not enough discussed in many studies related to analytical modeling of mining conveyors, see e.g. [6] and [2]. To the best knowledge of the authors of this paper, this is one of the first attempts on applying a heuristic optimization technique for automatic tuning of parameters of longwall scraper conveyor model. The main goal of this study is the introduction of the new approach based evolutionary tuning of the mathematical

The research presented in the paper was financed by the National Centre of Research and Development (Poland) within the framework of the project titled "An integrated shell decision support system for systems of monitoring processes, equipment and hazards" carried out in the path B of Applied Research Programme - grant No. PBS2/B9/20/2013. This publication is financed from the statutory funds of the Faculty of Mechanical Engineering of the Silesian University of Technology in 2016.

model of the longwall scraper conveyor. This approach allows for adjusting the values of uncertain parameters of the model, and thus make it fully defined. This, in turn, allows for using this model for modeling of various scenarios of operation as well as for diagnosing the considered machine using the model-based diagnostics approach.

II. Concept of the evolutionary algorithm for tuning of conveyor model

The mathematical model of the considered scraper conveyor consists of the following submodels: model of a doubled main drive, model of auxiliary drive, model of mine breaker and contactor control, power supply model, model of equations of motion, model of masses, and model of motion resistance. A detailed mathematical and phenomenological description can be found in [2].

The dynamic behavior of the described model strongly depends on values of key parameters corresponding to physical properties of the real conveyor system. The total number of these parameters can be declared as D. As it is mentioned above, the most of them can be easily established in the direct way because there is the possibility for gauging and quantifying physical properties of the plant. On the other hand, the rest values declared as $\mathbf{x} = [x_1\ x_2\ \ldots\ x_d]$ can be indirectly found using signals of process variables which are collected during monitoring of the object and simulation of the model. We assume that the number of observed signals (real and simulated) is equal to J, whereas the number of experiments (scenarios) is I. Henceforth, real (r) and modeled (m) time series that are needed for tuning purposes may be denoted as

$$\mathbf{y}_r^{ij} = \left[y_r^{ij}(1)\ y_r^{ij}(2)\ \ldots\ y_r^{ij}(K)\right],$$

and

$$\mathbf{y}_m^{ij}(\mathbf{x}) = \left[y_m^{ij}(1,\mathbf{x})\ y_m^{ij}(2,\mathbf{x})\ \ldots\ y_m^{ij}(K,\mathbf{x})\right],$$

where j is the j-th signal (real or artificial) collected in the i-th experiment scenario, K is the number of samples.

The main objective of the tuning procedure is to adjust the values of the parameters x_1, x_2, …, x_d in order to obtain the smallest difference between the response of the system and the response predicted by the proposed model for each scenario. Therefore, the optimization problem can be written as follows:

$$\begin{array}{ll}\text{Minimize} & C(\mathbf{x}) = f\left[\mathbf{y}_r^{ij}, \mathbf{y}_m^{ij}(\mathbf{x}), I, J, K\right] \\ \text{subject to} & \Omega(\mathbf{x}),\end{array} \tag{1}$$

where f represents a function (with constant arguments I, J, K) for comparison of two time series, whilst Ω denotes boundaries and constraints in the optimization process. The optimal solution $\mathbf{x}^*$ is found if the criterion function C has a relative minimum value at $\mathbf{x} = \mathbf{x}^*$, that means if

$$\mathbf{x}^* = \arg\min_{\mathbf{x}\in\Omega} C(\mathbf{x}). \tag{2}$$

As to be expected, the criterion function C can be formulated in several ways. In this study, the authors propose two variants of this function. The first one is prepared applying the Minkowski distance of order p

$$C(\mathbf{x}, p) = \sum_{i=1}^{I}\sum_{j=1}^{J}\sqrt[p]{\sum_{k=1}^{K}\left|y_r^{ij}(k) - y_m^{ij}(k,\mathbf{x})\right|^p}. \tag{3}$$

This measure is a metric in a normed vector space which is considered as a generalization of both the Euclidean distance and the Manhattan distance. The second function is composed of three sub-criteria

$$C(\mathbf{x}, \mathbf{w}) = w_1 C_1(\mathbf{x}) + w_2 C_2(\mathbf{x}) + w_3 C_3(\mathbf{x}), \tag{4}$$

where w_1, w_2 and w_3 are used in order to control the significance of each component. The first criterion function in equation (4) is grounded on the mean absolute percent error and therefore it can be written as

$$C_1(\mathbf{x}) = \frac{100}{IJK}\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\left|\frac{y_r^{ij}(k) - y_m^{ij}(k,\mathbf{x})}{z^j}\right|, \tag{5}$$

where z^j corresponds to the range of the j-th sensor.

The second function is declared making use of cross-correlation as the base of a measure of the total lag τ between real and simulated signals

$$C_2(\mathbf{x}) = \tau = \sum_{i=1}^{I}\sum_{j=1}^{J}\tau^{ij}\left[R_{xy}\left(\mathbf{y}_r^{ij}, \mathbf{y}_m^{ij}(\mathbf{x})\right)\right]. \tag{6}$$

The last component is used in order to find the aggregate value of the maximum absolute errors which are present in signals collected under all scenarios

$$C_3(\mathbf{x}) = \sum_{i=1}^{I}\sum_{j=1}^{J}\max\left|\frac{\mathbf{y}_r^{ij} - \mathbf{y}_m^{ij}(\mathbf{x})}{z^j}\right|. \tag{7}$$

The solution of the optimization problem can be found using a limited number of strategies. Derivative-based approaches cannot be employed in this paper, mainly due to the form of the objective function C. Moreover, one can easily observe, that the return value of this function depends on the measurement noise in the real-world data as well as the virtual measurements (with simulated disturbances, noise and computing errors as well) obtained during numerical computations. In contrast, pure stochastic optimization methods, for example, Monte Carlo techniques will not be able to find an accurate solution with guaranteeing polynomial-time convergence because of the time of numerical computations of the conveyor model. Therefore, the authors decided to use the evolutionary algorithm which is known as one of the most common heuristic optimization methods.

III. CASE STUDY

A. Description of JOY BLS conveyor and its operating scenarios

The developed mathematical model was based on construction and parameters of the scraper conveyor of type JOY® BLS with the doubled and auxiliary drives. The parametrization of the described simulator was performed based on technical documentation of the modeled conveyor, and data available in the literature [7], [8]. The developed model is characterised by nearly fifty adjustable parameters. The operational parameters were determined theoretically or selected basing on the experts' knowledge.

The simulator of a scraper conveyor provides a possibility of simulation of eight operating scenarios, which represent the characteristic considering the operations performed during work such as idle run-up, idle run-up and run-down, etc. For the performed study four of them were selected due to the significant differences between these scenarios.

One should mention that some of operational parameters cannot be measured and were assumed according to literature data and technical data sheets. Therefore, it is essential to tune up the model in order to simulate its behavior during realization of considered scenarios properly. The five parameters ($d = 5$) that are subject to the tuning process in this study are as follows: the efficiency of the drive system: $x_1 = \eta$ [-]; the damping factor: $x_2 = \mu$ [-]; the torque losses of flexible and hydrodynamic couplings: $x_3 = \Delta M_p$ [Nm]; the approximation friction coefficient: $x_4 = a$ [-]; the unitary mass of excavated material: $x_5 = m_u$ [kg/m]. These parameters are selected since it is not possible to obtain their values in the direct way.

B. Tuning experiments and results

The verification tests were carried out with the assumption corresponding to the process variables collected during operational states of the machine. It was decided that only nine process variables ($J = 9$) could be used for tuning: the load of the engines $y_1 = M_{gn}$ [Nm]; the torque of the engines $y_2 = M_n$ [Nm]; the linear velocity of the chain $y_3 = v_n$ [m/s]; the phase currents of the first and second engine $y_4 = INZ1A$, $y_5 = INZ1B$, ..., $y_9 = INZ2C$ [A] (see [2] for details).

It was also assumed that the sampling rate of the sensors was equal to 500Hz, whereas the analog-to-digital converter resolution was set to 32bits. The noise powers of selected signals were as follows: $\sigma_{v_n} = 10\text{E}{-}08$, $\sigma_{Mgn} = \sigma_{Mpg} = 100$, $\sigma_I = 10\text{E}{-}02$. The reference signals $\mathbf{y}_r^{ij}$ were gathered while simulation of the model for selected scenarios ($I = 4$). The optimal values of parameters were chosen as follows: $x_1^r = 0.957$, $x_2^r = 0.1$, $x_3^r = 5$, $x_4^r = 3$ and $x_5^r = 461.54$. The time of the simulation was set to 40s. The error measure in the form of

$$\delta x = \frac{100\%}{d}\sum_{i=1}^{d}\delta x_i = \frac{100\%}{d}\sum_{i=1}^{d}\left|\frac{x_i^r - x_i^*}{x_i^r}\right|, \qquad (8)$$

was defined in order to evaluate the performance results.

The evolutionary algorithm implemented in Global Optimization Toolbox of Matlab® software was applied in this paper. For each optimization experiment, the boundary values of parameters were chosen taking into account literature data: $0.8 \le x_1 \le 0.99$, $0 \le x_2 \le 10$, $2 \le x_3 \le 8$, $2 \le x_4 \le 4$, $300 \le x_5 \le 570$. In the first step, the performance of the evolutionary algorithm was examined through analysing the influence of the variant of the criterion function C and the values of its parameters. The feasible population method was adapted to create a random well-dispersed initial population satisfying all constraints and bounds. Fitness scaling was done using the rank method, whereas the selection of the parents to the next generation was obtained by means of the stochastic uniform method. The elite count $\delta_s = 2$ and crossover fraction $p_c = 0.8$ were chosen. The heuristic crossover function was employed with the user-defined parameter $\lambda_h = 1.2$. The remaining individuals were mutated with the use of the adaptive feasible method. The population size was equal to 50, whilst the total number of generations was set to 20. Nine trials were performed in this part of the study: trials from 1 to 5 - fitness function was declared using Eq. (3) for $p = \{1, 2, \ldots, 5\}$; trials from 6 to 9 - fitness function was declared using Eq. (4) for $\mathbf{w} = [1\ 0\ 0]$, $\mathbf{w} = [0\ 1\ 0]$, $\mathbf{w} = [0\ 0\ 1]$, $\mathbf{w} = [0.9\ 0.01\ 0.09]$.

The results of experiments from this stage of the study are included in the first part of Table I. It can be stated that in the average sense, the minor errors can be achieved using the criterion function C in the form of Eq. 3. For this function, in each case beyond the 5th trial the mean error δx was close to 10%. Nevertheless, the smallest error was reached for the second criterion function that has been declared using Eq. 4 with $\mathbf{w} = [0.9\ 0.01\ 0.09]$. Hence, this variant of the fitness function was used in the next two steps.

In the second step, the authors analysed the influence of the population size and the crossover probability on the performance of the evolutionary algorithm. In order to examine this issue six experiments were conducted: trials 10 and 11 - fitness function was declared using Eq. (4) for $\mathbf{w} = [0.9\ 0.01\ 0.09]$ and the population size was equal to $\{30, 40\}$, the rest properties were the same as in the first step; trials from 12 to 14 - fitness function was declared using Eq. (4) for $\mathbf{w} = [0.9\ 0.01\ 0.09]$ and the crossover fraction was equal to $\{0.6, 0.7, 0.9\}$, the rest properties were the same as in the first step; trial 15 - fitness function was declared using Eq. (4) for $\mathbf{w} = [0.9\ 0.01\ 0.09]$, the population size was equal to 50 and the crossover fraction was equal to 0.7, the rest properties were the same as in the first step. This part of the research led us to state that, the smallest error could be achieved with the use of 20 individuals in the population, whereas the crossover fraction should be set to 0.7. It is easy to observe that, these settings can provide the mean error result smaller than 5%.

In the last part of the study, the analysis of the accuracy of the evolutionary optimization was carried out in the context of the sampling rate and resolution as well as the number

TABLE I: The errors and final values of parameters determined by the evolutionary optimization algorithm

Trial No.	x_1^* [−] δx_1 [%]	x_2^* [−] δx_2 [%]	x_3^* [Nm] δx_3 [%]	x_4^* [−] δx_4 [%]	x_5^* [kg] δx_5 [%]	δx [%]
1st step						
1	0.949 0.75	0.13 30.64	6.33 26.76	3.01 0.46	457.95 0.78	11.88
2	0.950 0.72	0.09 1.76	2.91 41.79	3.02 0.76	455.64 1.28	9.26
3	0.958 0.14	0.11 13.97	7.00 40.09	3.00 0.08	462.21 0.15	10.89
4	0.956 0.09	0.08 12.26	3.37 32.49	3.00 0.02	461.23 0.07	8.98
5	0.941 1.60	0.20 102.19	4.59 8.08	2.93 2.26	471.84 2.23	23.27
6	0.919 3.96	0.27 170.47	2.25 54.87	2.67 10.79	507.01 9.85	49.99
7	0.873 8.76	0.29 192.92	5.32 6.48	2.33 22.02	405.70 12.10	48.46
8	0.899 6.03	0.061 38.74	2.01 59.62	3.06 2.08	409.92 11.18	23.53
9	0.955 0.21	0.10 4.78	6.18 23.72	2.87 4.13	486.54 5.42	7.65
2nd step						
10	0.958 0.18	0.09 3.27	7.88 57.67	2.99 0.19	463.24 0.37	12.33
11	0.934 2.36	0.06 32.17	2.84 43.11	2.98 0.66	450.75 2.34	16.13
12	0.959 0.23	0.26 165.68	3.90 21.93	2.86 4.53	492.54 6.72	39.82
13	0.957 0.06	0.10 6.74	5.66 13.34	2.99 0.19	463.24 0.37	4.14
14	0.941 1.64	0.14 46.92	6.34 26.90	3.04 1.65	440.88 4.48	16.32
15	0.958 0.19	0.10 4.83	6.92 38.49	3.00 0.10	461.55 0.00	8.72
3rd step						
16	0.965 0.92	0.18 80.97	7.86 57.29	2.96 1.22	474.40 2.79	28.64
17	0.954 0.22	0.09 1.63	3.52 29.60	3.00 0.09	459.83 0.37	6.38
18	0.943 1.38	0.08 13.24	2.53 49.31	2.97 0.69	456.95 0.99	13.12

of the available sensors. The last three trials were done as follows: trial 16 - fitness function was declared using Eq. (4) for $\mathbf{w} = [0.9\ 0.01\ 0.09]$, only three sensors were used v_n, $INZ1A$ and $INZ2A$, the rest properties were the same as in the second step; trial 17 - fitness function declared using Eq. (4) for $\mathbf{w} = [0.9\ 0.01\ 0.09]$, all sensors were used, the sampling rate was equal to 1Hz, the resolution was set to 16bits, the rest properties were the same as in the second step; trial 18 - fitness function declared using Eq. (4) for $\mathbf{w} = [0.9\ 0.01\ 0.09]$, only three sensors were used v_n, $INZ1A$ and $INZ2A$, the sampling rate was equal to 1Hz, the resolution was set to 16bits, the rest properties were the same as in the second step.

The last analysis is very important from a technical point of view. It can be pointed out, that in mining engineering practice it is possible to use measuring devices with lower sampling rate and resolution for accurate tuning of a longwall scraper conveyor model. The sampling rate equals to 1Hz with the resolution equals to 16bits can be enough for this kind of problems to have the mean error significantly less than 10% provided that it involves the required number of measuring sensors.

Taking into account overall results of the study presented in Table I it can be concluded that the most important parameter is the damping factor. Even a small change in the value of this parameter can have a strong influence on the results of the simulation. On the other hand, the value of the loss of the torques of flexible and hydrodynamic couplings has almost no effect on the simulation.

IV. Conclusions

In this paper, the authors proposed and verified a concept of the heuristic optimization method for tuning values of parameters of longwall scraper conveyor model. In the considered model five of parameters were defined as uncertain, and the tuning problem was defined over these parameters. In turn, nine output parameters of the longwall scraper conveyor model were selected for tuning procedures. The tuning procedure was performed using four selected operational scenarios which were the most representative for the performed task.

The authors have formulated the optimization problem and applied the evolutionary computation in order to find the final solution. Two kinds of the criterion function was proposed. The first one was based on Minkowski's distance, whilst the second was prepared by means of weighted sub-criteria such as the mean absolute percent error, the cross-correlation function and the aggregate value of the maximum absolute errors. It was shown that the second type of the criterion function should be used in order to obtain the smallest mean errors during searching optimal values of parameters. The validation tasks confirm that the proposed tuning method is characterized by the high precision of tuning of uncertain parameters of the model, and may be successfully applied in real mining conditions.

References

[1] K. Cenacewicz and P. Przystałka, "Conveyor belt simulator with fault models," *Modelling in Engineering*, vol. 24, no. 55, pp. 13–20, 2015, [in Polish].

[2] K. Cenacewicz and A. Katunin, "Modeling and simulation of longwall scraper conveyor considering operational faults," *Studia Geotechnica et Mechanica*, vol. 38, no. 2, pp. 15–27, 2016. [Online]. Available: http://dx.doi.org/10.1515/sgem-2016-0015

[3] A. Mucherino and O. Seref, *Advances in Modeling Agricultural Systems*. Boston, MA: Springer US, 2009, ch. Modeling and Solving Real-Life Global Optimization Problems with Meta-heuristic Methods, pp. 403–419. [Online]. Available: http://dx.doi.org/10.1007/978-0-387-75181-8_19

[4] P. M. Vasant and P. M. Vasant, *Handbook of Research on Novel Soft Computing Intelligent Algorithms: Theory and Practical Applications*, 1st ed. Hershey, PA, USA: IGI Global, 2013.

[5] J. Valadi and P. Siarry, *Applications of Metaheuristics in Process Engineering*. Springer Publishing Company, Incorporated, 2014.

[6] M. Dolipski, E. Remiorz, and P. Sobota, "Determination of dynamic loads of sprocket drum teeth and seats by means of a mathematical model of the longwall conveyor," *Archives of Mining Sciences*, vol. 57, no. 4, pp. 1101–1119, 2012. [Online]. Available: http://dx.doi.org/10.2478/v10267-012-0073-7

[7] M. Dolipski, *Dynamics of chain conveyors*. Gliwice: The Silesian University of Technology Press, 1997, [in Polish].

[8] A. M. Plamitzer, *Electrical machines*, 4th ed. Warsaw: WNT, 1976, [in Polish].

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 605–608
DOI: 10.15439/2016F10
ACSIS, Vol. 8. ISSN 2300-5963

Facility Location Models forVehicle Sharing Systems

Alain Quilliot
LIMOS CNRS, Labex IMOBS3
Université Blaise Pascal
63000 Clermont-Ferrand, France

Antoine Sarbinowski
LIMOS CNRS, Labex IMOBS3
Université Blaise Pascal
63000 Clermont-Ferrand, France.

Email: alain.quilliot@isima.fr

***Abstract*—Designing a vehicle sharing system means locating stations which allow users to pick up and give back vehicles. One takes this strategic level decision while anticipating related rebalancing costs. We study here a strategic related bi-level *Vehicle Sharing Station Location* (VSSL) model, which involves as slave problem a static *Vehicle Sharing Rebalancing* (VSR) model.**

I. Introduction

Vehicle Sharing systems (see [4]), involving bikes or electric cars, are among the systems which currently strive in order to find their place in the urban mobility landscape as a compromise between full individual transportation and rigid public transportation. They most often work as one-way systems: customers should be allowed to pick up a *vehicle* at any station and give it back at any other station. But the system may fast become unbalanced, with either empty or overfilled stations, making arise two decision problems:

- a *strategic* level problem (see [4]) about the way stations are located and capacitated.
- an *operational* (or tactical) level problem (see [2, 3, 5, 6]), about the way the *Rebalancing Process* is performed.

This contribution deals with the *strategic level* problem, which has been scarcely studied, and which we refer to as the *Vehicle Sharing Station Location* (VSSL). We link it with the *operational level* known as the Vehicle Sharing Rebalancing Problem (VSRP).

Efficiently locating the *stations* of the system means:

- locating the stations close to the origins and destinations of the users, in such a way that a global *Access Demand* be maximized, or at least some target value be reached;
- minimizing investment and infrastructure costs;
- making in such a way that the expected running costs due to the periodic *Rebalancing Process* be the smallest possible.

While the two first criteria yield standard *Facility Location* models, (see [9]), dealing with the last one leads us to explicit those expected running costs due to the periodic *Rebalancing Process:* This process consists in periodically picking up some *vehicles* at *excess* stations, that means stations which may be considered as containing more than enough *vehicles*, and move them to *deficit* stations, while using *carriers* (trucks, self-platoon convoys…). Optimizing this process gives rise to *Vehicle Sharing Rebalancing* (VSR) models. While *on line* VSR models received very little attention, being only handled through application of empirical decision rules (see [5, 7]), several *static* VSR models (see [2, 3, 7, 8]) have already been proposed and studied through heuristics and ILP models.

Our purpose is here to cast operational VSR as a *slave* sub-problem of a strategic *Vehicle Sharing Station Location* (VSSL) model. We first consider that the input data for VSSl problem mainly consists in an origin/destination matrix OD, and in additional information about demands and costs, and derive (Section II) a bi-level *Vehicle Sharing Station Location* (VSSL) whose master problem IS a *Facility Location* model and slave sub-model is some static VSR model. Next we propose (Section III) a related bi-level algorithmic resolution scheme which decomposes in turn the VSR model into a simple *Min-Cost Assignment master* model and a *slave* PDP: *Pick up and Delivery* model (see [1, 5]) model. We end by providing a VSR lower bound and performing numerical experiments (Section IV).

II. The VSSL Model

A. Vehicle Sharing Station Location Instances

VSSL (*Vehicle Sharing Station Location*) input is a set *VS* of *virtual stations*, given together with:

- a *demand* matrix *OD*: for any *x, y* in *VS*, *OD(x, y)* means the *access demand* to the system in *x, y*, that means the number of *vehicles* which should be

picked up at station x and given back at station y by the users during a *reference* period P.
- a distance matrix $DIST$: $DIST(x,y)$ means the distance (time required) from x to y.

Demands and Costs: Solving VSSL means computing a real station subset X of VS and its related capacity function C Given a subset X of VS and a station u in VS, we denote by $Prox(u, X)$ the element x in X which is the closest to u Then the *Access Demand* $Acc(x, y, X)$ which is induced by X between two stations x and y of X is given by:
- $Acc(x, y, X) = \Sigma_{u, v \text{ in } VS \text{ such that } x = Prox(u, X), y = Prox(v, X)} OD(u,v).\Phi(Dist(u, Prox(u, X)).\Phi(Dist(v, Prox(v, X))$, Φ being a decreasing [0, 1]-valued function.

We set: $Global\text{-}Demand(X) = \Sigma_{x,y \text{ in } X} Acc(x, y, X)$.
This *Access Demand* induces, for any station x in X, a residual quantity $Res(x, X)$:
- $Res(x, X) = \Sigma_y Acc(y, x, X) - \Sigma_y Acc(x, y, X)$.

This *residual* quantity means the number of vehicles which is likely to be in excess ($Res(x, X) > 0$) or in deficit at station x at the end of standard period P.
The *Top Demand* in station $x \in X$, i.e. the variation between the least and the largest numbers of *vehicles* in station x during period P, is given by:

$Top(x, X) = Q(x, X).H(x, X)$ with :

$Q(x, X) = Sup(\Sigma_y Acc(y, x, X)), \Sigma_y Acc(x, y, X))$

$H(x, X) = \Pi(|Res(x, X)|/Q(x, X))$,

Π being a decreasing $[\alpha, 1]$-valued function, $\alpha > 0$.

Setting a station at node x in VS with capacity $C = C(x)$, has a fixed cost $Fix(x)$, augmented with a flexible cost $C.Prop(x)$, which linearly depends on C.
Besides, since running the system defined by X and function C periodically requires relocating *vehicles* from *excess* stations to *deficit* stations, we denote by $Run\text{-}Cost(X, C)$ the cost of this *rebalancing process*.

Constraints: $X \subseteq VS$ and C are subject to:
- *Capacity Constraints*: for any $x \in X$, $Top(x, X) \leq C(x)$;
- *Demand Constraints*: *Global Demand* should be at least equal to some target level *Goal*: $Global\text{-}Demand(X) \geq Goal$.

Then the VSSL model comes as follows:

VSSL Model : {Compute the subset X, the *Depot* station D, and the capacity function C in such a way that the *Capacity and Demand Constraints* be satisfied and that:
- $Cost = \Sigma_{x \in X} (Fix(x) + C(x).Prop(x)) + Depot\text{-}Cost(D) + Run\text{-}Cost(X, C, D)$ is minimal.

We denote by *Relax*-VSSL the restriction of VSSL which is obtained by removing the *Run-Cost* quantity.

B. The Vehicle Sharing Rebalancing Problem: VSR

Let us suppose that X and C are given, together with the *Depot* station D. For any station x, $v(x) = Res(x, X)$ *vehicles* are in *excess* at station x: if $v(x) < 0$, we talk about *deficit*. We suppose $\Sigma_{x \in X} v(x) = 0$, which means that D may bring additional *vehicles* to the system. *K-Max* is the number of available *carriers*, all with capacity *CAP* and initially located at D. This defines the *VSR* instance (X, v, C, D, *K-Max*).

VSR Feasible Solutions: A VSR tour γ is a finite sequence $\gamma_{Route} = \{x_0 = D, x_1, .., x_{n(\gamma)} = D\}$, *of stations,* given together with a *loading strategy,* that means with 2 sequences $\gamma_{Load} = \{L_0, L_1, .., L_{n(\gamma)}\}$ and $\gamma_{Time} = \{T_0 = 0, T_1, .., T_{n(\gamma)}\}$ of coefficients whose meaning is: a *carrier* which follows the route γ_{Route} loads, at time T_i , L_i *vehicles* at *station* x_i (unloads in case $L_i < 0$). This VSR tour γ is *feasible* if:
- For any $i = 0, .., n(\gamma)\text{-}1$,
$$T_{i+1} \geq T_i + DIST(x_i, x_{i+1}); \quad \text{(E1)}$$
- For any $i = 0, .., n(\gamma)\text{-}1$,
$$L^*_i = \Sigma_{j = 0..i} L_j \leq CAP; \quad \text{(E2)}$$
- $\Sigma_{j = 0..n(\gamma)} L_j = 0$; (E3)
- For any j such that $v(x_j) \geq 0$, $v(x_j) \geq L_j \geq 0$; (E4)
- For any j such that $v(x_j) \leq 0$, $v(x_j) \leq L_j \leq 0$. (E5)

Then *a feasible solution* for the *VSR* instance (X, v, C, D, *K-Max*, *DIST*) is a collection $\Gamma = (\Gamma(k), k = 1..K \leq K\text{-}Max)$ of *feasible tours*, such that, for any station x: $\Sigma_k \Sigma_{i \text{ such that } x(k)i = x} L(k)_i = v(x)$. (E6)
The cost of Γ is given by: $R\text{-}Cost(\Gamma) = \alpha.K +$
$\beta.\ Sup_k\ T(k)_{n(\Gamma(k))} + \chi.\ \Sigma_k\ T(k)_{n(\Gamma(k))}$
$+ \delta.(\ \Sigma_k \Sigma_j\ DIST(x(k)_j, x(k)_{j+1}).L^*_j)$,
where α, β, χ, δ are some scaling coefficients.

We derive the following **VSR Model**: {Compute a *feasible* VSR solution $\Gamma = (\Gamma(k), k = 1..K)$ which minimizes the above quantity $R\text{-}Cost(\Gamma)$}.

Remark 1: The $Run\text{-}Cost(X, C, D)$ quantity of the VSSL model is the optimal value of this *VSR* model.

III. ALGORITHMS

We deal with the VSSL model according to a GRASP hierarchical decomposition scheme:

VSSL-GRASP Scheme
Initialize X and C while solving *Relax*-VSSL;
Not *Stop*; While Not *Stop* do
Solve the slave VSR model induced by X; (*)
Derive an additional constraint $C\text{-}Aux(X)$, and update X, C through local search;

We implement the first instruction by adapting *Facility Location* algorithms (see [8]) into a *Relax*-VSSL procedure, while observing that the capacity function C derives from X and the *Top Demand* function x, $X \rightarrow Top(x, X)$ in a straighforward way. The resulting procedure is a GRASP Algorithm, which involves local search operators *Insert*(x) *Remove*(x), *Replace*(x, y) and *Merge*(x, y).

X being given, let us now explain how we deal with the resulting VSR sub-problem ((*) instruction).

A. Decomposing VSR into Min Cost Assignment and PDP: The Distance Strategy.

In case we could decide, for any pair (x, y), x *excess*, y *deficit* station, which quantity $Q_{x,y}$ has to move from x to y in order to achieve the *rebalancing process*, then we derive a VSR solution by solving the *Load Splitable* PDP instance (see [1]) defined by:

- *Requests* correspond to the 3-uple ($o(j) = x$, $d(j) = y$, $Load(j) = Q_{x,y} \neq 0$)
- Minimize $\alpha.K + \beta.$ Sup $_k$ $Length(\Gamma(k)) + \gamma.\ \Sigma_k$ $Length(\Gamma(k)) + \delta.\ \Sigma_j$ $Ride(j)$: k denotes the vehicles, ($\Gamma(k)$) the related PDP tours and $Ride(j)$ is the time spent by $Load(j)$ inside a *truc*k.

We check that:

Theorem 1: *We may restrict ourselves to vectors $Q = (Q_{x,y}$, x excess station, y deficit station) which are vertices of the Assignment polyhedron P-Assign:*

P-Assign: {$Z = (Z_{x,y}$, x excess, y deficit) such that:

- *For any excess station x, $\Sigma_{y\ deficit} Z_{x,y} = v(x)$;*
- *For any excess station x, $\Sigma_{x\ excess} Z_{x,y} = - v(x)$}*

This leads us to handle VSR through the following decomposition scheme:

***VSR Assignment/PDP Decomposition Scheme*:**

Initialize cost vector Q; Not *Stop*;

While Not *Stop* do

Derive Z and the *Request* set $J = J(Z)$)

Solve the *Load Splitable* PDP related instance;

Update Q;

The Distance Strategy: Initializing Q comes in a natural way by setting: for any x, y, x *excess*, y *deficit* stations, $Q_{x,y} = DIST(x, y)$. We call this strategy, the *Distance Strategy*. We may state:

Theorem 2: *If K is fixed and β, δ equal to 0 (we minimize the carrier riding time), then the Distance strategy induces a VSR approximation ratio of (1+CAP). This is the best possible ratio*.

Theorem 3: *If K is fixed and χ, δ equal to 0 (we minimize the makespan), then the Distance Strategy induces a VSR approximation ratio of (1+K.CAP). This is the best possible ratio*.

B. VSR-Assignment/PDP Algorithm

We follow the guideline of the previously described hierarchical decomposition scheme. As a matter of fact, we revisit it as follows

***VSR-Assignment/PDP Algorithm*:**

Initialize cost vector Q; Derive Z and the *Request* set $J = J(Z)$; Solve the *Load Splitable* PDP related instance through some generic *Insertion* algorithm and get a current VSR solution Γ; Not *Stop*;

While Not *Stop* do

Update cost vector Q and the *Request* set $J = J(Z)$; Let J_0 the set of formerly existing requests which have been removed from J and J_1 the set of newly created requests; Remove J_0 and next Reinsert J_1, in the sense of the *PDP Insertion* algorithm, into current solution Γ;

Cost vector Q and related *Request* set J are updated by:

- 1 th Step: Identify a subset $J_0 \subseteq J$ of *poorly inserted requests* (those with a large gap between cost $Q_{x,y}$ and mean *riding time* $R_{x,y}$);
- 2 th step: Set, for any x, y involved into J_0, x *excess*, y *deficit*, $Q_{x,y} = (Q_{x,y} + R_{x,y})/2$.

C. Retrieving Sensitivity Constraint C-Aux(X)

A key instruction inside the main loop of the *VSSL-GRASP* algorithm is the following:

"*Derive an additional constraint C-Aux(X)…*"

We implement it while using the dual solution λ_x, $x \in VS$ of the *Min-Cost Assignment* problem related to current vector Q, as a sub-gradient vector and derive the following Bender's like constraint *C-Aux*(X):

$\Sigma_{x \in VS}$ Res(x, X). $\lambda_x \leq \Sigma_{x \in VS}$ Res(x, X_0). λ_x .

D. A Lower Bound for the VSR model

We get a VSR lower bound *LB* by introducing (see [8]) a network with time indexed nodes and turning *Preemptive* VSR (*carriers* may exchange *vehicles* while performing the *Rebalancing* process) into a network flow model, which involves an integral *carrier* flow vector dominating some rational *vehicle* flow vector. Practically, we compute *LB* while using an ILP solver and applying some rounding process when the size of G is too large.

IV. Numerical Experiments

Since we can't provide exact reference values for the VSSL model, we separately evaluate the distinct components of the VSSL-GRASP Algorithm.

A. Testing VSR-Assignment/PDP and Relax-VSSL

A VSR instance is identified by the numbers *n*, *nd*, *K-Max*, by the matrix *DIST*, and by function *v*. *T-Max* is set to 480. We compute, for any instance:

- the value *LB* of the lower bound of Section III;
- the value *V-NP-Dist* (*V-NP*) of the solution related to the *Distance* strategy (*VSR-Assignment/PDP*) and its related CPU time;

Assignment/PDP) and its related CPU time; We get (on PC AMD Opteron 2.1GHz, while using gcc 4.1 compiler and the CPLEX12 library):

Table I:
Testing *VSR-Assignment/PDP*

n , n_A, K-Max	*V-NP-Dist/LB(%)*	*V-NP-Dist-CPU(s)*	*VNP/LB (%)*	*NP-CPU*
48, 20, 3	19.6	4.6	14.6	17.6
48, 30, 3	24.0	6.7	16.2	20.5
48, 40, 3	12.4	8.9	10.4	29.0
96, 20, 5	13.8	7.1	10.8	30.5
96, 60, 5	22.6	10.3	14.8	48.8
148, 40, 7	11.8	10.1	8.5	20.2
148, 80, 7	15.7	18.3	14.7	48.4

Comment: The LB value provides us with a rather good approximation. Though the *Distance* strategy is rather efficient, we improve V-NP values in a significant way by fully performing local search.

In order to test *Relax-VSSL*, we generate a set *VS* of *n* points of the Euclidian space R^2, (so *DIST* means the Euclidian distance), and an origin/destination matrix *OD*, with all values *OD(x, y)* between 0 and a given parameter *S*, and uniformly distributed. Functions Φ and Π are piecewise linear. We compute, for every instance, the gap *G* between the CPLEX optimal solution and the of *Relax-VSSL* together with related CPU times *T-ILP* and *T-Rel*. Then we get, while always setting *S* to 10:

Table II:
Testing *Relax-VSSL*

n , M	*G (%)*	*T-ILP (s)*	*T-Rel*
10, 5	1.1	29.6	4.1
15, 5	0	126	6.9
15, 10	0	*	14.0
20, 5	2.8	1588	10.7
20, 10	0	*	21.2

B. Testing VSSL-GRASP

A VSSL test is identified here by:

- the coefficients *n* (cardinality of *VS*, *S* (top *OD(x, y)* value), *q* (relative weight of *Run-Cost* inside the global cost of a solution);
- the number *M* of replications of the *VSSL-GRASP* scheme;
- the length L of the main loop of *VSSL-GRASP.*

We compute, for any instance, the gap *G* between the initial cost obtained through *Relax-VSSL* and the final cost obtained through *VSSL-GRASP*, together with related CPU times *T0* and *T1*. Then we get:

Table III:
Testing *VSSL-GRASP*

n , S, q (%) M, L	*G (%)*	*T0(s)*	*T1*
20, 10, 25%, 20, 50	8.5	42.5	288.4
20, 20, 50%, 20, 50	17.6	43.1	360.8
20, 30, 75%, 5, 10	30.6	11.2	152.6
20, 30, 75%, 20, 50	35.9	43.6	506.2
50, 10, 25%, 5, 20	5.0	29.6	304.0
50, 10, 50%, 5, 20	12.6	29.6	334.8
50, 10, 75%, 5, 20	28.4	29.6	350.6
50, 10, 75%, 10, 50	31.3	58.8	1482.0

Comment: Computing costs increase with the *S* value.

References

[1] C. Archetti, M. Speranza: Vehicle routing problems with split deliveries; p 3-22, ITOR, (2012). http://dx.doi.org/10.1111/j.1475-3995.2011.00811.x

[2] M. Benchimol, P. Benchimol, B. Chappert, A. De la Taille, F. Laroche, F. Meunier, L. Robinet: Balancing the stations of a self service bike hiring systems, *RAIRO-RO* 45, p 37-61, (2011). http://dx.doi.org/10.1051/ro/2011102

[3] D. Chemla, F. Meunier, R. Wolfler Calvo: Bike sharing systems: solving the static rebalancing problem; *Discrete Optimization* 10 (2), p. 120-146, (2013). http://dx.doi.org/10.1016/j.disopt.2012.11.005

[4] D. Gavalas, C. Konstantopoulos, G. Pantziou: Design and management of vehicle sharing systems: a survey of algorithmic approaches; *ArXiv e-prints*, October 2015. https://arxiv.org/abs/1510.01158v1

[5] G. H. Kek, R. L. Cheu, Q. Meng, C. Ha Fung: A study on the vehicle size and transfer policy for car rental problems *Transportation Res. E: Logistics and Transp. Review* 64 (1), p 110-121, (2014), http://dx.doi.org/10.1016/j.tre.2014.01.007

[6] M. Nourinedjad, M. J. Roorda: A dynamic carsharing decision support system; *Transp. Res. E*, 66, p 36-50, 2014). http://dx.doi.org/10.1016/j.tre.2014.03.003

[7] B. Bernay, S. Deleplanque, A. Quilliot: Routing in Dynamic Networks: Grasp Versus Genetics, *7 th WCO Workshop*, FEDCIS Conf, Warwsaw, p 487, 492, (2014) http://dx.doi.org/10.15439/978-83-60810-58-3

[8] A. Klose, A. Drexl: "Facility location models for distribution systems"; *EJOR* 162, p 429-449, 2005. http://dx.doi.org/10.1016/j.ejor.2003.10.031.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 609–617
DOI: 10.15439/2016F124
ACSIS, Vol. 8. ISSN 2300-5963

Formulation and Practical Solution for the Optimization of Memory Accesses in Embedded Vision Systems

Khadija HADJ SALEM
LCIS Laboratory
50 rue Barthélémy de Laffemas BP 54
26902 Valence, France
khadija.hadj-salem@lcis.grenoble-inp.fr

Yann KIEFFER
LCIS Laboratory
50 rue Barthélémy de Laffemas BP 54
26902 Valence, France
yann.kieffer@lcis.grenoble-inp.fr

Stéphane MANCINI
TIMA Laboratory
46 avenue Félix Viallet
38031 Grenoble, France
stephane.mancini@imag.fr

***Abstract*—The design of embedded vision systems carries a difficult challenge regarding the access times of memories holding image data for some particular cases of image treatments. This paper studies the optimization challenge reflecting the efficient operation of adhoc memory systems proposed by electronic designers to alleviate this problem. New algorithms are proposed for producing solutions to this 3-objective problem, and numerical experiments are conducted on real-world data for validating their efficiency.**

I. Introduction

The design of embedded vision systems carries many challenges, one of which is the efficient access to the image memory. An architectural solution was proposed by Mancini and al. [1] in the form of a software tool that creates an ad-hoc memory hierarchy for non-linear image accesses. But operating this kind of systems is itself an optimization challenge, involving 3 objectives reflecting 3 main electronic design parameters. To the best of our knowledge, this problem has not been studied before in the optimization literature.

The remaining of this paper is organized as follows. After describing the Memory Management Optimization design software, we explain the related optimization problematic set by the efficient operation of the circuits produced by this tool, and give a formal multi-objective mathematical model for this problem, as well as several sub-problems of interest. We then review the state of the art. After giving lower bounds for the 3-objective problem, we analyze the complexity of some of the mono-objective sub-problems. The description of new approaches, including a simple heuristic and two algorithms, is then given. Numerical experiments follow, which are conducted on real-world data in order to validate their efficiency, and a conclusion and perspectives section closes the paper.

II. Embedded Vision Systems Context: Architectural Solution and Optimization Problematic

A. Memory Management Optimization Tool

Among modern-day electronic devices, embedded vision systems such as picture and video cameras represent a specific design challenge with respect to memory management. Image sizes are measured in 100's of kbs or even Mbs, while the access times must be short enough to allow the quick handling of the data. For example, a live video feed may have 30 frames per second, meaning that the handling of one image (frame) must take less than 1/30 s. But it is a well-known fact in electronic memory design that access times grow with the size of the memory to be accessed. Due to this fact, it is not possible to reach the performance needed with the simple use of memories. Something has to be added to improve the access times.

For digital image treatments (also called *kernels*) that have linear access patterns to the memory addresses, usual caches as used for CPUs will solve this problem. But for non-linear access patterns, the problem remains a big hurdle for the easy and efficient design of kernel circuit designs. An illustration of a non-linear kernel is given in Fig. 1.

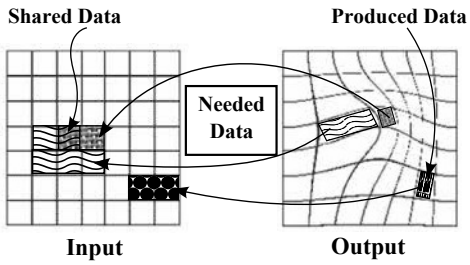


Fig. 1. Example of a non-linear kernel

To address this problematic, Mancini and al. [1] have designed a software generator of memory hierarchies tailored to one particular non-linear kernel. Their solution, called *Memory Management Optimization* (MMOpt), takes as input a non-linear kernel for which the memory hierarchy is to be produced, such as the one shown in Fig. 1; it analyzes its access patterns; it then designs a run-time behavior for the whole resulting block *Tile Processing Unit* (TPU); and it finally outputs the design of the TPU, together with the

information needed to orchestrate its operational behavior.

We give some details about the architecture of the TPU, as shown in Fig. 2. It is made of (a) a *Prefetching Unit* (PU) that loads data from external memory to local buffers, and (b) a *Processing Engine* (PE), that computes output data using prefetched input ones.

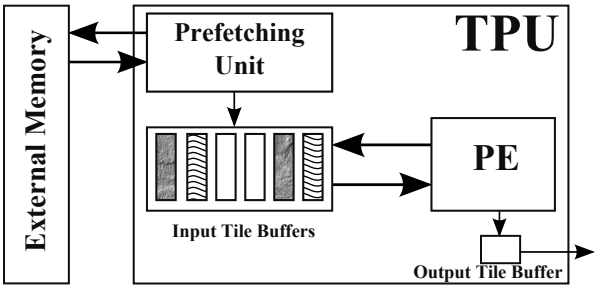


Fig. 2. Architecture template of the TPU

MMOpt computes and encodes into the TPU a schedule of prefetches and a schedule of computations. Hence memory accesses in the final system are deterministic (i.e. independent of pixel values), and this is a requirement of the input kernel for the whole MMOpt scheme to work out.

B. Optimization for MMOpt

When designing electronic circuits, some of the important design criteria are the area of the circuit produced, since it is directly related to production costs; the energy consumption, which may be limited, and which conditions the battery life for battery-powered devices; and the performance, which is usually a design parameter reflecting reactivity, and fluidity in the case of moving images.

TPUs produced by MMOpt embed schedules for the prefetches of input tiles and computations of output tiles (see Fig. 3).

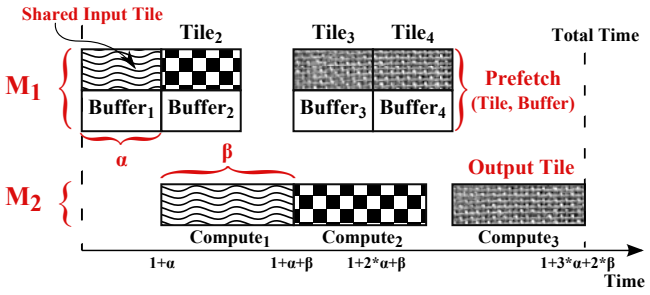


Fig. 3. Prefetches and computations schedules

The architecture of the TPU and those schedules will impact those three design parameters in the following way: the number of buffers of the TPU will account for most of its area; the number of prefetches reflects the main part of the energy consumption; and the performance is related to the completion time of the whole prefetches-computation schedule for the computation of one image.

Since MMOpt is a fully automatic electronic design software, computing good schedules is both a necessity and an opportunity for the circuit designers to deliver, with the help of MMOpt, low-cost, low-energy and efficient TPUs.

III. The 3-objective Process Scheduling and Data Prefetching Problem

As you may know, the Integer Linear Programming formulation is used in the operations research as a modeling approach for the optimization problems. In this study, we have chosen to formulate our optimization issue by an off-line, 3-objective model with clearly delineated inputs and outputs, which we now present. This multi-objective mathematical formulation is a very flexible modeling approach that allows a preciseness for dealing with many specific sub-problems.

A. Problem Statement and Assumptions

The main multi-objective optimization problem considered in this paper is called *3-objective Process Scheduling and Data Prefetching Problem* (3-PSDPP). It involves the definition of the number of buffers of the TPU, the scheduling of output tiles computations, and the scheduling of input tiles prefetches, while respecting a requirement constraint between prefetches and computations.

The main assumptions that apply to the 3-PSDPP are the following:

1) Input tile sizes are identical and each input tile fits exactly into one buffer.
2) There is no distinction between buffers, i.e. any input tile may be prefetched into any buffer.
3) All input (respectively output) tiles and the subset of input tiles required to compute each output tile are known in advance.
4) Only one input (output) tile can be prefetched (computed) at a time.
5) The prefetch operations and the computation steps may be carried out simultaneously.
6) Input (output) tile prefetch (respectively computation) times are constant and identical.

B. Formulation for 3-PSDPP

The 3-PSDPP problem consists of finding an appropriate computation sequence in which output tile will be computed, and an associated prefetch sequence in which input tile will be prefetched from the external memory to the buffers, that simultaneously minimizes the number of prefetches, the number of buffers, and the completion time. We now describe the input data of our problem 3-PSDPP, the expected output, the constraints, and the 3 formal objectives reflecting the 3 electronic design parameters (Mathematical Formulation given in Table I).

1) Inputs: a 3-PSDPP instance is represented by a 5-tuple $(\mathcal{X}, \mathcal{Y}, (\mathcal{R}_y)_{y\in\mathcal{Y}}, \alpha, \beta)$ where $\mathcal{X}$ is the set of input tiles to be prefeteched, and $\mathcal{Y}$ is the set of output tiles to be computed. Each output tile y requires its own set of input tiles, denoted

TABLE I
MATHEMATICAL FORMULATION FOR 3-PSDPP

Inputs	$\mathcal{X}= \{1,\ldots,X\}, \mathcal{Y}= \{1,\ldots,Y\}$, where $X, Y \in \mathbb{N}^*$ $\mathcal{R}_y \subseteq \mathcal{X}, \forall\ y \in \mathcal{Y}$ $\alpha, \beta\ \in \mathbb{N}^*$
Outputs	(N, Z, Δ), where $N, Z, \Delta \in \mathbb{N}^*$ $(p_i)_{i\in\mathcal{N}}$, where $p_i = (d_i, b_i, t_i)$ $(c_j)_{j\in\mathcal{M}}$, where $c_j = (s_j, u_j)$
Constraints	**(1)** $\forall\ y\ \in\mathcal{Y}, \exists\ j \in\mathcal{M}\ /\ s_j = y$ **(2)** $\forall\ j \in\mathcal{M}, \forall\ x \in\mathcal{X}, x \in \mathcal{R}_{s_j} \Rightarrow$ $(\exists\ a \in \{1,\ldots,u_j - \alpha\}, \exists\ i \in \mathcal{N}/t_i = a, d_i = x\ \&$ $(\forall a' \in \{a+\alpha,\ldots,u_j+\beta-1\}, \forall i' \in \{i+1,\ldots,N\},$ $t_{i'} = a' \Rightarrow b_{i'} \neq b_i))$ **(3)** $\forall\ i \in\mathcal{N}\backslash\{1\}, t_i\ \geq\ t_{i-1} + \alpha$ **(4)** $\forall\ j \in\mathcal{M}\backslash\{1\}, u_j\ \geq\ u_{j-1} + \beta$
Objectives	$\min\ Z, \min\ N, \min\ \Delta$

by $\mathcal{R}_y$. Also, the duration of a prefetch step α, and that of a computation step β, have to be given as input.

Remark 1. *The set of input tiles $(\mathcal{R}_y)_{y\in\mathcal{Y}}$ must be present in the buffers during the whole computation step of the tile y.*

Remark 2. *Each input tile x already prefetched earlier may be reused if it is still present in the buffer.*

2) Outputs: a feasible solution to such an instance is defined by $((p_i)_{i\in\mathcal{N}}, (c_j)_{j\in\mathcal{M}}, Z, N, \Delta)$.

- Configuration of the prefetched input tiles:
  we denote by $(p_i)_{i\in\mathcal{N}}$ the prefetch sequence, where $p_i = (d_i, b_i, t_i)$ encodes that input tile d_i is prefetched in the buffer b_i at the time t_i.
- Configuration of the computed output tiles:
  we denote by $(c_j)_{j\in\mathcal{M}}$ the computation sequence, where $c_j = (s_j, u_j)$ encodes that output tile s_j is to be computed at the time u_j.
- The values for the three criteria (Z, N, Δ):
  we denote by Z the number of buffers; N is the total number of prefetched input tiles; and Δ is the completion time, meaning the total time it takes for the whole operation of the TPU from the beginning of the first prefetch to the end of the last computation of one full image.

3) Constraints: the first constraint (1) on solutions is that for each output tile y, there exists a computation step j in which this output tile is computed. The second and main constraint (2) ensures that all the input tiles $\mathcal{R}_y$ required by y have to be prefetched from the external memory to the internal storage area (buffers) before the start date u_j of its associated computation step, and will not be overwritten until its end date. Input tiles already prefetched earlier can be reused, provided they have not been overwritten. Constraints (3) and (4) guarantee that different input (output) tiles cannot be pre-fetched (computed) simultaneously.

4) Objectives: in the formulation above, three objectives have to be minimized: the number of prefetches N reflects the main part of the energy consumption; the number of buffers Z of the TPU will account for most of its area, and is related to cost; and the completion time Δ accounts for the performance of the TPU.

C. Sub-Problems of 3-PSDPP

From this multi-objective problem 3-PSDPP, we derive several mono and bi-objective sub-problems. The mono-objective sub-problems we consider are *Minimum Buffers of 3-PSDPP* (MB-PSDPP), in which the number of buffers is to be minimized; *Minimum Prefetches of 3-PSDPP* (MP-PSDPP) in which the number of prefetches is to be minimized; and *Minimum Completion Time of 3-PSDPP* (MCT-PSDPP), in which the completion time is to be minimized.

For the remaining sub-problems to be presented, the number of buffers Z will be fixed as input data. Hence, we consider the *Prefetching and Scheduling Problem* (PSP), where the number of buffers is fixed as input, and the number of prefetches N is to be minimized. In addition, the variant of PSP, where the computation sequence is given as part of the input, is called the *Data Prefetching Problem* (DPP).

We will also consider the bi-objective sub-problem of 3-PSDPP, 2-PSDPP, where the number of buffers Z is fixed, and both N and Δ are to be minimized.

D. Related Work

In the study by Mancini and al. [2], two algorithms for optimizing the running of the TPU produced by the MMOpt tool are proposed.

The first algorithm is M_1, for which both the number of prefetches N and the completion time Δ are minimized. The second one, called M_2, aims at minimizing both the number of prefetches N and the number of buffers Z.

The algorithm M_1 proceeds in two steps which are, respectively, *Computations* and *Prefetches*. Furthermore, M_2 comprises three steps, the first two of which are those of M_1. The third step is *Delay Computations*.

For both algorithms M_1 and M_2, the authors [2] add a *Destinations* step for determining the sequence of destination buffers.

These different steps are outlined in Fig. 4.

1) Computations: this step encodes the order in which a batch of output tiles has to be successively computed, one at a time. The traffic to the external memory is then minimized by optimizing the obtained scheduling.

To construct the computation sequence, the authors [2] solve an instance of the *Asymmetric Traveling Salesman Problem* (ATSP) to find a Hamiltonian path in the complete directed graph $\overrightarrow{G}$ whose vertices are the set of output tiles, and whose arcs are weighted by $\varphi(k, l)$, where (k, l) is a pair of output tiles, in which the output tile k will be computed before the output tile l. The function $\varphi(k, l)$ defines the number of additional input tiles to be prefetched for computing the output tile l, when all input tiles shared between k and l are already prefetched.

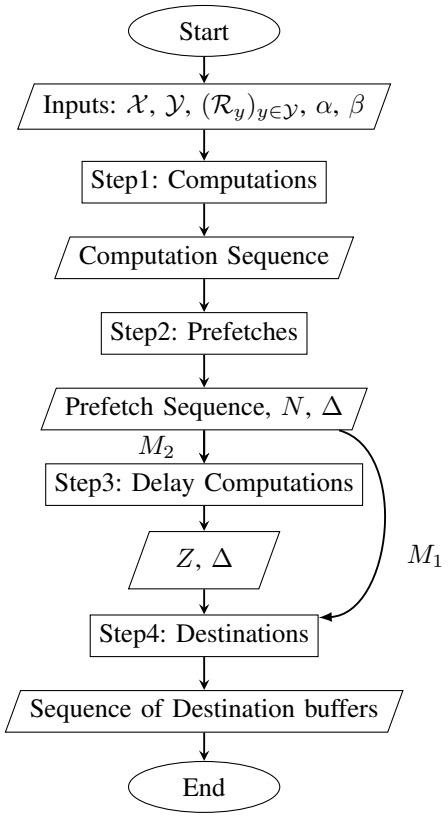


Fig. 4. Flowchart for algorithms M_1 and M_2

2) Prefetches: in this step, the authors [2] determine the schedule of prefetches associated to the computation sequence given by step 1. This schedule encodes which input tile should be prefetched from the external memory to the buffers at each moment. In fact, in parallel to each computation step, they prefetch the additional input tiles needed for the next computation.

3) Delay Computations: in this step, in order to reduce buffer usage, the authors [2] simply delay some computations by inserting fake computations when necessary. New trade-offs between the embedded memory area and the computing time can be then reached.

4) Destinations: this step simply consists in deciding in which buffer to place each prefetched input tile.

To the best of our knowledge, the 3-PSDPP problem has not been studied before in the operations research literature. We now relate some sub-problems of the 3-PSDPP problem to similar problems in the literature. Thus, we focus here on the uniform *Tool Switching Problem* (ToSP) arising in the flexible manufacturing context. The ToSP consists of finding an appropriate job sequence in which jobs will be executed, and an associated sequence of tool switches that minimizes the number of tool loading/unloading operations in the tool magazine of a single tool-switching machine. The general ToSP was first considered by Tang and Denardo [3]. They showed that the ToSP can be solved in polynomial time for a fixed job sequence, *Tooling Problem* (TP), using the *Keep Tools Needed Soonest* (KTNS) algorithm. On the other hand, when the job sequence is to be determined, Crama and al. showed that the ToSP is already NP-Hard for any fixed tool magazine capacity larger than or equal to 2 [4]. Different optimization techniques, including exact and heuristics methods, have been applied to its resolution (see Bard [5]; Privault and Finke [6]; Laporte and al. [7]; Konak and al. [8]; Amaya and al. [9]; Catanzaro and al. [10]).

IV. Models Analysis

For validating the efficiency of the proposed approaches, we develop three lower bounds lb_N, lb_Z, and lb_Δ for the different optimization criteria (N, Z, Δ). We then give complexity results for some of the mono-objective sub-problems of 3-PSDPP problem described in Section III-C.

A. Lower Bounds

Proposition 1. *$X - |\Omega|$ is a lower bound on the number of prefetches for the 3-PSDPP, where Ω denotes the set of input tiles which are not required by any output tile.*

Proof: For any solution to some given instance of 3-PSDPP, all input tiles that are required at least once for the computation of an output tile have to be prefetched at least once to some buffer. Hence the total number of prefetches cannot be less than $X - |\Omega|$. ■

Proposition 2. *$\max_{y\in\mathcal{Y}} |\mathcal{R}_y|$ is a lower bound on the number of buffers for the 3-PSDPP.*

Proof: Fix an instance of 3-PSDPP, and a feasible solution for that instance. When an output tile is computed, all the required input tiles have to be present in the buffers. Hence $\max_{y\in\mathcal{Y}} |\mathcal{R}_y|$ is a lower bound for the number of buffers in the solution. ■

Proposition 3. *$lb_1 = \alpha * X' + \beta$ and $lb_2 = \alpha + \beta * Y$ are lower bounds on the completion time Δ for the 3-PSDPP.*

Proof: Fix an instance of 3-PSDPP, and a feasible solution for that instance. Since all input tiles have to be loaded before the last computation starts, the completion time is at least $\alpha * X'$ (for the prefetches) plus β (for the computation of the last output tile).

Likewise, all output tiles have to be computed, and no computation can start before a first input tile has been prefetched. Hence the completion time is lower bounded by $\beta * Y$ (computation time for all output tiles) plus α (prefetch time for the first prefetch). ■

Thus, the completion time Δ is lower bounded by the maximum lb_1 and lb_2 ($lb_\Delta = \max\{lb_1, lb_2\}$).

B. Complexity Analysis

To prove that MB-PSDPP can be solved in polynomial time, we give an algorithm for which the number of buffers Z equals its lower bound ($Z_{\min} = lb_Z$). In this algorithm, we first fix the number of buffers Z to $\max_{y\in\mathcal{Y}} |\mathcal{R}_y|$. Then, for each output tile, we prefetch all its required input tiles into the Z buffers before

the corresponding computation step starts. The idea here is that a prefetch step and a computation step are not carried out in parallel.

To prove also that MP-PSDPP is solvable in polynomial time, we give another algorithm for which the number of prefetches N equals its lower bound ($N_{\min} = lb_N$). In this method, we first prefetch successively all the input tiles in $\mathcal{X}'$, where $\mathcal{X}' = \mathcal{X} \backslash \Omega$ and Ω denotes the set of the input tiles which are not required by any output tile. Then, when the prefetch steps are finished, all output tiles are successively computed.

For the MCT-PSDPP sub-problem, we have not yet been able to determine its complexity.

We now examine the two sub-problems of 3-PSDPP where N is to be minimized, and Z is fixed as input, namely PSP and DPP. We then prove the equivalence between PSP and the tool switching problem ToSP. In the description of the PSP problem, both input and output tiles ($\mathcal{X}$, $\mathcal{Y}$) are regarded as ToSP data (tools, jobs). The incidence matrix Tools×Jobs can then be regarded as the requirements of input tiles needed to compute all the output tiles $(\mathcal{R}_y)_{y\in\mathcal{Y}}$. The fixed number of buffers Z is the analogue of the capacity of the tool magazine. In addition, finding a computation sequence for minimizing the total number of prefetches corresponds to finding a job sequence for minimizing the total number of tools loadings. Thus, PSP is NP-Complete. When the computation sequence is given as input data, the same polynomial reduction works to prove the equivalence of DPP and Tooling Problem (TP). Hence, DPP is polynomially solvable.

This equivalence allows us to adapt the KTNS algorithm, as described by Tang and Denardo for solving the TP [3], to give an optimal solution for DPP. We call this adaptation *KTNS Adapted to DPP* (KAD). On the other hand, in the case of PSP, we developed an algorithm, named *KTNS Adapted to PSP* (KAP), to solve it.

We will also consider the bi-objective problem 2-PSDPP where the number of buffers Z is fixed, and both N and Δ are to be minimized. For solving the 2-PSDPP sub-problem, we developed a new solution approach called *Shifted Prefetches for bi-PSDPP* (SPbP).

To introduce both KAP and SPbP algorithms, respectively, for PSP and 2-PSDPP sub-problems of 3-PSDPP, we now present the specific adaptation KAD that it is an intermediate step in both KAP and SPbP algorithms.

V. Solution Methods

A. Constructive Heuristic

We propose now a constructive heuristic called H_1 for solving the 3-PSDPP, in which the number of buffers Z equals its number of required input tiles $X - |\Omega|$ and both N and Δ are to be minimized. In this method, the number of prefetches is optimum and we try to get the best possible completion time.

This algorithm proceeds in three phases. For each input tile x, we calculate its number of occurrences $O_c(x)$ in $(\mathcal{R}_y)_{y\in\mathcal{Y}}$. Then, the input tiles are sequenced in their decreasing order of $O_c(x), \forall x \in \mathcal{X}$. Finally, for each computation is determined when it can be scheduled at the earliest. The corresponding date is the end of the loading of the latest prefeteched tile among its required input tiles. Computations are scheduled greedily in this order, while making sure to respect these "at earliest" dates.

B. KTNS Adapted to DPP: KAD

We present an adaptation of the KTNS algorithm for solving the mono-objective sub-problem DPP (described in Section III-C). The KAD adaptation will be the second step of both KAP and SPbP algorithms presented in the following subsections.

We first restate the KTNS policy established by Tang and Denardo [3]. In our case, we can state the KTNS policy in this way:

1) At any instant, no input tile is prefetched unless it is required by the next output tile.
2) If an input tile must be prefetched, the input tiles that are kept (not removed) are those needed the soonest.

The pseudo-code of the KAD algorithm can be summarized as follows:

Algorithm 1 KAD

Input: $\mathcal{X}$, $\mathcal{Y}$, $(\mathcal{R}_y)_{y\in\mathcal{Y}}$, $(s_j)_{j\in\mathcal{M}}$, Z, α, β
Output: $(p_i)_{i\in\mathcal{N}}$, N, Δ

1: $M \leftarrow$ Incidence_Matrix ($\mathcal{X}$, $\mathcal{Y}$, $(\mathcal{R}_y)_{y\in\mathcal{Y}}$)
2: $P \leftarrow$ Permute (M, $(s_j)_{j\in\mathcal{M}}$)
3: $P' \leftarrow$ Flip_Blocks (P, Z)
4: $N, (d_i)_{i\in\mathcal{N}} \leftarrow$ Prefetches (P')
5: $(b_i)_{i\in\mathcal{N}} \leftarrow$ Destinations ($(d_i)_{i\in\mathcal{N}}$, P, Z)
6: $(t_i)_{i\in\mathcal{N}} \leftarrow$ Prefetches_StartDate ($(d_i)_{i\in\mathcal{N}}$, α, β)
7: $(u_j)_{j\in\mathcal{M}} \leftarrow$ Computations_StartDate ($(s_j)_{j\in\mathcal{M}}$, $(t_i)_{i\in\mathcal{N}}$, α, β)
8: $\Delta \leftarrow$ Completion_Time ($(u_j)_{j\in\mathcal{M}}$, β)

We now give some explanations about the pseudo-code. Given the computation sequence s_j, a number of buffers Z, and an $X \times Y$ *input tile-output tile* matrix M, where M_{xy} is 1 if output tile y requires x ($x \in (\mathcal{R}_y)_{y\in\mathcal{Y}}$) and 0 otherwise, we first determine the new incidence matrix P by permuting the columns of M according to the $(s_j)_{j\in\mathcal{M}}$.

In the second step, we determine the set of 0-blocks of P, as shown in Fig. 5. A 0-block is defined as a maximal subset of consecutive zeroes in a row of P. Intuitively, a 0-block is a maximal time interval which input tile i is not needed, but it is needed before and after this interval. Assume now that the 0-blocks of P have been ordered in increasing order of the index of their last column in P, the routine Flip_Blocks (P, Z) flips to 1 as many 0-blocks of P as possible, as long as each column $j (j = 1, \ldots, Y)$ of P contains no more than Z ones ($Z = 2$ in this example), as shown in Fig. 6.

The resulting matrix is denoted by P'. Then, the routine Prefetches(P') determines the total number of prefetches N, by counting all the blocks of 1 in P'. A 1-block can be defined

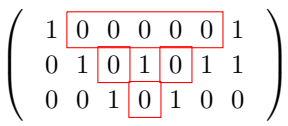


Fig. 5. KTNS-Initial Matrix

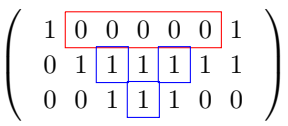


Fig. 6. KTNS-Final Matrix

as a maximal subset of consecutive ones in a row of P', for which an input tile i is not needed before and after this interval.

To construct the prefetch sequence $(d_i)_{i\in\mathcal{N}}$, we affect, for each output tile (column j in matrix P'), the set of associated input tiles (index of 1-blocks starting in column j) to be prefetched before the corresponding computation step. We determine then the associated sequence of destination buffers $(d_i)_{i\in\mathcal{N}}$, by affecting for each prefetched input tile a free buffer from the Z buffers. After that, we construct the associated schedules of prefetches-computations, in which the input tiles are prefetched one after the other. The same applies for computing the output tiles sequence. In the resulting prefetches-computations schedules, a prefetch step and a computation step are not carried out in parallel. Indeed, each computation step begins on the date where all its required input tiles have been prefetched. The routine Prefetches_StartDate() determines the start dates of the prefetches schedules. Similarly, the routine Computations_StartDate() determines the start dates of the computations schedules. The routine Completion_Time() computes then the completion time Δ.

C. Algorithm KAP for PSP

We have developed the KAP algorithm for solving the PSP sub-problem of 3-PSDPP, in which the number of buffers Z is fixed ($Z \geq \max_{y\in\mathcal{Y}}|\mathcal{R}_y|$) and the number of prefetches N is to be minimized. In contrast to DPP, the order in which the computations have to be carried out is not given as input, and has to be determined.

The KAP algorithm proceeds in two steps as follows:

1) Step1-Find a Computation Sequence (lines 1–3 of pseudo-code): in this step, we determine the computation sequence by solving the same instance of an ATSP as step 1 of both algorithms M_1 and M_2 (see Section III-D) [2].

2) Step2-KAD Algorithm (line 4 of pseudo-code): in second step, we resolve the DPP (described in Section V-B), by the KAD algorithm in order to determine the schedules of both the prefetches and computations with an optimal number of prefetches N.

The pseudo-code of the KAP algorithm can be summarized as follows:

D. Algorithm SPbP for 2-PSDPP

We have developed also the SPbP algorithm for solving the 2-PSDPP sub-problem of 3-PSDPP, in which the number of buffers Z is fixed ($Z \geq \max_{y\in\mathcal{Y}}|\mathcal{R}_y|$) and both the number of prefetches N and the completion time Δ are to be simultaneously minimized.

The SPbP algorithm proceeds in three steps as follows:

Algorithm 2 KAP

Input: $\mathcal{X}$, $\mathcal{Y}$, $(\mathcal{R}_y)_{y\in\mathcal{Y}}$, Z, α, β
Output: $(p_i)_{i\in\mathcal{N}}$, $(c_j)_{j\in\mathcal{M}}$, N, Δ

1: Let $\overrightarrow{G}=(Y,A)$ be a complete directed graph on Y
2: Let φ: $A \rightarrow \mathbb{N}$, $\overrightarrow{kl} \mapsto |\mathcal{R}_l \backslash \mathcal{R}_k|$
3: $(s_j)_{j\in\mathcal{M}} \leftarrow$ Short_Hamiltonian_Cycle ($\overrightarrow{G}$, φ)
4: N, $(d_i)_{i\in\mathcal{N}}$, $(b_i)_{i\in\mathcal{N}}$, $(t_i)_{i\in\mathcal{N}}$, $(u_j)_{j\in\mathcal{M}}$, $\Delta \leftarrow$ KAD ($\mathcal{X}$, $\mathcal{Y}$, $(\mathcal{R}_y)_{y\in\mathcal{Y}}$, $(s_j)_{j\in\mathcal{M}}$, Z, α, β)

1) Step1-Find a Computation Sequence (line 1 of pseudo-code): this step is the first one of KAP algorithm described in Section V-C.

2) Step2-KAD Algorithm (lines 2 of pseudo-code): this step is the second one of the KAP algorithm described in Section V-C

We give now a detailed description of the third step "Shifting Prefetches".

3) Step3-"Shifting Prefetches" (lines 3–9 of pseudo-code): In the resulting schedules of prefetches-computations, we apply the idea of "Shifting Prefetches", in which the completion time Δ is to be minimized. The originality of this step is to allow an overlap between the prefetch and computation steps. In this step, we shift in the prefetches schedule those that can be carried out in parallel with the previous computation step, by checking that no required input tiles of the output tile in this step were overwritten until its end date.

The pseudo-code of the SPbP algorithm can be summarized as follows:

Algorithm 3 SPbP

Input: $\mathcal{X}$, $\mathcal{Y}$, $(\mathcal{R}_y)_{y\in\mathcal{Y}}$, α, β, Z
Output: $(p_i)_{i\in\mathcal{N}}$, $(c_j)_{j\in\mathcal{M}}$, N, Δ

1: $(s_j)_{j\in\mathcal{M}} \leftarrow$ Computation_Sequence($\mathcal{X}$, $\mathcal{Y}$, $(\mathcal{R}_y)_{y\in\mathcal{Y}}$)
2: N, $(d_i)_{i\in\mathcal{N}}$, $(b_i)_{i\in\mathcal{N}}$, $(t_i)_{i\in\mathcal{N}}$, $(u_j)_{j\in\mathcal{M}}$, $\Delta \leftarrow$ KAD ($\mathcal{X}$, $\mathcal{Y}$, $(\mathcal{R}_y)_{y\in\mathcal{Y}}$, $(s_j)_{j\in\mathcal{M}}$, α, β, Z)
3: **for** $j = 2$ To $\mathcal{M}$ **do**
4: Prefetch[j] $\leftarrow \{i \in \mathcal{N}/P'[j][d_i] - P'[j-1][d_i] = 1\}$
5: Buffer[j] $\leftarrow \{b_i/i \in$ Prefetch[j]$\}$
6: Advanced_Prefetch[j] $\leftarrow$ $\{l/l \in$ Prefetch[j]&Buffer[j][l] $\notin b_i(\mathcal{R}_{s_{j-1}})\}$
7: Not_Advanced_Prefetch[j] $\leftarrow$ Prefetch[j]$\backslash$Advanced_Prefetch[j]
8: **end for**
9: $(t_i)_{i\in\mathcal{N}}$, $(u_j)_{j\in\mathcal{M}}$, $\Delta \leftarrow$ Schedule_Prefetches_Computations ($(s_j)_{j\in\mathcal{M}}$,Advanced_Prefetch[],Not_Advanced_Prefetch[], α, β)

E. Solutions for 3-PSDPP

Though our 3-PSDPP is a multi-objective optimization problem, we have developed two approaches. The first one

is the KAP algorithm for solving the mono-objective sub-problem PSP, where the number of buffers Z is fixed as input, and the number of prefetches N is to be minimized. The second one is the SPbP algorithm for solving the 2-PSDPP problem, in which the number of buffers Z is fixed, and both N and Δ are to be minimized.

The different steps of both both KAP and SPbP algorithms are outlined in Fig. 7. As shown in this figure, the KAP algorithm represents the two first steps of the SPbP algorithm.

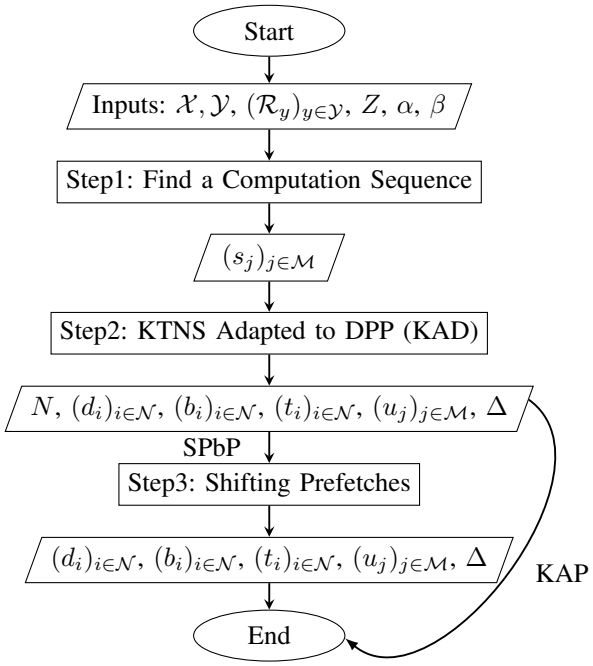


Fig. 7. Flowchart of KAP and SPbP algorithms

Therefore, we can use both KAP and SPbP as two methods for producing solutions to the main multi-objective optimization problem 3-PSDPP. By varying the number of buffers Z, both KAP and SPbP algorithms give a set of different solutions, in order to let the circuit designer pick his favorite compromise solution. It is an asset of these two methods that the circuit designer may choose his solution.

F. Example

In order to illustrate the process of both KAP and SPbP algorithms for solving two sub-problems of 3-PSDPP, where $z = 4$ buffers, let us consider the input data given in Fig. 8 for the case where $x = 5$ input tiles, $y = 3$ output tiles, $\alpha = 2$, and $\beta = 3$ time units.

We first determine the computation sequence $s_j = Y_2, Y_1, Y_3$, by finding a short Hamiltonian cycle in the graph $\overrightarrow{G}$ (see Fig. 9).

By running the *KAP* algorithm on our example, Fig. 10 gives the corresponding prefetches and computations schedules together with values of the outputs.

$$M = \begin{array}{c} \\ X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \end{array} \begin{array}{c} \begin{array}{ccc} Y_1 & Y_2 & Y_3 \end{array} \\ \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 0 & 0 \end{pmatrix} \end{array}$$

Fig. 8. $X \times Y$ Matrix

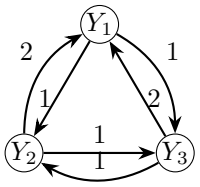


Fig. 9. Graph $\overrightarrow{G}$

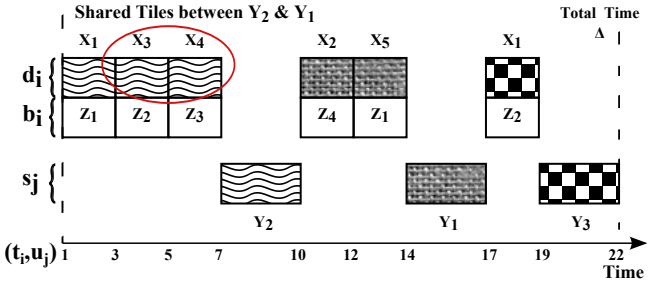


Fig. 10. KAP Schedules

Finally, we apply the **"Shifting Prefetches"** for minimizing the completion time Δ. The Fig. 11 shows that the completion time Δ is reduced by 2 units of time.

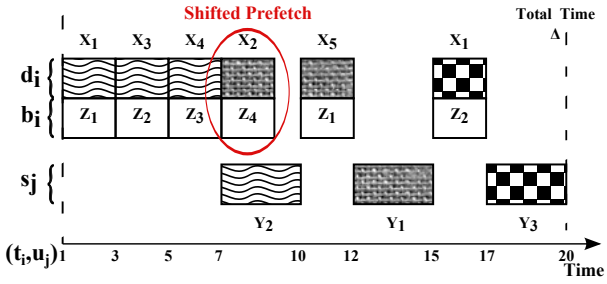


Fig. 11. SPbP Schedules

VI. Experiments and Results

In order to illustrate the practicability and efficiency of both KAP and SPbP algorithms, and the heuristic H_1, we use a set of 5 benchmarks from real-life non-linear image processing kernels already used by Mancini and al. [1].

A. Data Sets

Table II shows the characteristics of the test instances, together with the values for X (number of input tiles) and Y (number of output tiles to be computed).

As summarized in Table II, the benchmarks are variations of four kernels (fisheye, polar, fd resize, and fd haar) for which the input data structure (multi-resolution (an)isotropic mipmap input data) is modified. In fact, the first three kernels represent geometric non-linear transformations [11], [12]. The fourth kernel creates a pyramidal multi-resolution image [13]. The last one represents a kernel of a face detection application based on haar features [13]. The number of the input image

tiles varies between 350 and 7000 input tiles, and the number of the output tiles varies between 150 and 1200 tiles.

TABLE II
PARAMETERS VALUES OF DATA SETS

N°	**Kernel**	**Input data type**	X	Y
1	Fisheye	mipmap isotropic	352	158
2	Fisheye	mipmap anisotropic	704	158
3	Polar	mipmap anisotropic	4225	112
4	Fd Resize	mipmap isotropic	1280	1186
5	Fd Haar	pyramidal integral image	7040	428

Table III shows for each kernel the values of different lower bounds ($lb_Z, lb_N, lb_1, lb_2, lb_\Delta$), which are developed in Section IV-A. These lower bounds allow us to evaluate the performances of both proposed approaches KAP and SPbP.

TABLE III
LOWER BOUNDS FOR N, Z, AND Δ

Kernel N°	lb_Z	lb_N	lb_1	lb_2	lb_Δ
1	13	224	452	477	477
2	21	360	724	477	724
3	20	244	492	339	492
4	13	429	862	3561	3561
5	96	2272	4548	1287	4548

B. Numerical Results

This section presents an experimental analysis of the performance of both KAP and SPbP algorithms, respectively the heuristic H_1. The algorithms just described were coded in Python, except the ATSP part which was re-encoded as a TSP, and run through Concorde's Chained Lin-Kernighan implementation [14]. Tests were run on a computer powered by an Intel Core i5 processor with 4 GB of RAM. All our tests were carried out for the case where $\alpha = 2$ and $\beta = 3$ time units.

Table IV summarizes the numerical results for both KAP and SPbP algorithms, where the number of buffers is fixed to Z_1, respectively to Z_2, and those of the algorithms M_1, and M_2 on different data sets described in Table II. For the 5 kernels (given in line 1), the running time of both KAP and SPbP algorithms is in the order of a few minutes. The third line gives the number of prefetches N, the number of buffers Z, and the completion time Δ for the algorithms M_1, and M_2 (line 2). For both cases Z_1, and Z_2 (line 3), the N and Δ achieved by the KAP algorithm are then given in line 5. The sixth line (Gain 1) shows the gains as measured relatively to the lower bounds (lb_N and lb_Δ) of KAP algorithm, for both N and Δ relatively to those achieved by algorithms M_1, and M_2. Similarly, both N and Δ achieved by the SPbP algorithm are then given in line 7. The eighth line (Gain 2) shows the gains relative to the lower bounds (lb_N and lb_Δ) of SPbP algorithm, for both N and Δ relatively to those achieved by algorithms M_1, and M_2. The three last lines give for each of the algorithms M_1, M_2, KAP, and SPbP, the ratio of the achieved completion time Δ to the lower bound lb_Δ (given in Table III). The column Average provides the average gains (%) for all the kernels in the case of Z_1, respectively, of Z_2.

As illustrated in Table IV, the fixed Z_1 is larger than its lb_Z, for which the algorithm M_1 reaches the maximum number of buffers and the completion time is minimized. Whereas, Z_2 gives the minimum number of buffers (Z_2 equals its lower bound lb_Z).

By running the KAP algorithm, the traffic to the external memory is reduced with an average reduction of 57.5% (Z_1), respectively, of 36.8% (Z_2). In contrast, the completion time is increased, with an average increase of 14.9% (Z_1), respectively, of 22%(Z_2). This is due to the absence of overlap between the prefetches and computations in the schedules produced by the KAP algorithm.

In addition, due to the reuse of the KAP algorithm as a subroutine of the SPbP algorithm, the traffic to the external memory is reduced with the same average reduction of 57.5% (Z_1), respectively, of 36.8% (Z_2). In contrast to KAP, minimizing Δ by SPbP leads to a 37.1% (Z_1), respectively, to a 25% (Z_2) decrease in average of the completion time.

In the same way, a comparison between the completion time Δ achieved by each of the algorithms M_1, M_2, KAP, and SPbP and the lower bound lb_Δ, is considered. A comparison against the lower bound provides a measure of deviation from optimality. It is used as a performance indicator, and calculated by taking the ratio of Δ to lb_Δ. As shown in Table IV, for both cases Z_1, and Z_2, the completion time Δ of SPbP algorithm is in average more closer to the value of lb_Δ than the different values given by each of the algorithms M_1, M_2, and KAP. It is at most twice the value of lb_Δ. This is explained by inserting "Fake Computation" sometimes to stop processing or prefetching in order to optimize the use of resources. This means also that the SPbP algorithm gives a better completion time Δ than the other methods.

In summary, these numerical experiments show that the schedules produced by both KAP and SPbP algorithms, for a given number of buffers Z, have the optimal number of prefetches N. In contrast to KAP, SPbP gives the best completion time Δ. Hence, these results are numerical evidence validating the efficiency of our proposed approaches as compared to the one currently in use in the MMOpt tool.

In the same way, Table V shows the completion time Δ of the heuristic method H_1 for all the non-linear kernels given in Table II. Thus, a comparison between the completion time Δ obtained by method H_1 and the lower bound lb_Δ, is considered. It is calculated by taking the ratio of Δ to lb_Δ.

TABLE V
COMPARISON BETWEEN Δ OF H_1 AND lb_Δ

Kernel N°	**1**	**2**	**3**	**4**	**5**
Δ (H_1)	547	736	551	3658	4551
Δ (H_1) / lb_Δ	1.14	1.01	1.11	1.02	1.00

As shown in Table V, the completion time Δ of H_1 is

TABLE IV
NUMERICAL RESULTS OF M_1, M_2, KAP, AND SPBP

Kernel N°		**1**		**2**		**3**		**4**		**5**		**Average (Z1) %**	**Average (Z2) %**
		M_1	M_2	M_1	M_2	M_1	M_2	M_1	M_2	M_1	M_2		
MMOpt	Z	20	13	29	21	28	20	18	13	139	96		
	N	395	395	640	640	478	478	1710	1710	3640	3640		
	Δ	907	976	1341	1457	1021	1081	5075	5129	7899	8070		
		Z_1	Z_2	Z_1	Z_2	Z_1	Z_2	Z_1	Z_2	Z_1	Z_2		
KAP (Fixed Z)	N	298	322	457	517	353	405	1283	1458	2560	2997		
	Δ	1071	1119	1389	1509	1043	1147	6125	6475	6405	7279		
Gain 1	N	56.7	42.6	65.3	43.9	53.4	31.1	33.3	19.6	78.9	47.0	57.5	36.8
	Δ	-38.1	-28.6	-7.7	-7.0	-4.1	-11.2	-69.3	-85.8	44.5	22.4	-14.9	-22.0
SPbP (Fixed Z)	N	298	322	457	517	353	405	1283	1458	2560	2997		
	Δ	795	871	1113	1226	851	947	4629	4916	5849	6789		
Gain 2	N	56.7	42.6	65.3	43.9	53.4	31.1	33.3	19.6	78.9	47.0	57.5	36.8
	Δ	26.0	21.0	36.9	31.5	32.1	22.7	29.4	13.5	61.1	36.3	37.1	25.0
Δ (MMOpt) / lb_Δ		1.90	2.04	1.85	2.01	2.07	2.19	1.42	1.44	1.73	1.77	1.79	1.89
Δ (KAP) / lb_Δ		2.24	2.34	1.91	2.08	2.11	2.33	1.72	1.81	1.40	1.60	1.87	2.03
Δ (SPbP) / lb_Δ		1.66	1.82	1.53	1.69	1.72	1.92	1.29	1.38	1.28	1.49	1.49	1.66

very closer to the value of lb_Δ. This means that the lb_Δ is a good lower bound on the completion time Δ for the 3-PSDPP problem.

VII. CONCLUSION AND FUTURE WORK

In this paper, we addressed the multi-objective optimization problem 3-PSDPP, that arises in the context of optimizing the memory hierarchy of non-linear kernels in order to enhance some electronic stakes (energy consumption, real time, and cost) in the MMOpt tool. We described some of its mono- and bi-objective variants, proved some lower bounds, and analyzed the complexity of some of them. We then presented a constructive heuristic to solve the 3-PSDPP problem, the KAP algorithm to solve the PSP sub-problem, and the SPbP algorithm to solve the 2-PSDPP sub-problem. The results on the same real-world data set as used by Mancini and al. [1] show a very significant improvement and reduce the amount of transferred data up to 30% and a reduction of the computing time up to 15%.

An interesting area for further research may be the improvement of the proposed methods, and the development of other approaches based on constructive heuristics and exact methods that would provide good solutions for other 3-PSDPP sub-problems. It would also seem interesting to use exact methods such as ILP for solving the NP-Hard variant of the PSP sub-problem. Additional research is also required to determine the complexity of the mono-objective sub-problem MCT-PSDPP, in which the completion time Δ is to be minimized.

REFERENCES

[1] S. Mancini and F. Rousseau, "Enhancing non-linear kernels by an optimized memory hierarchy in a high level synthesis flow," in *Proceedings of the Conference on Design, Automation and Test in Europe*. EDA Consortium, 2012, pp. 1130–1133.

[2] ——, "Optimisation d'accélérateurs matériels de traitement par incorporation d'un gestionnaire de données et de contrôle dans un flot de HLS," in *Conférence en Parallélisme, Architecture et Système*, 2013.

[3] C. Tang and E. Denardo, "Models arising from a flexible manufacturing machine, part i: Minimization of the number of tool switches," *Operations Research*, vol. 36, no. 5, pp. 767–777, 1988.

[4] Y. Crama, A. Kolen, A. Oerlemans, and F. Spieksma, "Minimizing the number of tool switches on a flexible machine," *International Journal of Flexible Manufacturing Systems*, vol. 6, no. 1, pp. 33–54, 1994.

[5] J. Bard, "A heuristic for minimizing the number of tool switches on a flexible machine," *IIE Transactions*, vol. 20, no. 4, pp. 382–391, 1988. [Online]. Available: http://dx.doi.org/10.1080/07408178808966195

[6] C. Privault and G. Finke, "Modelling a tool switching problem on a single nc-machine," *Journal of Intelligent Manufacturing*, vol. 6, no. 2, pp. 87–94, 1995. [Online]. Available: http://dx.doi.org/10.1007/BF00123680

[7] G. Laporte, J. Salazar-Gonzalez, and F. Semet, "Exact algorithms for the job sequencing and tool switching problem," *IIE Transactions*, vol. 36, no. 1, pp. 37–45, 2004. [Online]. Available: http://dx.doi.org/10.1080/07408170490257871

[8] A. Konak and S. Kulturel-Konak, "An ant colony optimization approach to the minimum tool switching instant problem in flexible manufacturing system," in *2007 IEEE Symposium on Computational Intelligence in Scheduling*, 2007.

[9] J. Amaya, C. Cotta, and A. Fernández, "A memetic algorithm for the tool switching problem," in *Hybrid metaheuristics*. Springer, 2008, pp. 190–202.

[10] D. Catanzaro, L. Gouveia, and M. Labbé, "Improved integer linear programming formulations for the job sequencing and tool switching problem," *European Journal of Operational Research*, vol. 244, no. 3, pp. 766–777, 2015.

[11] A. Thornton and S. Sangwine, "Log-polar sampling incorporating a novel spatially variant filter to improve object recognition," in *Sixth International Conference on Image Processing and Its Applications*, vol. 2, 1997, pp. 776–779.

[12] N. Bellas, S. Chai, M. Dwyer, and D. Linzmeier, "Real-time fisheye lens distortion correction using automatically generated streaming accelerators," in *17th IEEE Symposium on Field Programmable Custom Computing Machines, FCCM'09.*, 2009, pp. 149–156.

[13] P. Viola and M. Jones, "Robust real-time face detection," *International journal of computer vision*, vol. 57, no. 2, pp. 137–154, 2004.

[14] D. Applegate, R. Bixby, W. Cook, and V. Chvátal, *On the solution of traveling salesman problems*. Rheinische Friedrich-Wilhelms-Universität Bonn, 1998.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 619–626
DOI: 10.15439/2016F389
ACSIS, Vol. 8. ISSN 2300-5963

Heuristic Optimization for the Resource Constrained Project Scheduling Problem: a Systematic Mapping

Aurelia Ciupe, Serban Meza, Bogdan Orza
Technical University of Cluj-Napoca, Multimedia Systems and Applications Lab., 15 Daicoviciu st.
Cluj-Napoca, Romania
Email: {aurelia.ciupe, serban.meza, bogdan.orza}@com.utcluj.ro

***Abstract*— Context: Heuristic optimization has been of strong focus in the recent modeling of the Resource Constrained Project Scheduling Problem (RCPSP), but lack of evidence exists in systematic assessments. New solution methods arise from random evaluation of existing studies. Objective: The current work conducts a secondary study, aiming to systemize existing primary studies in heuristic optimization techniques applied to solving classes of RCPSPs. Method: The systemizing framework consists of performing a systematic mapping study (SM), following a 3-steped protocol. Results: 371 primary studies have been depicted from the multi-stage search and filtering process, to which inclusion and exclusion criteria have been applied. Results have been visually mapped in several distributions. Conclusions: Specific RCPSP classes have been grounded and therefore a rigorous classification is required before performing a systematic mapping. Focusing on recent developments of the RCPSP (2010-2015, a strong interest has been acknowledged on solution methods incorporating AI techniques in meta- and hyper-heuristic algorithms.**

I. Introduction

DEALING with resource allocation in a constrained project scheduling space has risen to a solid multidisciplinary topic under the (project) scheduling theory, addressed since late 60s, as the Resource(-)Constrained Project Scheduling Problem (RCPSP). Empirically, the goal of the RCPSP is to determine a time and resource-feasible schedule, with given precedence or resource constraints, following an objective function. A comprehensive characterization of the RCPSP positioning in the research literature has been given by Mohring et. al (2003), where RCPSP ‘‘one of the most intractable problems in Operations Research’’, has become ‘‘a popular playground for the latest optimization techniques, including virtually all local search paradigms’’.

A. Problem context and motivation

According to the (computational) complexity theory, the RCPSP has been widely characterized as belonging to the class of NP-hard (in the strong sense) problems, with NP-complete (in the strong sense) decision version, therefore combinatorial optimization techniques do apply. Several dimensions, namely research themes, fundament a typology construction for the Project Scheduling Problem (PSP)/ RCPSP space, mainly consisting of: project type, problem approaches, solution methods. Two main implications, can be therefore considered:

1) Proposing new solution methods require a solid mapping of recent and past advancements in solution techniques with an extensive literature search on multiple RCPSP dimensions. In such a context, a literature review process carried out to propose a new solution, involves a cross-domain scanning of a large amount of existing evidence in connected fields (Operations Research, Artificial Intelligence, Applied Mathematics), where computational comparison and benchmarking is of use in performance assessment.

2) Proposing new RCPSP solution methods follows trends of converging research areas, lacking in a systematic assessment in the RCPSP problem space to identify trends and gaps. Current solution methods move forward with connected fields, and include Intelligent Systems and AI techniques, without a stand-alone specific process to identify gaps, based on existing trends. Even though RCPSP research themes have been identified and fully characterized, current secondary studies in RCPSP have the form of reviews, surveys, classifications, and are to be included in the primary studies’ class, as they do not follow a rigorous systemic approach in synthesizing literature review, but provide more of narrative descriptions.

B. Proposed approach and outline

The current work applies the principles of systematic mapping studies (SM) to the RCPSP, with the aim of tracking existing solution methods and identifying trends and gaps based on several studies’ distributions with a focus stands in classic heuristic algorithms and its extended classes (meta-heuristics, hyper heuristics). Section 2 provides an overview of existing guidelines for the systematic mapping and implications in Operations Research and Scheduling. Section 3 describes the research protocol applied and the studies’ frequency resulted in each stage. Section 4 provides an analysis of several studies’ mappings according to time-based and problem space-based distributions. Suggestions for further improvement are described in Section 5.

978-83-60810-90-3/$25.00

II. BACKGROUND AND EVIDENCE

A. Systematic mapping and implications in scheduling

Although large-scale reviews are conducted under the Systematic Literature Review (SLR) umbrella, most of the large-scale SLRs in Software Engineering (SE) and IT are conducted using the Systematic Mapping guidelines [1]. Derived from Kitchenham's guidelines [2], Petersen and Budgen [1], have proposed an extension of guidelines for constructing systematic maps in software engineering. Additionaly, a more detailed presentation of the included studies in Kitchenham's [3] tertiary review, has been offered with respect to the SLRs characteristics. Of relevance, is the comparison between the number of potentially relevant studies and the effective number of relevant studies included in each aformentioned SLRs. According to Petersen [4], Dyba et al. performed a SLR based on 5453 potentially relevant studies, from which 78 have been extracted. The same number of primary studies was used by Hannay et al., Kampenes et al. and Sjøberg et al. with a result of 24, 78 and 103 finally selected studies. MacDonell & Shepperd used the least number of potentially studies (185), to a achieve a relevant set of 10 studies. Most of the aforementioned SLRs have included between 1-5% of the primary studies (Dyba et al. – 1.43%, Grimstad et al. – 2.49%, Hannay et al. – 0.44%, Kampenes et al. – 1.43%, Kitchenham et al. – 0.74%, Sjøberg et al. – 1.88%, MacDonell & Shepperd – 5.45%, Davis et al.- 4.6%), with an exception of Mendes (49%). Correlated to Kitchenham's [3] contribution in aggregating SLRs under a tertiary study, Bugen [5] has applied Kitchenham's guidelines [2] to SMs, and conducted informal tertiary studies.

Latest SM literature, focuses on refining guidelines and protocols, as well as providing improvement strategies. While emphasizing the benefits of using SM in educational environments, as means of building students' transferable research skills, 3 challenges are positively assessed by Kitchenham [6]: 1) no matter the level of studies (post-, undergraduate), students have the required skills of conducting SMs 2) SMs are valuable means of organizing evidence of current state of literature 3) SMs can be treated as solid projects as preparation for research careers. Several further directions, referred as improvements [7], still need to be covered in the future: search and selection process improvement; quality assessment grounding, studies aggregation process improvement. Based on these improvement statements, Petersen [8] offers a valuable update to the existing SM framework applied to software engineering, by mapping evidence regarding 4 research questions: frequency of publications in the SE field, covered topics, revenues of publications, and review process. A need of conducting SMs on existing or extracted topic-specific classifications has been pointed out, observing that the majority of SM studies deliver new classification schemes themselves, rather than building on existing ones.

In terms of extending its field applicability [8], the SM process can be alternatively used as means of evidence gathering in operational fields, for growing disciplines with multiple facets. Applicability of systematic studies to operational fields has been recently acknowledged in the manufacturing [9] and industrial software processes [10]. With respect to the scheduling field, a SLR has been conducted for staffing and scheduling problems in software projects, where 52 papers have been extracted and analyzed with the purpose of identifying main issues in adopting resource scheduling features in industry. Specific to software scheduling, a SLR has been performed [11] addressing uncertainty assessment in software projects, where 165 studies have been analyzed in 5 distributions.

B. Comparative studies in heuristic optimization for solving the RCPSP problem

The Project Scheduling field, specifically the RCPSP, can therefore be introduced as a successful candidate for extending MSs to new fields. In the fields of complexity, discrete and combinatorial mathematics [19], [20], the RCPSP's main characterization is the generalization of the job-shop scheduling problem [21], although the job shop scheduling problem is sometimes treated as a "special case" of the RCPSP [22]. Basically, the optimization problem is to define a start time for each project activity, based on precedence and resource constraints, while following a uni or multi-obective (project makespan minimization, time/cost/resource trade-offs etc).

Large amount of effort has been put during the past decade in building RCPSP classification schemes. Two main dimensions have been extensively tackled: general problem description [12][13] and solution methods[14]–[16], while [17] extends the attributes of the classification scheme to: type of constraints, type of precedence relations, type of resources, source, type of activity splitting, number of execution modes, number of objectives, type of objective function, level of information, distribution of information.

Heuristic methods have been considered as state-of-art solution techniques and a comprehensive description with computational analysis has been presented since late '90 [14]. Based on the Artificial Intelligent and Mathematic Optimization advancements, in the field of Computer Science (CS), classic heuristics applied to RCPSP (exact methods) have evolved into modern heuristics: meta-heuristics (MH) (including evolutionary MH, memetic algorithms, nature-inspired MH), hyper-heuristics, simheuristics and hybrid-heuristics. An attempt of mapping the solution space of the RCPSP, in terms of solution methods, has been presented by [18], where 174 studies (1971-2012) have been classified based on a defined taxonomy. No systematic assessment has been carried, though. Increased interest in the use of meta-heurisitics (both derived and applied or just applied) has been shown during the past 5 years: a computational comparison based on the PSPLIB J30, J60 and J120 problem sets has been presented by [19], [20]. MHs applied to the multi-mode, multi-skill and multi-objective RCPSP have been compared by [21]–[24]. Analysis and comparison of combined meta-heuristics (hybrid-MH) applied to the RCPSP and dynamic RCPSP has been summarized in [21]–[26].

III. Research Methodology

To assess the heuristic techniques applied to solving scheduling problems, with focus on the RCPSP formulations and recent developments, a systematic mapping has been conducted following guidelines of Kitchenham [27] as well as updates provided in [28], [4], [8].

A. Review question

On the basis of systemizing evidence regarding heuristic optimization applied to the RCPSP, the following research question has been considered:

RQ: Which is the evolution of the heuristic algorithms used as solution methods for the RCPSP and to which RCPSPs have they been applied?

B. Search strategy

The search strategy represents a combination between a set of search strings and databases on which to be applied. 4 databases have been selected: IEEE Xplore (http://ieeexplore.ieee.org/Xplore), Elsevier Science Direct (http://www.sciencedirect.com), SpringerLink (http://link.springer.com/), SCOPUS (http://www.scopus.com/) (Table 1). As the RQ of the current study addresses solution methods in forms of techniques and algorithms, IEEE Xplore has been considered relevant. On the other hand, the RCPSP is an OR specific problem, therefore, 2 multidisciplinary databases have been chosen: Springer Link, Science Direct.

TABLE I.
RATIONALE FOR DATABASE CHOICE

Database	Rationale and limitations
IEEE Xplore	multi-format export (includes .csv, .bibtex key), all metadata export (includes abstract); number of entries/export limited at 100
Science Direct	multi-format export (no .csv export, includes .bibtex key), metadata does not include abstract); number of entries/export limited at 100
Springer Link	multi-format export (includes .csv, .bibtex key), metadata does not include abstract); number of entries/export limited at 100
SCOPUS	aggregates literature from multiple databases; multi-format export (includes .csv, .bibtex key), metadata does includes abstract); number of entries/export limited at 100

Six search strings have been constructed based on the PICO strategy [8] (Population, Intervention, Comparison and Outcomes) and submitted for inquiry to each database: S1: project scheduling *heuristic*; S2:"project scheduling" *heuristic*; S3: resource constrained project scheduling *heuristic*; S4: resource-constrained project scheduling *heuristic*; S5: "resource constrained project scheduling" *heuristic*; S6: RCPSP* *heuristic*. The string generation strategy follows the principle of the selection process, where, firstly, a list of the literature-proposed heuristic algorithms was extracted, followed by a list of possible applicable algorithms and a mapping of solution methods on RCPS problem types. No manual search or snowballing has been performed. Special characters have been used to compress search strings, where applicable.

C. Study selection and frequency distribution

The process of study selection as been divided in 5 major steps.

The first phase focuses on preliminary study selection and assessment (Fig.1). Constructed search strings have been applied on the 4 databases individually, as an evidence of the studies frequency, in each case, was of interest. For each search string a list of studies has been exported automatically from each database, using built-in exporting options. Lists have been aggregated for each database (IN1) and duplicates have been removed automatically (IN2). Filtering based on title keywords has been applied ("project" AND "schedul"; RCPS), following a word-form derivation (i.e. "schedul": "scheduling", "schedule (s)", to extract studies relevant for the project scheduling area (IN3A), specific, RCPSP (IN3B). Results have been merged and duplicates removed (IN4). Studies referring to other scheduling scenarios (i.e. machine scheduling, network scheduling, etc.) have been excluded, being of no interest for the current work. A second filtering based on title keywords extracted studies that only address solution techniques. General keywords have been considered (i.e. "heuristic" or "algorithm") (IN5A). Solution-specific algorithms (i.e. metaheuristics optimization algorithm: Tabu Search, Simulated Annealing, Ant Colony Optimization etc.) do not contain any intuitive common keywords, therefore a specific filtering has been applied based on metaheuristic optimization taxonomies [29]–[32](IN5B).

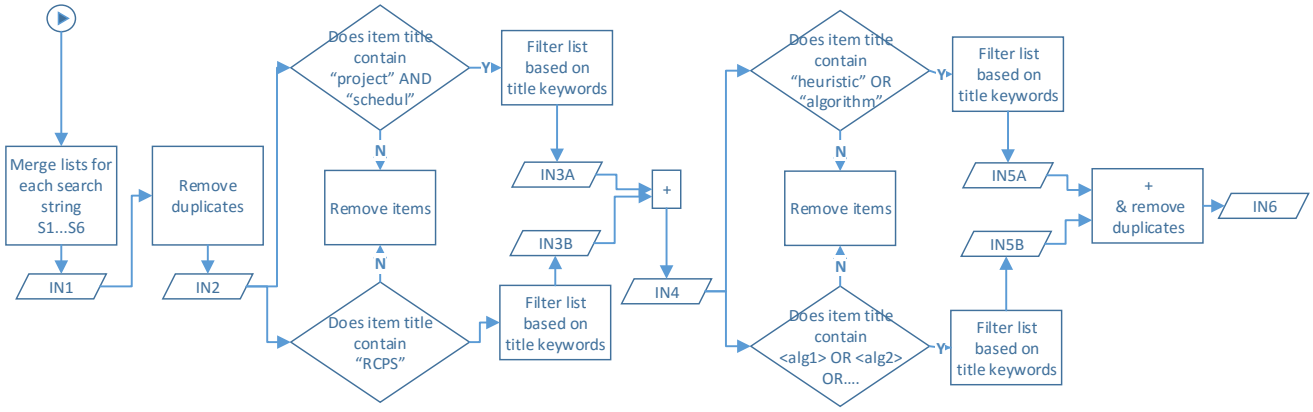


Fig. 1 MS preliminary study selection and assessment workflow

Based on the given references, Table II includes the criteria based on which filtering has been performed to identify specific modern heuristics.

TABLE II.
STUDIED FREQUENCY IN PRELIMINARY SELECTION AND ASSESSMENT

General solution method	Algorithm-specific keywords
General solution method	"heuristic" OR "algorithm"
PS/RCPS Modern Heuristic	"local search" OR "grasp" OR "variable neighbourhood" OR "guided local" OR "iterated local" OR "basic local" OR "simulated annealing" OR "hill climbing" OR "tabu search" OR "random optimization" OR "evolution" OR "genetic" OR "memetic" OR "swarm" OR "stochastic scatter" OR "ant colony" OR "particle" OR "bee colony" OR immune" OR "neural" OR "hybrid"
Modern heuristics	"electromagnetism" OR "frog" OR "multi-pass" OR "filter and fan"

Resulted lists from each database for all search strings (S1, S2…S6) have been merged in a single list and duplicates have been automatically removed (IN6). Studies distribution for each step are presented in Table III.

TABLE III.
STUDIED FREQUENCY IN PRELIMINARY SELECTION AND ASSESSMENT

Database	IN1	IN2	IN3	IN4	IN5	IN6
IEEE Xplore	817	289	127	133	63	51
Science Direct	7633	4768	330	342	102	70
Springer Link	11129	6679	343	373	94	77
SCOPUS	3864	1239	624	659	324	185

Relevancy-based study selection and assessment has been conducted in the second phase (Fig.2). Resulted lists from each database have been merged and duplicates have been removed automatically based on title. A casual manual check has been undertaken to remove duplicates. Filtering based on title keywords has been applied, to extract studies that addressed resource-oriented project scheduling (i.e. "resource"; RCPSP). Both keywords have been separately applied as the studies frequency in each case was of interest. Another relevancy-based criterion tested the relevancy of the solution method. Studies referencing the "algorithm" keyword in their titles have been manually checked for relevancy to the current work. All the 188 entries, have been though found relevant, therefore included in the further process. A total number of 359 entries have been obtained and submitted as input for the next phase

Language and content type exclusion (EC) and inclusion criteria (IC) have been applied in the 3rd step. The language criterion has been considered eliminatory. From entries that had language specified in metadata, studies in other languages than English, such as Spanish, Chinese have been excluded, (EC). From the resulted entries, only English studies were extracted (IC). In case of filled in language attribute, that did not respect the EC, the language EC was updated and entries re-checked to pass the IC. Entries with blank language field, have been checked manually based on the publication language. The language field has been updated for all the remained items. All studies with updated language field have been English-written, therefore successfully passed EC/IC check and included in the final list. A list of 339 studies have been obtained by applying the language EC/IC.

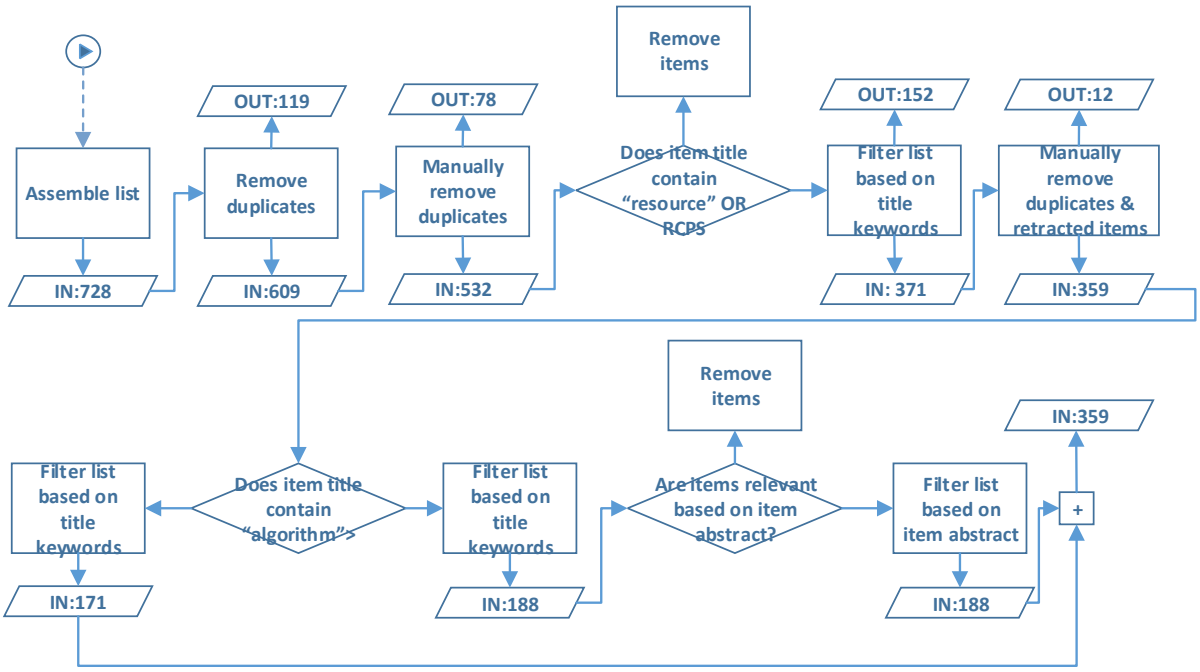


Fig. 2 MS relevancy study selection and assessment workflow

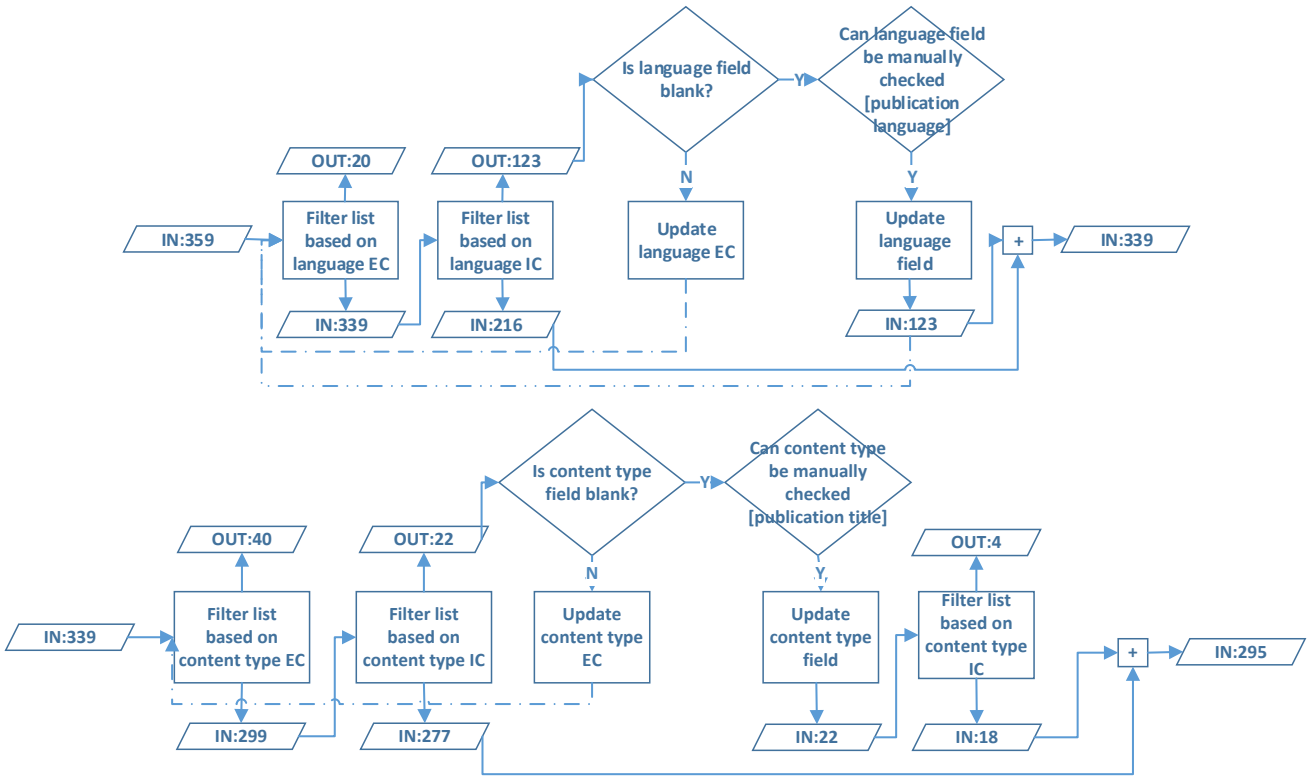


Fig. 3 MS relevancy study selection and assessment based on a) language IC/EC b) content type IC/EC

The next phase, consisted of applying a content type-based language EC/IC, in a similar way to the language-based EC/IC. Studies that have been categorized as technical papers or reports, reviews, workshop papers or book chapters have been excluded (EC). Papers published in peer-reviews publications (journals) or conference proceedings have been included (IC). Papers not categorized have been submitted for manual checking based on the "Publication title" attribute. 22 studies had the content type updated. Based on the IC criteria, 18 studies have been selected and 4 removed. A total number of 295 have been therefore obtained. Both phases of EC/IC (language and content type) are illustrated in Fig.3.

A manual refinement has been undertaken as a last step of the selection process, based on abstracts' screening. 4 studies had the full content published in Chinese and have been excluded based on language (4 studies). Studies that were proposing new solution approaches were of interest, therefore, (computational or experimental) evaluation or comparison papers, as well as classification papers, have been excluded (15 studies). Where in doubt, full text has been screened. Generalized studies that have not addressed a particular RCPS problem or solution method (i.e. generalized algorithms not addressing any heuristic/meta-heuristic solution method), as well as studies that could not be positioned in a classification scheme (either irrelevant or not accessible) have also been excluded (10 studies). A final list of 252 studies has been furtherly considered for mapping and analysis.

IV. Studies Analysis and Mapping

To answer the proposed research question (RQ), several types of mappings have been extracted: studies identification on RCPS problem types and solution algorithms classes, frequency of publications for classes of RCPSP solution methods and algorithms, as well as algorithm mapping on the specific RCPS problem types. Statistics have been extracted based on title, abstract or brief content screening (when required), without a full evaluation, as in case of SLRs [4].

In terms of visual representations, several assumptions have been made in the SM/SLR literature. With respect to SLRs, Kitchenham [27] emphasizes on the use of forest plots as being the most common mechanism for presenting quantitative results, while vulnerability to publication bias is likely to be assessed by using funnel plots. On the other hand, Petersen [8], in the study of SM guidelines updating, identifies 6 approaches to visualize SM mappings. Based on the analysis, bar plots and bubble plots proved to be the most common representations, while heatmaps, although rare, were considered interesting ways to directly visualize the relative amount of publications in different categories [8]. Based on the above observations, the current study proposes bar charts and bubble plots as visual representations. Three categories of mapping have been considered relevant to identify the evolution of heuristic optimization applied to the RCPSP: a) distribution of heuristic classes on years b) distribution of heuristic algorithms on years c) heuristic algorithms solving classes of RCPSP problems

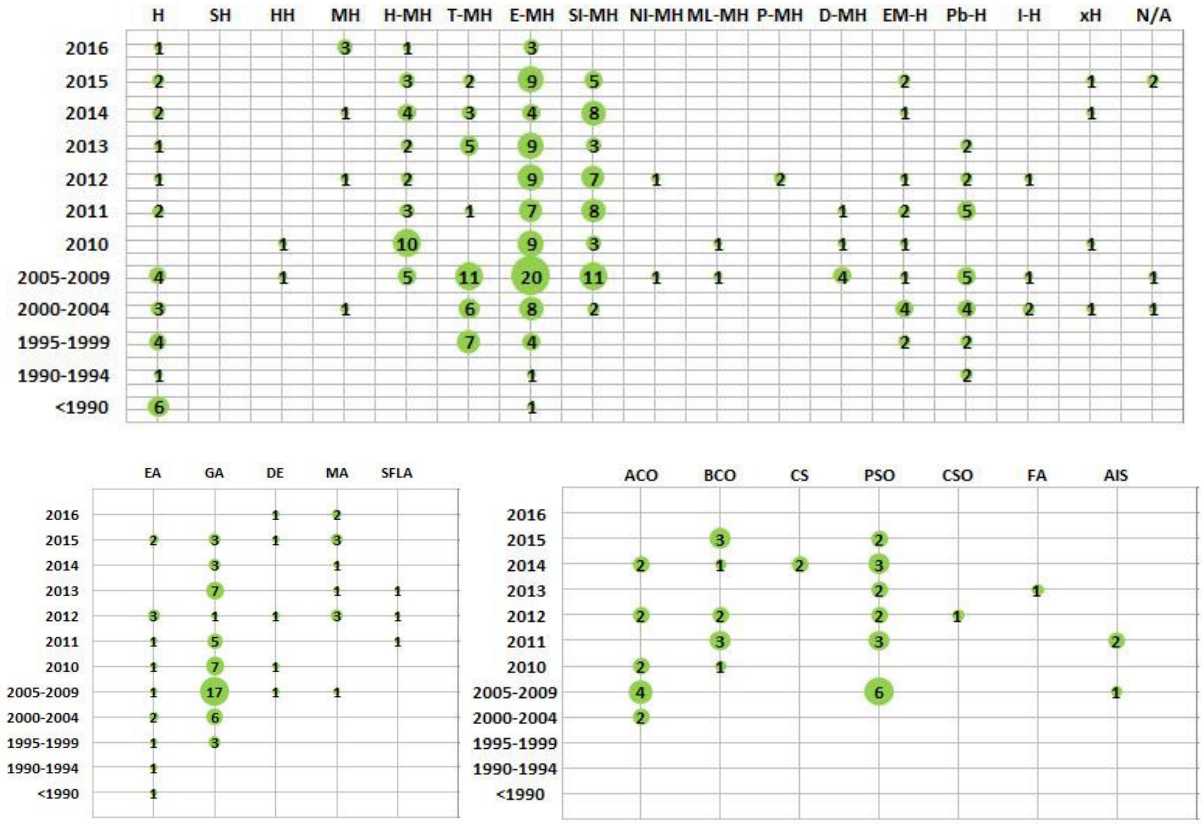


Fig. 4 Distribution of a) classes of heuristic algorithms on years b) Evolutionary methods on years c) Swarm Intelligence algorithms on years

Several year-based scales have been considered to be relevant for the studies mapping: unique discrete year values have been considered for the 2010 – 2016 (first decade) interval; 4 intervals have been considered between 1990 – 2009; one category for studies published before 1990. Acronyms have been used for RCPS problems definition or heuristic optimization classes and algorithms, following existing taxonomies. A bubble chart representation has been preferred, according to updated guidelines [8], where bubble plots are recommended ways of representing frequency distributions [4]. Bubbles have been positioned in the intersection of two x-y scatter plots. The size of the bubble indicates the number of studies for the given intersection point.

The first mapping, answers the given RQ by presenting a time-distribution of classes of heuristic optimization algorithms (Fig.4). The mapping process was conducted based on title, and when in doubt, the abstract was scanned. Studies addressing more than one solution method have been mapped to multiple categories. Several classes of algorithms have been identified: 2 classes referencing more generalized studies (H: general heuristics, MH: general metaheuristics); 2 classes of algorithms addressing extended heurisitics (SH: simheuristics, HH: hyper-heuristics); xH: other heuristics; N/A: not assigned heuristics. All the other classes (H-MH: hybrid MH; T-MH:trajectory-based MH; E-MH:evolutionary MH; SI-MH:swarm intelligence-based MH; NI-MH:nature-inspired MH excluding SI-MH; ML-MH: machine learning-inspired MH; P-MH: probability-based MH; D-MH: deterministic MH; EM-H: exact methods H; Pb-H: priority-based H; I-H: improvement H). Studies have been included in a separate category, when more than one algorithm was addressed as “improvement” of one class, and properties of H-MH; HH, MA (memetic algorithms) were applied (screening of abstracts). 252 studies have been extended to 288, some entries presenting more than one algorithm.

2 metaheuristic categories proved to be dominant: E-MH (29.1%) and SI-MH (16.3%), followed by the class of T-MH (11.4%) and H-MH (10.4%). In both cases, an increased interested is shown during 2013-2016: E-MH (25 studies), SI-MH (16 studies) compared to the period of 2005-2009, referenced in the literature: E-MH (20 studies), SI-MH (16 studies). From the state-of-the art heuristics, as advertised, Pb-H are the most popular algorithms, including: scheduling schemes and priority rules (7.6%).

Both E-MH and SI-MH classes have been decomposed on specific optimization algorithms (Fig.4.b and Fig.4.c). Studies belonging to the E-MH classes included solution methods such as: EA (evolutionary algorithms and strategies), GA (genetic algorithms), DE (differential evolution); MA (memetic algorithms), SFLA (shuffled frog leaping algorithm). Studies categorized as belonging to the class of SI-MH, addressed algorithms such as: ACO (ant colony optimization), BCO (bee colony optimization); CS (cuckoo search), CSO (cat swarm optimization), FA (firefly algorithm), AIS (artificial immune system).

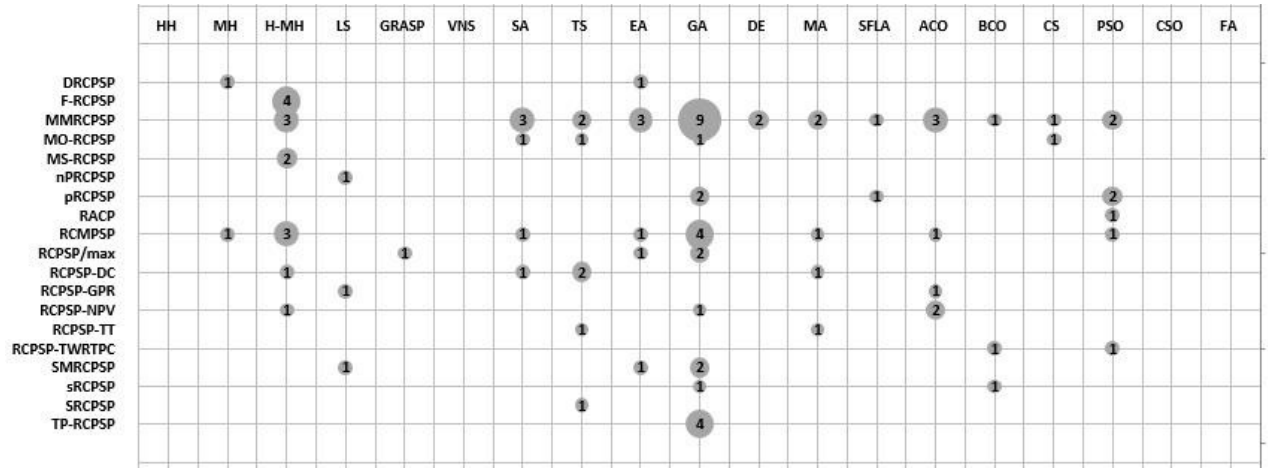


Fig. 5 Mapping of modern heuristic algorithms on RCPS classes of problems

According to the year-based distributions, preferred algorithms are GA (52 studies) and PSO (18 studies), followed by EA (13 studies) and ACO (12 studies). Fewer attempts for developing RCPSP solutions included: derivations of CSO, CS, FA. Compared to other classes of algorithms, the SI-MH appear to be a recent class of solving algorithms, having a debut in the 2000-2004 period. Several independent studies present directions for future interest: a) despite the specific identified algorithms, 2 studies have been depicted addressing solving approaches based on learning strategies b) 1 agent-based solution method.

According to the RCPSP solution space, a solution framework consists of both solution method and the problem definition addressed. Fig.4. maps the distribution of general modern metaheuristics (HH, MH, H-MH), the specific algorithms belonging to the E-MH, SI-MH classes, as well as T-MH class (LS: local search, GRASP: greedy randomized adaptive search, VNS: variable search neighborhood, SA: simulated annealing, TS: tabu search) on RCPS problems. 19 specific problem classes have been identified in the extracted list of entries, based on existing classification schemes [17]. Additionally, 91 studies addressed a general class of the RCPSP, while 62 studies did not address the specific RCPSP. 4 classes of algorithms, where found of not addressing any RCPS problem (xH, HH, Ni-MH, xH). The frequency distribution presents the MMRCPSP (multi-mode RCPSP) (14.2%) and the RCMPS (mult-project RCPSP) (6.2%) as the most addressed problems. Solution methods for the MMRSPSP are popular within E-MH (17 studies) and SI-MH (7 studies). In terms of general heuristic optimization classes, H-MH (14 studies) algorithms prove an increased popularity in addressing specific RCPS problems (F-RCPSP, MMRCPSP, MS-RCPSP), comparable to the T-MH class.

V. Conclusions

While current studies that seek to provide a representation for the RCPS problem space are more oriented on taxonomy construction, the current work adopts the structured process of SM studies to provide an overview and decomposition on the evolution of solution methods and their applicability to specific RCPS problems. Several contributions are to be claimed: 1) providing a framework under which the SMs can be extended to other fields, specifically validation of SM applicability to the RCPSP 2) validation of the solution methods and problems classification based on an extensive database of studies 3) trend identification and decomposition based on existing solution evolution. To reduce bias in assessment we recommend a further evaluation of quality metrics, an optional step in conducting SMs (opposite to SLRs, where quality assessment is mandatory to validate study selection and filtering). A second proposed improvement addresses construction of initial search strings based on words synonymy or derived attributes (i.e. “strategies”, “methods” when referring to “algorithms”). For a complete validation and for the purpose of transforming a SM in a work of reference for the RCPSP, computational comparison is required to evaluate algorithms performance, based on specific setups. Several existing RCPSP libraries (PSPLIB, Peterson set) provide project descriptions (number of activities, types of activities and resources) on which proposed algorithms are tested for computational parameters (average/min/std. deviation, computational time etc), in different scenarios (number of iterations). Therefore, the current study provides the input for a further SLR, being a successful candidate for a meta-analysis, that would provide a more solid background in algorithmic benchmarking.

VI. References

[1] M. Turner, B. Kitchenham, D. Budgen, and O. P. Brereton, *Lessons Learnt Undertaking a Large-Scale Systematic Literature Review*, Proc. EASE, 2008, pp. 110–118.

[2] B. Kitchenham and S. Charters, *Guidelines for performing Systematic Literature Reviews in Software Engineering*, 2007.

[3] B. Kitchenham, O. Pearl Brereton, D. Budgen, M. Turner, J. Bailey, and S. Linkman, *Systematic literature reviews in software engineering - A systematic literature review,* Inf. Softw. Technol., 2009, vol. 51, no. 1, pp. 7–15

[4] K. Petersen, R. Feldt, S. Mujtaba, and M. Mattsson, *Systematic mapping studies in software engineering*, EASE’08 Proc. 12th Int. Conf. Eval. Assess. Softw. Eng., 2008, pp. 68–77.

[5] D. Budgen, T. Mark, P. Brereton, and B. Kitchenham, *Product-Focused Software Process Improvement*, Proc. PPIG, 2009, vol. 32, pp. 195–204.
[6] B. Kitchenham, P. Brereton, and D. Budgen, *The educational value of mapping studies of software engineering literature*, 2010 ACM/IEEE 32nd Int. Conf. Softw. Eng., 2010, vol. 1, pp. 589–598.
[7] B. Kitchenham, *Systematic review in software engineering: where we are and where we should be going*, Proc. 2nd Int. Work. Evidential Assess. Softw. Technol. (EAST '12), 2012, pp. 1–2.
[8] K. Petersen, S. Vakkalanka, and L. Kuzniarz, *Guidelines for conducting systematic mapping studies in software engineering: An update*, Inf. Softw. Technol., 2015, vol. 64, pp. 1–18.
[9] A. Negahban and J. S. Smith, *Simulation for manufacturing system design and operation: Literature review and analysis*, J. Manuf. Syst., 2014, vol. 33, no. 2, pp. 241–261.
[10] N. Bin Ali, K. Petersen, and C. Wohlin, *The Journal of Systems and Software A systematic literature review on the industrial use of software process simulation*, J. Syst. Softw., 2014, vol. 97, pp. 65–85.
[11] M. Marinho, S. Sampaio, T. Lima, and H. De Moura, *A Systematic Review Of Uncertenties in Software Project Management Projects*, International Journal of Software Engineering & Applications (IJSEA), 2014, vol. 5, no. 6, pp. 1–21.
[12] W. Herroelen, E. Demeulemeester, and B. Reyck, *A classification scheme for project scheduling*, Proj. Sched., 1999, vol. 14, no. 9727, pp. 1–26.
[13] P. Brucker, A. Drexl, M. Rolf, E. Pesch, and K. Neumann, *Resource-constrained project scheduling: Notation, classification, models and methods*," 1999, vol. 112, pp. 3-41
[14] R. Kolisch and S. Hartmann, *Heuristic Algorithms for the Resource-Constrained Project Scheduling Problem: Classification and Computational Analysis*, Proj. Sched. SE, 1999, vol. 14, pp. 147–178.
[15] R. Kolisch, *Experimental evaluation of state-of-the-art heuristics for the resource-constrained project scheduling problem,* 2000, vol. 127, pp. 394–407,
[16] R. Kolisch and S. Hartmann, *Experimental investigation of heuristics for resource-constrained project scheduling: An update*, Eur. J. Oper. Res, 2006, vol. 174, no. 1, pp. 23–37.
[17] C. Schwindt and J. Zimmermann, *Handbook on Project Management and Scheduling Vol.1*", Eds. Cham: Springer International Publishing, 2015, pp. 57–74.
[18] M. Abdolshah, *A Review of Resource-Constrained Project Scheduling Problems (RCPSP) Approaches and Solutions*, Int.Trans. Journal of Engineering, Management, & Applied Sciences & Technologies 2014, vol. 5, no. 4, pp. 253 - 286
[19] P. P. Das and S. Acharyya, *Meta-heuristic approaches for solving Resource Constrained Project Scheduling Problem: A Comparative study*, Comput. Sci. Autom. Eng. (CSAE), 2011 IEEE Int. Conf., 2011, vol. 2, pp. 474–478.
[20] A. Lim, H. Ma, B. Rodrigues, S. T. Tan, and F. Xiao, *New meta-heuristics for the resource-constrained project scheduling problem*, Flex. Serv. Manuf. J., 2011, pp. 48–73.
[21] P. Myszkowski, *Novel heuristic solutions for Multi-Skill Resource-Constrained Project Scheduling Problem*, Comput. Sci. Inf. Syst., 2013, pp. 159–166.
[22] H. Cristiano and F. De Assis, *Multi-objective metaheuristic algorithms for the resource-constrained project scheduling problem with precedence relations*, Comput. Oper. Res., 2014, vol. 44, pp. 92–104.
[23] V. Van Peteghem and M. Vanhoucke, *An experimental investigation of metaheuristics for the multi-mode resource-constrained project scheduling problem on new dataset instances*, Eur. J. Oper. Res., 2014, vol. 235, no. 1, pp. 62–72.
[24] M. Beckmann, H. P. Kiinzi, F. Wirtschaftswissenschaften, and F. Hagen, Lecture Notes in Economics and Mathematical Systems.
[25] C. Tchao and S. L. Martins, *Hybrid heuristics for multi-mode resource-constrained project scheduling*, Learning and Intelligent Optimization, Springer, 2007, pp. 234–242.
[26] R. Villela and L. S. Ochi, *Hybrid Heuristics for Dynamic Resource-Constrained Project Scheduling Problem*, Lecture Notes in Computer Science, 2010, Vol.6373, pp 73-87.
[27] B. Kitchenham, *Procedures for performing systematic reviews*, Keele, UK, Keele Univ., 2004, vol. 33, no. TR/SE-0401, p. 28.
[28] B. Budgen, David Turner, Mark Brereton, Pearl Kitchenham, *Using Mapping Studies in Software Engineering*, Proc. PPIG, 2008, vol. 2, pp. 195–204
[29] S. Binitha and S. S. Sathya, *A Survey of Bio inspired Optimization Algorithms*, Int. J. Soft Comput. Eng., 2012, vol. 2, no. 2, pp. 137–151.
[30] A. Colorni, M. Dorigo, F. Ma, V. Maniezzo, G. Righini, M. Trubian, and P. Milano, *Heuristics from Nature For Hard Combinatorial Optimization Problems*, Int.Transactions on Operational Research, 1996, pp. 1–38
[31] E. Talbi, *A Taxonomy of Hybrid Metaheuristics*, J. of Heuristics, 2002, vol. 45, pp. 1–45.
[32] S. Nesmachnow, *An Overview of Metaheuristics: Accurate and Efficient Methods for Optimisation*, Int. J. Metaheuristics, 2014, vol. 3, no. 4, pp. 320–347.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 627–635
DOI: 10.15439/2016F447
ACSIS, Vol. 8. ISSN 2300-5963

Minimizing Total Completion Time in Flowshop with Availability Constraint on the First Machine

Yumei Huo
Department of Computer Science
College of Staten Island, CUNY
Staten Island, New York 10314, USA,
Email: yumei.huo@csi.cuny.edu

Hairong Zhao
Department of Mathematics, Computer Science & Statistics
Purdue University Northwest
2200 169th Street, IN 46323, USA
Email: hairong@pnw.edu

***Abstract*—We study the problem of minimizing total completion time in 2-stage flowshop with availability constraint. This problem is NP-hard in the strong sense even if both machines are always available. With availability constraint, although a bulk of research papers have studied the makespan minimization problem, there is no research done on the total completion time minimization. This paper is the first attempt to tackle this problem. We focus on the case that there is a single unavailable interval on the first machine only. We show that several special cases can be solved optimally or approximated within a constant factor. For the general case, we develop some lower bounds and dominance rules. Then we design and implement a branch and bound algorithm. We investigate the effectiveness of different lower bounds and the dominance rules by computational experiments. We also study how the start time and the duration of the unavailable interval affects the efficiency of the branch and bound algorithm.**

I. Introduction

SCHEDULING with machine availability constraint has attracted more and more research effort since early nineties. Machine availability constraint is prevalent in all real industrial settings. A machine may be unavailable due to breakdown, preventive maintenance, or processing unfinished jobs from a previous scheduling horizon. With the machine availability constraint, many classical problems have to be reconsidered. While some of the problems can still be easily solved, many of them become more complicated and new optimal algorithms/heuristics need to be designed. Many papers have been devoted to various scheduling problems with machine availability constraint, see the survey papers by Lee, Lei and Pinedo ([18]), Sanlaville and Schmidt ([21]), Schmidt ([22]), Ma, Chu and Zuo ([20]), etc.

For flowshop scheduling with availability constraint, most research is done for two-stage flowshop problems([20]): there are two machines, machine 1 and machine 2. Each machine may have one or more unavailable intervals; There are n jobs, 1, 2, ..., n. Each job j, $1 \leq j \leq n$, has two operations a_j and b_j which have to be processed on machine 1 and on machine 2, respectively. The operation b_j cannot start on machine 2 before a_j finishes on machine 1. We want to find a schedule of the jobs so that some objectives are optimized. The bulk of flow shop research in the last decades has been focused on the minimization of the maximum of the job completion time, i.e. the length or makespan of a schedule. However, Gupta and Dudek [8] pleaded that criteria in which the costs of each job are reflected have a better economic interpretation than the makespan objective has. This paper deals with the minimization of the sum of the job completion times in a two-machine flow shop.

Based on the effect of the availability constraint to the disrupted job's processing, researchers discuss three cases, namely resumable, nonresumable, and semiresumable. Resumable and nonresumable cases are defined by Lee in [16]. In the resumable case, preemption is allowed thus jobs can be resumed after being interrupted by the unavailable interval. In the nonresumable case, a job has to be restarted if interrupted by the interval. Semiresumable case is defined by Lee in [19] and is between the resumable case and nonresumble case. In this case, a job doesn't have to be restarted from scratch, instead, a fraction of the job needs to be reprocessed after the machine is available.

A. Literature Review

Almost all research papers about flow shop scheduling with availability constraint are with respect to makespan criterion ([20]). Lee ([17]) showed that if there is only one unavailable interval, either on the first machine or on the second machine,the problem is NP-hard in the ordinary sense. If there are an arbitrary number of unavailable intervals on the first machine but the second machine is always available, or there there are an arbitrary number of unavailable intervals on the second machine but the first machine is always available, the problem becomes strongly NP-hard ([15]) in either case. Furthermore, for the former case, Johnson's rule gives a 2-approximation regardless of the number of unavailable intervals; On the other hand, for the latter case, there is no polynomial time constant approximation for any constant even if there are only two unavailable intervals on the second machine and the first machine is always available. Many heuristics, meta-heuristics and exact methods are developed for the general problem, see the surveys [18],[21], [22], [20] and the references therein for more details.

For the problem of minimizing the total completion time with availability constraint, we are not aware of any research so far. On the other hand, when both machines are available, the problem has attracted a lot of attention because of its

notorious intractability. A lot of research effort has been focused on finding the exact solution using branch and bound algorithm. In the following, we will review related literature.

The first research on this problem was done by Ignall and Schrage ([13]) in 1965. They gave two lower bounds and developed a branch-and-bound algorithm. For the experiments, they limited the number of jobs to be 10. In 1967, Conway et. al ([3]) showed that it is sufficient to study permutation schedules, i.e. a schedule in which both machines have the same job sequence with no unnecessary idle time between operations. By using local search and other heuristics to generate a good initial upper bound for the branch and bound algorithm, Kohler and Steiglitz ([14]) further improved the algorithm by Ignall and Schrage. They did the experiments on instances of size 10 to 50 jobs. For most instances of more than 15 jobs, only suboptimal solutions is obtained within preset time limit. In 1976, this problem was shown to be NP-Hard in the strong sense by Gary, John and Sethi([6]).

Since 1990s, the problem was studied again by a group of research papers, including [23], [10], [5], [4], [1], [12]. These papers try to improve the lower bounds by using Lagrangian relaxation ([23], [10]) or networking formulation ([1]), and/or propose some dominance rules ([5],[4]), consequently improve the performance of the branch-and-bound procedure and solve bigger problems. Della Croce et al.'s ([4]) branch-and-bound algorithm can solve up to 45 (30) job problems when processing times are uniformly distributed in the [1,10] ([1,100]) range. In 2004, using a lower bound scheme based on a network formulation, the branch and bound algorithm by Akkan and Karabati ([1]) can solve problems with up to 60 (45) jobs, where processing times are uniformly distributed in the [1,10] ([1,100]) range. At about the same time, Hoogeveen et al. ([12]) also used improved lower bounds by LP to solve instances of 40 jobs within reasonable time.

Very few papers studied the constant approximation algorithms or solvable cases. The first one is by Gonzales and Sahni ([7]) who showed that SPT (Shortest Processing Time first) rule gives an m-approximation for m stage flowshops. Thus for 2-stage flowshop, SPT is a 2-approximation. Later, Hoogeveen and Kawaguvhi ([9] refined this bound for 2-stage flowshop and they showed that the approximation ratio of SPT is $2\beta/(\alpha+\beta)$, where $\alpha = \min\{a_j, b_j\}$, $\beta = \max\{a_j, b_j\}$. They also studied some special cases. Specifically, they showed that

1) if $a_j = a$ for all jobs j, the problem remains NP-hard in the strong sense and SPT rule gives $4/3$- approximation schedule and the bound is tight.
2) if $b_j = b$ for all jobs j, then SPT rule generates an optimal schedule.
3) if $a_j \geq b_j$ for all jobs j, scheduling the jobs in non-decreasing order of a_j gives an optimal schedule.
4) if $a_j \leq b_j$ for all jobs j, the problem can be solved optimally in $O(n^3)$ time.

B. New Contributions

Our paper is the first attempt to tackle the total completion time minimization problem in the 2-stage flowshop with availability constraint. Given the complexity of the problem, we focus on the case that there is a single unavailable interval on the first machine. We consider the resumable case only. We first show that SPT still provides a 2-approximation if the single unavailable interval starts early so that no a_j can finish before it. We also show that some other special cases can be solved or approximated within a constant factor.

For the general case, we give some lower bounds and dominance rules and develop a branch and bound algorithm. We investigate the effectiveness of different lower bounds and the dominance rules by computational experiments. We also study how the start time and the duration of the unavailable interval affects the efficiency of the branch and bound algorithm.

C. Organization

The paper is organized as follows. We first give some preliminary results in Section II. In Section III, we give a mathematical formulation of the problem and develop five lower bounds of the optimal solution. In Section IV, we develop some dominance rules and a branch and bound algorithm. Then we give analysis of the experimental results. Finally, we conclude in section V.

II. Preliminary results

Let us first introduce some notations. To indicate the machine unavailability constraint, most literature extends the $\alpha \mid \beta \mid \gamma$ notation, by adding two new components. In the α field, h_{ik} represents the problem with k unavailable intervals on the ith machine. So h_{11} and h_{21} represents the problem with one unavailable interval on the first machine and on the second machine respectively. The second component is in the β field, we use $r-a$, $nr-a$ and $sr-a$ to denote resumable, nonresumable and semiresumable availability constraints, respectively. Thus, our problem, minimizing the total completion time with a single unavailable interval on machine 1, can be denoted as $F2, h_{11} \mid r-a \mid \sum C_j$.

Like the classical model, one can show that there is an optimal schedule that is a permutation schedule. So we only consider permutation schedules. Let S be a permutation schedule of the jobs, for job j, we use $C_{a,j}(S)$, $C_{b,j}(S)$ to denote the completion time of the first operation a_j and the second operation b_j in S, respectively. The completion time of job j in S, is denoted by $C_j(S)$ which is equal to $C_{b,j}(S)$. We use $[j]$ to denote the job scheduled in position j on both machines, and use $C_{a,[j]}(S)$ and $C_{b,[j]}(S)$ to represent the completion of $a_{[j]}$ and $b_{[j]}$ in S, respectively. If it is clear from the context, S may be omitted.

We frequently sort the jobs, for convenience, we use SPT_a, SPT_b, SPT to denote the rule that schedules the jobs in non-decreasing order with respect to a_j, b_j, and $(a_j + b_j)$ respectively.

From [7], we know that SPT is a 2-approximation algorithm when both machines are available. When there is a single

unavailable interval, we show in the following that the performance of SPT depends on the start time of the interval.

Lemma 2.1: Let $[s,t]$ be the unavailable interval for $F2,h_{11} \mid r-a \mid \sum C_j$, if $s \geq \sum_{j=1}^{n} a_j$ or $s < \min\{a_j\}$, SPT is a 2-approximation.

Proof: First, if $s \geq \sum_{j=1}^{n} a_j$, then all a_j-s can finish before the unavailable interval, thus the interval has no effect on the jobs at all. By [7], SPT is a 2-approximation.

Now, let us consider the case $s < \min\{a_j\}$. Suppose the jobs are indexed so that $(a_j + b_j) \leq (a_{j+1} + b_{j+1})$ for $1 \leq j \leq n-1$. We denote $P1$ to be the problem of scheduling of n jobs with processing time $(a_j + b_j)$ on a single machine with the unavailable interval $[s,t]$. Then one can easily show that the optimal schedule S_1 for $P1$ is obtained by SPT rule, and the completion time of j-th job will be $C_{[j]}(S_1) = t + \sum_{k=1}^{j}(a_k + b_k)$.

Let S be the schedule generated by SPT rule for $F2,h_{11} \mid r-a \mid \sum C_j$. Apparently, we have

$$C_j(S) \leq C_{[j]}(S_1) = t + \sum_{k=1}^{j}(a_k + b_k).$$

Now, consider the relaxed flowshop problem $P2$ where a_j and b_j can be concurrently executed and the completion time is the time when both a_j and b_j finish. Let S_2 be the optimal schedule for this relaxed flowshop problem. We must have that

$$C_{[j]}(S_2) \geq \tfrac{1}{2}(t + \sum_{k=1}^{j}(a_{[k]} + b_{[k]})) \geq \tfrac{1}{2}(t + \sum_{k=1}^{j}(a_k + b_k)).$$

Let S^* be the optimal schedule for $F2,h_{11} \mid r-a \mid \sum C_j$. Then we have

$$C_{[j]}(S^*) \geq C_{[j]}(S_2) \geq \frac{1}{2}(t + \sum_{k=1}^{j}(a_k + b_k)) \geq \frac{1}{2}C_{[j]}(S).$$

Thus, $\sum_{j=1}^{n} C_j(S) \leq \sum_{j=1}^{n} 2C_{[j]}(S^*)$. This completes the proof. ∎

Unfortunately, if s is arbitrary, SPT may generate a schedule that has a very large approximation ratio.

Lemma 2.2: For $F2,h_{11} \mid r-a \mid \sum C_j$ with a single unavailable interval $[s,t]$, if $s \geq a_1$ where $a_1 + b_1 = \min\{a_j + b_j\}$, SPT can generate a schedule whose approximation ratio is n in the worst case.

Proof: Given an instance I of the problem with a single unavailable interval $[s,t]$. Let I' be the corresponding instance of $F2 \mid\mid \sum C_j$ which is obtained from I by removing the unavailable interval. Let C_{opt} and C'_{opt} be the minimum total completion time for I and I', respectively. It is obvious that $C'_{opt} \leq C_{opt}$.

Let S be the schedule generated by SPT rule for instance I and S' be the SPT schedule for I'. For convenience, suppose the jobs are indexed in non-decreasing order of $(a_j + b_j)$, $1 \leq j \leq n$. Let i be the last job such that $C_{a,i}(S) \leq s$. Since $a_1 \leq s$, $i \geq 1$. It is clear that for $j \leq i$, we have $C_j(S) = C_j(S')$; for all $j > i$, we have $C_{a,j}(S) = C_{a,j}(S') + (t-s)$, so $C_j(S) = C_{b,j}(S) \leq C_{b,j}(S') + (t-s)$. Thus,

$$\begin{aligned}\sum_{j=1}^{n} C_j(S) &\leq \sum_{j=1}^{i} C_j(S') + \sum_{j=i+1}^{n}(C_j(S') + (t-s)) \\ &\leq \left(\sum_{j=1}^{n} C_j(S')\right) + (n-i)(t-s).\end{aligned}$$

On the other hand, since $\sum_{j=1}^{n} a_j > s$, at least one job must finish after t in any schedule. Thus the completion time of the last job in the optimal schedule is

$$C_{[n]}(S^*) \geq ((t-s) + \sum_{j=1}^{n}(a_j + b_j))/2 \geq \tfrac{(t-s)}{2} + \tfrac{C_n(S')}{2}.$$

For j-th job in S^*, $j \neq n$, we have that $C_{[j]}(S^*) \geq C_j(S')/2$. Thus we have $\sum_{j=1}^{n} C_j(S^*) \geq \sum_{j=1}^{n} C_j(S')/2 + (t-s)/2$ and

$$\begin{aligned}\sum_{j=1}^{n} C_j(S) &\leq \left(\sum_{j=1}^{n} C_j(S')\right) + (n-i)(t-s) \\ &\leq (2\left(\sum_{j=1}^{n} C_j(S^*) - (t-s)\right) + (n-i)(t-s) \\ &\leq 2\sum_{j=1}^{n} C_j(S^*) + (n-i-1)(t-s) \\ &\leq (n-i+1)\sum_{j=1}^{n} C_j(S^*).\end{aligned}$$

Since we assume $i \geq 1$, $\sum_{j=1}^{n} C_j(S) \leq nC_{opt}$. ∎

Next, we show that for some special cases in terms of a_j-s and/or b_j-s, we can use SPT_a or SPT_b to solve the problem optimally or get a good approximation.

Lemma 2.3: SPT_a generates a schedule whose total completion time is minimum for

(a) $F2,h_{11} \mid r-a, b_j = b \mid \sum C_j$; and
(b) $F2,h_{11} \mid r-a, a_j \geq b_j \mid \sum C_j$

Proof: First consider case (a): $F2,h_{11} \mid r-a, b_j = b \mid \sum C_j$. Let S be an arbitrary schedule and suppose that the jobs are scheduled in the order of 1, 2, $\cdots$. Define $C_0 = 0$, then the completion time of job i in S is

$$C_i(S) = \max(C_{i-1}, C_{a,i}) + b,$$

where

$$C_{a,i} = \begin{cases} \sum_{j=1}^{i} a_j, & \text{if } \sum_{j=1}^{i} a_j \leq s \\ \sum_{j=1}^{i} a_j + (t-s), & \text{if } \sum_{j=1}^{i} a_j > s. \end{cases}$$

Apparently scheduling the jobs in SPT_a minimizes the total completion time.

Now we consider case (b): $F2,h_{11} \mid r-a, a_j \geq b_j \mid \sum C_j$. Suppose that $a_1 \leq a_2 \leq \cdots \leq a_n$. Since $a_j \geq a_{j-1} \geq b_{j-1}$, in the schedule generated by SPT_a, b_j can be scheduled immediately after a_j completes, thus minimizing the total completion

time is the same as minimizing the total completion time of the a_j-s which is obtained by SPT_a rule.

■

Lemma 2.4: SPT_b generates a schedule for $F2, h_{11} \mid r - a, a_j = a \mid \sum C_j$ whose total completion time is at most $7/3$ times that of the optimal schedule.

Proof: Given an instance I of problem $F2, h_{11} \mid r - a, a_j = a \mid \sum C_j$ with the unavailable interval $[s, t]$, let C_{opt} be its minimum total completion time for I. Let S be the schedule generated by the SPT rule. Let I' be the instance of $F2 \mid a_j = a \mid \sum C_j$ obtained from I by removing the unavailable interval. Let C'_{opt} be the minimum total completion time for I'. Apparently, we have $C'_{opt} \leq C_{opt}$. Let S' be the SPT schedule for I'. From [9], we know that $\sum_{j=1}^{n} C_j(S') \leq \frac{4}{3} C'_{opt} \leq \frac{4}{3} C_{opt}$. Let $i = \lfloor \frac{s}{a} \rfloor$. Then only i jobs can finish before t on the first machine in any schedule which implies that $C_{opt} > (n-i)t$. Thus we have $C_j(S) = C_j(S')$ for $1 \leq j \leq i$. For $j > i$, its completion time in S is increased by at most the length of the interval compared with that in S', thus, $C_j(S) \leq C_j(S') + (t - s)$. So we have

$$\begin{aligned}\sum_{j=1}^{n} C_j(S) &\leq \left(\sum_{j=1}^{n} C_j(S')\right) + (n-i)(t-s) \\ &\leq \tfrac{4}{3} C'_{opt} + C_{opt} \\ &\leq \tfrac{4}{3} C_{opt} + C_{opt} \\ &\leq \tfrac{7}{3} C_{opt},\end{aligned}$$

and this completes the proof. ■

III. Mathematical Formulation

Our problem can can be formulated as follows: $\min \sum_{j=1}^{n} C_{b,j}$ subject to

$$C_{a,j} \geq a_j \quad \forall j \tag{1}$$

$$C_{b,j} \geq C_{a,j} + b_j \quad \forall j \tag{2}$$

$$C_{a,i} \geq C_{a,j} + a_i \bigvee C_{a,j} \geq C_{a,i} + a_j \quad \forall i, j,\ i \neq j \tag{3}$$

$$C_{b,i} \geq C_{b,j} + b_i \bigvee C_{b,j} \geq C_{b,i} + b_j \quad \forall i, j,\ i \neq j \tag{4}$$

$$C_{a,j} \leq s \bigvee C_{a,j} > t \quad \forall j \tag{5}$$

The schedule is a permutation schedule. (6)

The first four constraints are the same as the classical model. Constraint (1) says that the processing of any job on first machine cannot start before time 0; and the constraint (2) specifies the second operation of any job can't start before the first operation finishes; and the constraints (3) and (4) show that a machine can not process more than one job at a time. The constraint (5) is unique to our model, it says that the completion time of the first operation is either before or after the breakdown. An additional redundant constraint is (7) or (8), which will be used when we develop lower bounds.

$$C_{b,j} \geq \left(\min_{1 \leq k \leq n} a_k\right) + b_j \quad \forall j \tag{7}$$

$$C_{b,j} \geq \left(\min_{1 \leq k \leq n} a_k\right) + (t - s) + b_j \quad \text{if } \min_{1 \leq k \leq n} a_k > s \ \ \forall j \tag{8}$$

A. Lower Bounds

Let C_{opt} denote the minimum total completion time of the jobs. Based on the formulation, we can develop several lower bounds for C_{opt} by relaxing some of the constraints. Quite a few lower bounds were developed in previous literature for the classical flowshop model, see [5] and the references therein. With breakdown, those lower bounds either need to be modified or not work at all. In the following, we will examine those lower bounds.

1) LB1: This lower bound corresponds to the first lower bound of Ignall and Schrage [13]. The first lower bound from [13] is obtained by relaxing the constraint (4) so that the second machine can simultaneously process as many jobs as needed. Thus, the jobs should be scheduled in SPT_a order on the first machine, and b_j-s can be scheduled immediately without any delay after a_j finishes. With breakdown, the same idea still works, except that we have to count the number of a_j-s that are scheduled after the breakdown.

Suppose $a_1 \leq a_2 \leq \cdots \leq a_n$ and k is the smallest integer such that $\sum_{j=1}^{k+1} a_k > s$ where s is the start time of the unavailable interval. It is easy to see that there are at least $(n - k)$ jobs finish after the breakdown on the first machine in any schedule. Thus

$$LB1 = \left(\sum_{j=1}^{n} \sum_{i=1}^{j} a_i\right) + (n-k)(t-s) + \sum_{j=1}^{n} b_j.$$

2) LB2: This corresponds to the second lower bound of Ignall and Schrage [13], which is obtained by relaxing the constraints (1) and (3). Thus, the jobs should be scheduled in SPT_b order subject to constraint (7). Again, with breakdown, we need to do a modification. If $\min\{a_i\} \leq s$, then the lower bound is not affected by the breakdown; otherwise, constraint (8) should be satisfied. Assume $b_{i_1} \leq b_{i_2} \leq \cdots \leq b_{i_n}$, then we have that

$$LB2 = \begin{cases} n \min\{a_i\} + \sum_{j=1}^{n} \sum_{p=1}^{j} b_{i_p}, & \text{if } \min a_i \leq s \\ n(\min\{a_i\} + (t-s)) + \sum_{j=1}^{n} \sum_{p=1}^{j} b_{i_p}, & \text{otherwise.} \end{cases}$$

3) LB3: this corresponds to the lower bound in [23], which is obtained by applying Lagrangian relaxation to constraint (2) to the classical model. The bound from [23] is based on the assumption that $C_{a,[j]} = \sum_{i=1}^{j} a[i]$ which is true for classical model but not true with breakdown. However, we can still use it to get lower bounds for partial schedules whose completion time on the first machine passes the unavailable interval. For the details on how to calculate/estimate the lower bound, see [23].

From the analysis, we can see LB4 givens a stronger bound than both LB1 and LB2.

4) LB4: This corresponds to LB_{DNT2} in [5] which dominates the first two lower bounds in classical model. We can adapt this bound for our model.

First note that with breakdown, it is still the case that for any schedule S and the j-th job in S, $C_{a,[j]}(SPT_a) \leq C_{a,[j]}(S)$,

where SPT_a represents the schedule generated by SPT_a rule. If we relax the constraint (3), then the problem is equivalent to a single machine problem with each job j has a release time a_j or $a_j + (t-s)$ if $a_j > s$, and processing time b_j. A lower bound of the new problem is given by schedule the jobs using $SRPT$ (Shortest Remaining Processing Time First). As with the classical model, it is still true that $C_{b,[j]}(SRPT) \leq C_{b,[j]}(S)$ for S and j.

Use similar argument as in [5], we can show that for any schedule S,

$$\sum_{j=1}^{n} C_j(S) \geq \sum_{j=1}^{n} \max\left(C_{a,[j]}(SPT_a) + b_{[j]}(S), C_{b,j}(SRPT)\right) = \left(\sum_{j=1}^{n} C_{b,j}(SRPT)\right) + \Delta,$$

where

$$\Delta = \sum_{j=1}^{n} \max(0, b_j(S) - (C_{b,[j]}(SRPT) - C_{a,[j]}(SPT_a))),$$

and Δ is minimized by sorting both b_j-s of S and $(C_{b,[j]}(SRPT) - C_{a,[j]}(SPT_a))$ in non-decreasing order. By the analysis, it is easy to see that LB4 generates a bound that is better than both LB1 and LB2.

5) LB5-Lower Bound by Linear Programming: If we know that the number of first operations that finish before breakdown in the optimal schedule is K, then we can modify the formulation from [12] as follows. Let w_j be the time that $b_{[j]}$ has to wait after $a_{[j]}$ finishes and $w_1 = 0$. Thus we have

$$C_{b,[j]} = C_{a,[j]} + b_{[j]} + w_j \text{ for each position } j = 1, \ldots, n.$$

Then the problem is to minimize

$$\sum_{j=1}^{n} (C_{a,[j]} + w_j + b_{[j]})$$

subject to

$$a_{[k]} + w_k \geq b_{[k-1]} + w_{k-1} \; \forall k = 2, \ldots, n, \text{ and } k \neq K+1 \quad (9)$$

$$a_{[K+1]} + w_{[K+1]} + (t-s) \geq b_{[K]} + w_K \quad (10)$$

$$\sum_{i=1}^{K} a_{[i]} \leq s \quad (11)$$

Let x_{ij} be the binary variable such that $x_{ij} = 1$ means job i is the j-th job in the schedule and $x_{ij} = 0$ otherwise. Then the problem can be formulated as to

minimize

$$\sum_{j=1}^{n} (n-j+1) \sum_{i=1}^{n} x_{ij} a_i + \sum_{j=1}^{n} w_j + \sum_{j=1}^{n} b_j + (n-K)(t-s)$$

subject to

$$\sum_{j=1}^{n} x_{ij} \geq 1, \forall i = 1, \ldots, n$$

$$\sum_{i=1}^{n} x_{ij} \leq 1, \forall j = 1, \ldots, n$$

$$\sum_{i=1}^{n} x_{ik} a_i + w_k - \left(\sum_{i=1}^{n} x_{ik-1} b_i + w_{k-1}\right) \geq 0, \forall k \neq K+1$$

$$\sum_{i=1}^{n} x_{ik} a_i + w_k + (t-s) - \left(\sum_{i=1}^{n} x_{ik-1} b_i + w_{k-1}\right) \geq 0, \; k = K+1$$

$$\sum_{j=1}^{K} \sum_{i=1}^{n} x_{ij} a_i \leq s$$

$$x_{ij} \in \{0,1\} \text{ for } i = 1, \ldots n; j = 1, \ldots, n$$

$$w_j \geq 0 \text{ for } j = 2, \ldots, n$$

The problem is that we don't know the magic number K. However, we can find its minimum possible value of $K_{\min}$ and its maximum possible value $K_{\max}$ by sorting the first operations in SPT_a and LPT_a (longest processing time first by a_js) respectively. Let $LP(k)$ be the lower bound for our problem with a specific $K = k$. Then the lower bound for our problem is $\min_{k=K_{\min}}^{K_{\max}} LP(k)$.

IV. Branch and Bound Algorithm

Our branch-and-bound procedure builds the search tree in a depth-first-search fashion. It starts with the root node at level 0. Each node at level k of the tree corresponds to an initial partial schedule in which k jobs have been put in the first k positions. For this node, at most $(n-k)$ child nodes will be created, one for each unscheduled job. The size of the search tree can be reduced by applying lower bounding technique and dominance rules at each node.

A. Upper Bound

To get an initial upper bound, we first generate some random schedules. We also generate the neighbors of these random schedules which are obtained by local interchanges. Then, we pick the best among these schedules and the schedules obtained by applying SPT, SPT_a, SPT_b. The upper bound is updated any time a leaf of the search tree results in a better schedule. At each internal node, we also calculate the upper bound using SPT, SPT_a, SPT_b based on the partial schedule, the upper bound is updated if a better schedule is obtained.

B. Lower Bound of a Partial Schedule

The LB1, LB2, LB3, LB4 and LB5 mentioned in previous section assume that no jobs have been scheduled yet. All of them can be adjusted if some jobs have been scheduled. However, it will be too expensive to calculate LB5 at each node. So we only calculate LB5 at the root node, and enhance the lower bound based on the partial schedule using the technique from [9]. All other lower bounds will be calculated at each node. If the maximum lower bound is greater than the current upper bound, the node will be cut from the search tree.

C. Dominance Rules

For the aim of improving the performance of the branch and bound algorithm, dominance rules can be used to reduce the size of the search space.

Given a (partial) schedule S, we use $C_A(S)$, $C_B(S)$ to represent the completion time of the last operation on machine 1 and 2 in S, respectively. Let $TC(S)$ be the total completion time of the jobs in S. The earliest available time for the remaining jobs on the second machine, denoted by $r(S)$, is $\max(C_B(S), C_A(S) + \min_{i\notin S} a_i)$ or $\max(C_B(S), C_A(S) + \min_{i\in S} a_i + (t-s))$ if $s < C_A(S) + \min_{i\notin S} a_i \leq t$.

We have the following dominance rules that can be applied to cut the branches of the search tree.

Lemma 4.1: Dominance rule 1. Given two partial schedule S_1 and S_2 of the same set of jobs. If $TC(S_1) \leq TC(S_2)$ and $C_B(S_1) \leq r(S_2)$, then to find the optimal schedule of all jobs, there is no need to consider the schedules based on S_2.

It is easy to see that the above lemma is true. Given a schedule S, a particular schedule that can be checked to see if it dominates S is the schedule resulting from Johnson's rule. If $C_A(S) \leq s$, we check the schedule generated by Johnson's rule for the same set of jobs. Otherwise, $C_A(S) > t$, we check the schedule that schedules the jobs in the same way as S before t, and schedule the remaining jobs after t using Johnson's rule. In both cases, the schedules being checked are guaranteed to have a makespan not greater than that of S.

Lemma 4.2: Dominance rule 2 ([4]). Given two partial schedules of the same set of jobs, S_1 and S_2, then S_2 can be pruned if both of the following inequalities hold:

$$TC(S_1) \leq TC(S_2)$$

and

$$(C_B(S_1) - C_B(S_2))q \leq TC(S_2) - TC(S_1),$$

where q is the number of unscheduled jobs.

The Lemma is obviously true, actually it holds no matter how many breakdowns on the first machine, as long as there is no breakdown on the second machine. The rule can be applied to a schedule S by checking any adjacent schedules of S. Adjacent schedule can be obtained by swapping any two jobs i and j or by inserting j before or after i, etc.

Lemma 4.3: Dominance rule 3. Given a partial schedule S and an unscheduled job i such that $a_i \leq b_i$. If for any other unscheduled job j , we have $a_i \leq a_j$ and $b_i \leq b_j$, and $C_A(S) + a_i + a_j \leq s$ or $C_A(S) + \min_{k\notin S} a_k > s$ then there exists an optimal schedule that job i is the first among all the unscheduled jobs.

Proof: A similar rule for the classical model without availability constraint is given by Croce et. al ([4]) (Dominance rule D2). One can prove the lemma by adapting the proof from [4]. ■

It should be noted that this dominance rule holds only if $C_A(S) + a_i + a_j \leq s$ or $C_A(S) + \min_{k\notin S} a_k > s$. For example, consider $a_1 = a_2 = a_3 = 2$, $b_1 = 2$, $b_2 = 3$ and $b_3 = 4$. Suppose the unavailable interval is $[5, 11]$. At the beginning, S is empty, the conditions are satisfied for $i = 1$, thus by Lemma 4.3, we know there must exist an optimal schedule starting with job 1. So we don't need to consider the schedules that start with job 2 or 3. On the other hand, if the unavailable interval is $[3, 9]$, neither $C_A(S) + a_i + a_j \leq s$ nor $C_A(S) + \min_{k\notin S} a_k > s$ holds for $i = 1$. So it is not clear that there exists an optimal schedule starting with job 1. It turns out the the optimal schedule starts with the last job 3, not job 1.

Lemma 4.4: Dominance rule 4. Given a partial schedule π and two unscheduled jobs i and j, such that $a_i \leq a_j$, $b_i \geq b_j$. Let $L = s - C_A(\pi)$, and π_1 and π_2 be sequences of unscheduled jobs. Let $S_1 = \pi i \pi_1 j \pi_2$ and $S_2 = \pi j \pi_1 i \pi_2$. Then S_1 dominates S_2 if one of the following cases is true,

(a) $0 < L \leq a_i \leq a_j$, and

$$\begin{aligned}&\max(C_A(\pi) + a_i + (t-s), C_B(\pi)) + b_i \\ &\leq \max(C_A(\pi) + a_j + (t-s), C_B(\pi)) + b_j;\end{aligned}$$

(b) $L \leq 0$ or $L > a_i$, and

$$\begin{aligned}&\max(C_A(\pi) + a_i, C_B(\pi)) + b_i \\ &\leq \max(C_A(\pi) + a_j, C_B(\pi)) + b_j.\end{aligned}$$

Proof: If we could show (1) $C_{b,i}(S_1) + C_{b,j}(S_1) \leq C_{b,i}(S_2) + C_{b,j}(S_2)$, and (2) $C_{b,k}(S_1) \leq C_{b,k}(S_2)$ for any $k \neq i, j$, then both the makespan and the total completion time of S_1 are less than or equal to that of S_2, thus the lemma is true.

We consider case (a) first. Apparently, for any $k \in \pi$, we have $C_{a,k}(S_1) = C_{a,k}(S_2)$ and $C_{b,k}(S_1) = C_{b,k}(S_2)$. For job i in S_1 and job j in S_2, the condition $0 < L \leq a_i \leq a_j$ means that neither a_i not a_j can finish before the unavailable interval given the partial schedule π. So for S_1, we have $C_{a,i}(S_1) = C_A(\pi) + a_i + (t-s)$, $C_{b,i}(S_1) = \max(C_{a,i}(S_1), C_B(\pi)) + b_i$.

Similarly for S_2, we have $C_{a,j}(S_2) = C_A(\pi) + a_j + (t-s)$, $C_{b,j}(S_2) = \max(C_{a,j}(S_2), C_B(\pi)) + b_j$.

Since $a_i \leq a_j$, we have $C_{a,i}(S_1) \leq C_{a,j}(S_2)$. The condition $\max(C_A(\pi) + a_i + (t-s), C_B(\pi)) + b_i \leq \max(C_A(\pi) + a_j + (t-s), C_B(\pi)) + b_j$ implies

$$C_{b,i}(S_1) \leq C_{b,j}(S_2).$$

Consequently, for any $k \in \pi_1$, it must be true that $C_{a,k}(S_1) \leq C_{a,k}(S_2)$ and $C_{b,k}(S_1) \leq C_{b,k}(S_2)$. Therefore $C_B(\pi i \pi_1) \leq C_B(\pi j \pi_1)$.

Clearly, $C_{a,j}(S_1) = C_{a,i}(S_2)$. Since $b_i \geq b_j$, we have $C_{b,j}(S_1) = \max(C_{a,j}(S_1), C_B(\pi i \pi_1) + b_j \leq C_{b,i}(S_2) = \max(C_{a,i}(S_1), C_B(\pi j \pi_1)) + b_i$ Thus, for any $k \in \pi_2$, $C_{a,k}(S_1) = C_{a,k}(S_2)$ and $C_{b,k}(S_1) \leq C_{b,k}(S_2)$.

For case (b), given the partial schedule π, if $L \leq 0$, then all the remaining jobs including i and j have to be scheduled after the unavailable interval. If $L \geq a_i$, then a_i can finish before the unavailable interval, but a_j may or may not finish before the unavailable interval. Either way, the inequality

$$\max(C_A(\pi)+a_i, C_B(\pi))+b_i \leq \max(C_A(\pi)+a_j, C_B(\pi))+b_j$$

guarantees that $C_{b,i}(S_1) \leq C_{b,j}(S_2)$. We can use similar argument as case (a) to prove that $C_{b,k}(S_1) \leq C_{b,k}(S_2)$ for any $k \neq i, j$, thus the lemma is true. ■

D. Experimental Results

The instances are generated as follows. For processing times, we use three distributions, $[1, 10]$, $[1, 50]$, $[1, 100]$. It is known that the distribution $[1, 100]$ generates the most difficult type of problem instances. The $[1, 10]$ distribution is the easiest and also the most practical distribution. The number of jobs n takes on values 10, 15, 20, 25, 30, 35. The start time s of the unavailable intervals are generated from four uniform distributions, $[0, \frac{1}{4}A]$, $[\frac{1}{4}A, \frac{1}{2}A]$, $[\frac{1}{2}A, \frac{3}{4}A]$, $[\frac{3}{4}A, A]$, where $A = \sum_{j=1}^{n} a_j$. The duration of the unavailable intervals, $(t-s)$, takes the values of 1% 20%, 40%, 60%, 80%, 100% of A. For each combination of n and a processing time distribution, we generate 25 sets of n jobs; then for each set of n jobs, we generate 16 instances by combining different distribution of s and $(t - s)$.

We pre-process the jobs in SPT, SPT_a, SPT_b and Johnson's rule and keep this order so that they can be accessed at each node of the tree. We set the termination criteria of the branch and bound algorithm by limiting the number of the nodes explored to be 5000000.

1) Comparison of Lower Bounds At the Root Node: From previous discussion, when $min_{1 \leq i \leq n} a_i < s$, the LB2 bound doesn't put much consideration of the unavailable interval on the first machine, so its is expected to be the worst lower bound. $LB3$ can only be applied for partial schedules passing the unavailable interval. We also know that LB4 dominates LB1 and LB2. LB5 is generated by linear programming.

Experiments show that at the root node, LB5 always gives the best lower bound, and LB4 is very close to LB5. On average, LB1 is not that bad. The performance of LB2 can vary a lot from instance to instance. Table I shows the ratios of the LB5 with LB1, LB2, and LB4 for the instances with $n = 35$. For each distribution of the processing times, we list the minimum, maximum, average ratios and the standard deviation.

We also notice that the gap of LB5 at the root and the optimal solutions is very small. Table II shows the value of LB5 at the root compared with the optimal solution (or the best solution found if optimal not found) for the instances with $n = 35$. We can see on average, LB5 is within 1% of the optimal (best) solution.

2) Lower Bounds at the Internal Node: Although LB5 is very good and close to the optimal solution, but its computation takes a lot of time, thus we cannot afford to compute at other nodes. So we use the method from [12] to enhance LB5 based on the partial schedule at each node. For each instance, we record the number of nodes it explores before it finds the optimal solution or reach the maximum node limit, 5000000. For each of the lower bound LB1, . . ., LB5, We also record the number of nodes such that the lower bound is maximum. We observed that LB4 provides the best lower bound at internal nodes for almost all instances. Fig.1 shows 125 instances of 35 jobs. For each instance, among all the nodes explored, the figure shows the number of times when each of LB3, LB4 and LB5 equals to $\max(LB3, LB4, LB5)$. We can clearly see that $LB4$ provides the best lower bounds at the majority of the nodes in the search tree, and $LB3$ provides good lower bound only for a very small portion of the nodes in the tree.

3) Dominance Rule: To find out the effectiveness of the dominance rules, we run experiment with and without applying dominance rules. For each instance, we compare the number of nodes explored by the algorithm in both cases. We compute the ratio of the number of nodes with dominance rules and the number of nodes without dominance rules. Table III lists minimum, maximum, and average ratio for different processing time distribution when $n = 20$. We can see on average, the number of the nodes with dominance rules is only 10% of the number of nodes with no dominance rules, i.e 10 times speed up of the algorithm.

4) Unavailable Interval: Fig. 2 shows how the starting time of the unavailable interval affects the size of the search tree. In the figure, we show the number of nodes explored for 120 instances of $n = 20$. The instances are divided into 4 groups of 30 instances whose unavailable interval starting times are drawn from the distribution $[0, \frac{1}{4}A]$, $[\frac{1}{4}A, \frac{1}{2}A]$, $[\frac{1}{2}A, \frac{3}{4}A]$, $[\frac{3}{4}A, A]$, respectively. The figure suggests that the search tree size is getting bigger as the unavailable interval starts later. We can also see that the LB3 plays a bigger role when the unavailable intervals starts before $\frac{1}{4}A$.

Fig. 3 shows how the length of the unavailable interval affects the size of the search tree. The effect is not that obvious.

5) Problems Size Solved: Our algorithm is implemented in C++, and Gurobi is used for linear programming. Within the maximum number of nodes 5000000, we were able to solve the instances of 30 jobs. When $n = 35$, we can solve the majority of the instance if the processing time distribution is from $[1, 10]$. For the other two distributions, $[1, 50]$ and $[1, 100]$, we can only solve some of the instances. For example, for the 125 instances we used from Table I and Table II, the number of solved instances are 88, 28, 19, respectively. On the other hand, from the Table II, we know that the best solutions found by the branch and bound algorithm are actually very close to the optimal solution.

V. Conclusion

In this paper, we studied the problem of minimizing total completion time subject to the constraint that there is a

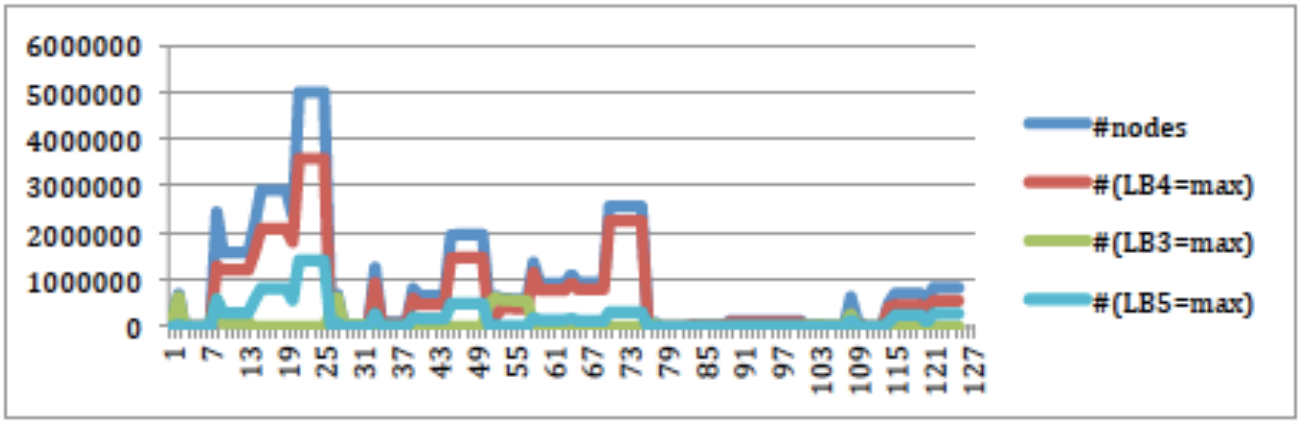


Fig. 1. Comparison of the lower Bounds at Internal Nodes

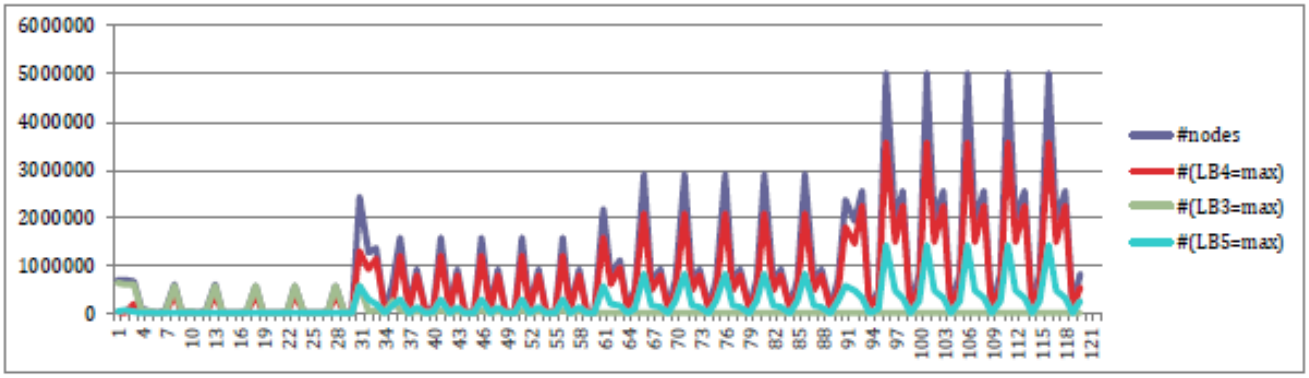


Fig. 2. Number of nodes explored by starting time of the unavailable interval, each of $[0, \frac{1}{4}A]$, $[\frac{1}{4}A, \frac{1}{2}A]$, $[\frac{1}{2}A, \frac{3}{4}A]$, $[\frac{3}{4}A, A]$ has 30 instances

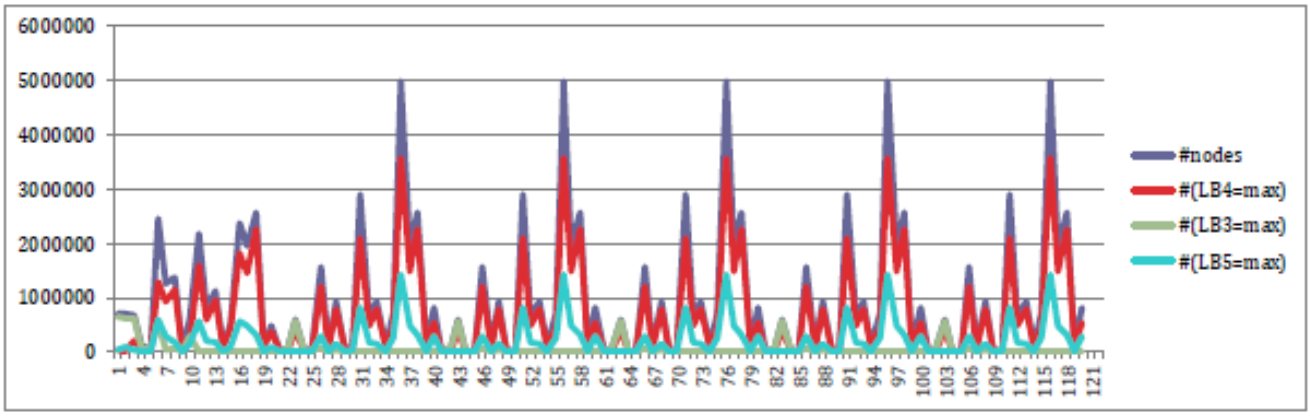


Fig. 3. Number of nodes explored by length of the unavailable interval, each of $1\%A$, $20\%A$, $40\%A$,$60\%A$,$80\%A$ and $100\%A$ has 20 instances

TABLE I
LOWER BOUNDS COMPARISON AT THE ROOT NODE

	[1,10] distribution			[1,50] distribution			[1,100] distribution		
	LB5/LB1	LB5 /LB2	LB5 / LB4	LB5/LB1	LB5 /LB2	LB5 / LB4	LB5/LB1	LB5 /LB2	LB5 / LB4
min	1.018	1.089	1.018	1.005	1.072	1.005	1.005	1.054	1.005
max	1.174	3.410	1.119	1.301	3.454	1.151	1.253	3.351	1.160
average	1.080	1.601	1.069	1.102	1.692	1.067	1.089	1.598	1.064
standard deviation	0.042	0.485	0.027	0.084	0.552	0.038	0.069	0.486	0.043

TABLE II
LB5 AT THE ROOT COMPARED WITH OPTIMAL SOLUTION/BEST SOLUTION FOUND

	[1,10]	[1,50]	[1,100]
min	0.001	0.002	0.002
max	0.013	0.046	0.028
average	0.001	0.014	0.012
standard deviation	0.003	0.009	0.007

TABLE III
THE RATIO OF NUMBERS OF NODES IN SEARCH TREES USING DOMINANCE RULES WITH THAT NOT USING DOMINANCE RULES

	[1,10]	[1,50]	[1,100]
min	0.002	0.011	0.004
max	0.149	0.714	0.244
average	0.027	0.117	0.069

single unavailable interval on the first machine. We show that some special cases can be solved optimally or within a constant factor; and for the general case, SPT rule gives an n-approximation. We performed computational experiments to investigate the effectiveness of different lower bounds and the dominance rules. It is observed that at the root node, LB5 provides the best lower bound with the expense of time, LB4 can provide almost good lower bounds but more efficiently at the root node and gives the best lower bounds at most internal nodes. The dominance rules speeds up the algorithm dramatically. We also observed that the earlier the the unavailable interval starts, the more efficient the algorithm is; on the other hand, the duration of the unavailable interval does not affect the efficiency of the branch and bound algorithm very much.

REFERENCES

[1] Akkan, C. and S. Karabati, The two-machine flowshop total completion time problem: Improved lower bounds and a branch-and-bound algorithm, European Journal of Operational Research, 159, 420-429, 2004, http://dx.doi.org/10.1016/S0377-2217(03)00415-6.

[2] Cadambi, B. W. and Y. S. Sathe, Two-machine flowshop scheduling to minimise mean flow time, Opsearch, 30, 35-41,1993.

[3] Conway, R. W., W. L. Maxwell, and L. W. Miller, Theory of Scheduling. Addison-Wesley, Reading, MA, 1967.

[4] Della Croce, F., M. Ghirardi, and R. Tadei, An improved branch-and-bound algorithm for the two machine total completion time flow shop problem, European Journal of Operational Research, 139, 293-301, 2002, http://dx.doi.org/10.1016/S0377-2217(01)00374-5.

[5] Della Croce, F., V. Narayan, and R. Tadei, The two-machine total completion time flow shop problem, European Journal of Operational Research, 90, 227-237, 1996, http://dx.doi.org/10.1016/0377-2217(95)00351-7.

[6] Garey, M. R., D. S. Johnson, and R. Sethi, The complexity of flowshop and jobshop scheduling, Mathematics of Operations Research, 13, 330-348, 1976, http://dx.doi.org/10.1287/moor.1.2.117.

[7] T. Gonzalez and S. Sahni, Flowshop and jobshop schedules: Complexity and approximation. Operations Research 26, 36-52, 1978,http://dx.doi.org/10.1287/opre.26.1.36.

[8] J.N.D. Gupta and R.A.Dudek, Optimality criteria for flowshop schedules, *AIIE Trans*, 3, 199-205, 1971, http://dx.doi.org/10.1080/05695557108974807.

[9] H. Hoogeveen and T. Kawaguchi, Minimizing total completion time in a two-machine flowshop: Analysis of special cases, Mathematics of Operations Research, Vol. 24 Issue 4, 887-910, 1999, http://dx.doi.org/10.1287/moor.24.4.887.

[10] Hoogeveen, J. A. and S. L. Van de Velde, Stronger Lagrangian bounds by use of slack variables: applications to machine scheduling problems, Mathematical Programming, 70, 173-190, 1995, http://dx.doi.org/10.1007/BF01585935.

[11] Hoogeveen, J. A. and S. L. Van de Velde, Scheduling by positional completion times: Analysis of a two-stage flow shop problem with a batching machine, Mathematical Programming, 82, 273-289, 1998, http://dx.doi.org/10.1007/BF01585876.

[12] Hoogeveen, J. A., L. van Norden and S. L. Van de Velde, Lower bounds for minimizing total completion time in a two-machine flow shop, Journal of Scheduling, 9: 559-568, 2006, http://dx.doi.org/10.1007/s10951-006-8789-x.

[13] Ignall, E., and L. E. Scharge, Application of the branch and bound technique to some flow-shop problems, Operations Research, 13, 400-412, 1965, http://dx.doi.org/10.1287/opre.13.3.400.

[14] Kohler, W. H. and K. Steiglitz, Exact, approximate and guaranteed accuracy algorithms for the flowshop problem n/2/F/F, Journal ACM, 22, 106 -114, 1975, http://dx.doi.org/10.1145/321864.321872.

[15] W. Kubiak, J.Blazewicz, P. Formanowicz, J. Breit and G. Schmidt, Two-machine flow shops with limited machine availability, European Journal of Operational Research Volume 136, Issue 3, pp. 528-540, 2002, http://dx.doi.org/10.1016/S0377-2217(01)00083-2.

[16] C.-Y. Lee, Machine scheduling with an availability constraints, Journal of Global Optimization, 9, pp. 363-382, 1996, http://dx.doi.org/10.1007/BF00121681.

[17] C.-Y. Lee, Minimizing the makespan in the two-machine flowshop scheduling problem with an availability constraint, Operations Research Letters, 20, pp. 129-139, 1997, http://dx.doi.org/10.1016/S0167-6377(96)00041-7

[18] C. Y. Lee, L. Lei and M. Pinedo, Current trends in deterministic scheduling, Annals of Operations Research,70(0) : 1-41, 1997.

[19] C. Y. Lee, Two-machine flowshop scheduling with availability constraints, European Journal of Operational Research, 114, pp. 420-429, 1999, http://dx.doi.org/10.1016/S0377-2217(97)00452-9

[20] Ma, Y., C. Chu and C. Zuo, A survey of scheduling with deterministic machine availability constraints, *Computers & Industrial Engineering*, 58(2), pp. 199-211, 2010, http://dx.doi.org/10.1016/j.cie.2009.04.014

[21] E. Sanlaville and G. Schmidt, Machine scheduling with availability constraints, Acta Informatica, 35, pp. 795-811, 1998.

[22] G. Schmidt. Scheduling with limited machine availability, *European Journal of Operational Research*, 121(1), pp. 1-15, 2000, http://dx.doi.org/10.1016/S0377-2217(98)00367-1

[23] Van de Velde, S. L, Minimizing the sum of the job completion times in the two-machine flow shop by Lagrangian relaxation, Annals of Operations Research, 26, 257 - 268, 1990.

Computer Science & Systems

CSS is a FedCSIS conference area aiming at integrating and creating synergy between FedCSIS events that thematically subscribe to more technical aspects of computer science and related disciplines. The CSNS area spans themes ranging from hardware issues close to the discipline of computer engineering via software issues tackled by the theory and applications of computer science and to communications issues of interest to distributed and network systems. Events that constitute CSNS are:

- AIPC'16—1st International Workshop on Advances in Image Processing and Colorization
- CANA'16—9th Computer Aspects of Numerical Algorithms
- CPORA'16—1st Workshop on Constraint Programming and Operation Research Applications
- IWCPS'16—3rd International Workshop on Cyber-Physical Systems
- MMAP'16—9th International Symposium on Multimedia Applications and Processing
- WSC'16—8th Workshop on Scalable Computing

1st International Workshop on Advances in Image Processing and Colorization

IMAGE Processing and Colorization are two fields which have shown a stable growth in the last decades. They have different impact in computer science and other research areas, e.g. biology, remote sensing, and medicine. Image processing is very important in medical applications, robotics, and other industrial applications. Recently, the study of "colorization" begins to attract attention. The colorization technique can colorize a monochrome image by giving a number of color pixels. Colorization is a computerized process of adding color to a black and white print, movie and TV program. Image Colorization has a great impact on different applications. Image processing and Colorization can also be combined in many applications. Different tools, from machine learning and mathematics, are greatly influencing the research areas in image processing and Colorization. Recently, bio-inspired optimization techniques have been used to improve the image processing techniques, segmentation and classification.

The purpose of the "Advances in Image Processing and Colorization" session is to bring scientists, researcher scholars, and students from academia and practitioners present ongoing research activities about both theoretical advances and applications of Image Processing and Colorization and to allow the exchange of new ideas and application experiences to face to face.

TOPICS

Topics of interest include but are not limited to:

- Bio-inspired based image classification,
- Bio-inspired based image segmentation,
- Thermal image processing,
- Biometric identification and authentication,
- Face and iris recognition,
- Bio-inspired based face and iris recognition,
- New colorization approach,
- Colorization image coding techniques,
- Color embedding,
- Applications of image colorization,
- Video colorization,
- Physics-based colorization,
- Interactive colorization method,
- Colorization using optimization techniques,
- Colorization for monochrome Image,
- Color to gray and back.

EVENT CHAIRS

- **Gaber, Tarek,** Faculty of Computers & Informatics, Suez Canal University, Ismailia, Egypt, and Scientific Research Group in Egypt, Egypt
- **Horiuchi, Takahiko,** Graduate School of Advanced Integration Science, Chiba University, Japan
- **Ibrahim, Abdelhameed,** Faculty of Engineering, Mansoura University, Mansoura, Egypt, and Scientific Research Group in Egypt, Egypt
- **Kromer, Pavel,** Dep. CS, VSB-Technical University of Ostrava, Czech Republic

PROGRAM COMMITTEE

- **Abed, Fidaa,** TU Muncih, Germany
- **Afifi, Ahmed,** Faculty of Computers and Information, Menofia University, Egypt
- **Celebi, M. Emre,** Dept. of Computer Science, Louisiana State University,, USA
- **Conci, Aura,** Computer Sc. Dep., Universidade Federal Fluminense - UFF, Brazil
- **Dey, Nilanjan,** Techno India College of Technology, India
- **Faria, Lincoln,** Department of Computer Science, Fluminense Federal University, Brazil
- **Ghoneim, Mohamed Elsayed,** Umm Al Qura University, Saudi Arabia
- **Hassanien, Aboul Ella,** Faculty of Computers and Information, Cairo University, Egypt
- **Hirai, Keita,** Graduate School of Advanced Integration Science, Chiba University, Japan
- **Khan, Zeeshan,** Qassim Private Colleges, Saudi Arabia
- **Kotera, Hiroaki,** Kotera Imaging Laboratory, Japan
- **Kotyk, Taras,** Ivano-Frankivsk National Medical University, Ukraine
- **Mosa, Abdelkhalik,** The university of Manchester, United Kingdom
- **Platos, Jan,** VŠB-Technical University of Ostrava, Czech Republic
- **Semary, Noura,** Faculty of Computers and Information, Menofia Univ, SRGE Member, Egypt
- **Singh, Aarti,** Maharishi Markandeshwar University
- **Smolka, Bogdan,** Silesian University of Technology, Poland
- **Snasel, Vaclav,** VSB -Technical University of Ostrava, Czech Republic
- **Tahoun, Mohamed,** PhD, Computer Science Department, Faculty Of Computers and Informatics, Suez Canal University, Egypt
- **Tharwat, Alaa,** Suez Canal University - Egypt, Egypt
- **Tsai, Pei-Wei,** Fujian University of Technology, China
- **Zubair, Muhammad,** Qassim Private Colleges, Saudi Arabia

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 641–644
DOI: 10.15439/2016F521
ACSIS, Vol. 8. ISSN 2300-5963

An Image Steganography Algorithm using Haar Discrete Wavelet Transform with Advanced Encryption System

Essam H. Houssein [1,*],Mona A. S. Ali [2,*], and Aboul Ella Hassanien[3,*]
[*] Faculty of Computers and Information
[1]Minia University [2]Benha University [3]Cairo University
[*]Scientific Research Group in Egypt (SRGE) http://www.egyptscience.net

***Abstract*—The security of data over the internet is a crucial thing specially if this data is personal or confidential. The transmitted data can be intercepted during its journey from device to another. For that reason, we are willing to develop a simple method to secure data. Data encryption is one method to secure the messages but the intruders can still try to crack it, in order to overcome this, steganography has been used to hide the data into a cover media(i.e. audio, image or video). Recently steganography attracts many researchers as a hot topic. This paper proposes an advanced technique for encrypting data using Advanced Encryption System (AES) and hiding the data using Haar Discreet Wavelet Transform (HDWT). HDWT aims to decrease the complexity in image steganology while providing less image distortion and lesser detectability. One-forth of the image carrying the details of the image in a region and other three regions carrying a less details of the image then the cipher text is concealed at most two Least Significant Bits (LSB) positions in the less detailed regions of the carrier image, if the message doesn't fit in the first LSB only it will use the second LSB. This proposed algorithm covers almost all type of symbols and alphabets.**

***Index Terms*—Steganography, Cryptography, Encryption, AES Encryption, LSB, DWT**

I. Introduction

STEGANOGRAPHY is a data hiding technique that has been mainly used in information security applications. It is similar to watermarking and cryptography techniques, but these three techniques are different in some aspects. Firstly, watermarking mainly tracks illegal copies or claims of the ownership of digital media. It is not geared for communication. Secondly, cryptography scrambles the data with the mixture of permutation(s) and substitution(s) so that unintended receivers cannot perceive the processed information. However, the fact that information has been embedded into a medium (i.e., watermarking) and communication has been carried out (i.e., cryptography) is known to everyone, or at least it is acceptable to reveal such a fact. Finally, steganography transmits information by embedding messages into innocuous looking cover objects, such as digital images, to conceal the very existence of communication. As a result, steganography is the art and science of data smuggling since its goal is to hide the presence of communication [1].

Each image hiding system consists of an embedding process and an extraction process. An innocuous-looking original image is used as the cover-image to conceal the secret data. The secret data are embedded into the cover-image by modifying the cover-image to form a stego-image. Cryptography [2], [3] and steganography [4] are the two important aspects of communications security. Although cryptography is a primary method of protecting valuable information by rendering the message unintelligible to outsiders [3], steganography is a step ahead by making the communication invisible. A possible formula of the process may be represented as: $Stegomedium = Covermedium + Embeddedmessage + Stegokey$

In this paper, an advanced technique for encrypting data is proposed using AES and hiding the data using Haar DWT technique, carrying the details of the image in a region and other three regions carrying a less details of the image then the cipher text is concealed at most two LSB positions. Extensive experiments show the effectiveness of the proposed method. The results obtained also show significant improvement than the method proposed in [5]. The remainder of the paper is organized as follows. Section II briefly describes the related work. In Section III, the related main knowledge is described. Section IV, discusses the features of the proposed technique. The performance is analyzed. Experimental results are given in Section V. Finally, Section VI concludes this paper.

II. Related Work

A few steganography approaches are briefly reviewed here.In [6], Tiegang et al. proposed a new image encryption algorithm based on hyper-chaos, which uses a new image total shuffling matrix to shuffle the pixel positions of the plain-image and then the states combination of hyper-chaos is used to change the grey values of the shuffled-image. In [7], Chin-Chen et al. proposed a new steganographic method to increase the message load in every block of the stego-image while keeping the stego-image quality acceptable. In [8], B. T. NilanjanDey, et al, proposed a method for hiding multiple images in an image based on DWT and DCT. In [9], Chen Po-Yueh et al. proposed a new steganography technique which embeds the secret messages using the DWT in the frequency domain to divide the image into 4 sub bands, and it will embed the secret data in the LSB of the lowest priority band and it wouldn't use the low frequency sub band that holds the most

details to preserve the quality of the image, and it have 2 modes and 5 cases because of the demands of the capacity or quality.

In [10], Chiang-Lung et al. proposed method adopts the complementary embedding strategy to reduce the loss of statistical property of the stego-image in spatial domain. In [11], Chi-Kwong Chan et al. proposed a data hiding method by simple LSB substitution with an optimal pixel adjustment process. The image quality of the stego-image can be greatly improved with low extra computational complexity. In [12], KokSheik et al. presented a novel Mod4 steganographic method in discrete cosine transform (DCT) domain. Mod4 is a blind steganographic method.

In [13], M. Juneja et al., proposed a secure methods of information security using steganography, the first method embeds In the LSB of the blue components partial green components of random positions in the edges of the green component, the second method is an adaptive method that uses the data of the MSB of the red, green and blue components to embed the message in the random pixels across smooth areas, the third method is a hybrid feature detection filter that predicts the edges in noisy conditions.

In [5], Avval et al, proposed steganography technique to embed audio into the edges of a color image, this technique uses the chaotic map to select the random edge pixels to embed the bits and to choose which random LSB bit location in the selected pixel. In [14], Hemalatha S. et al. proposed a method to hide multiple secret keys and images in a color image using integer wavelet transform (IWT). In [11], T. Garima, proposed an approach to encrypt message using RSA 1024 algorithm then it will be embedded in a cover image by modifying the LSB technique. In [15], L. S. Ahmed et al., proposed a method of encrypting the message by a substitution cipher then it will be embedded using LSB insertion techniques to achieve high capacity to be embedded and security of the steganography method. In [10], Liu Lung et al. proposed a jpeg steganographic using complementary embedding technique, this method is achieved by dividing the quantized DCT coefficients and the secret bits into two parts according to a predefined partition ratio. The two parts of DCT coefficients are used to embed the corresponding parts of secret bits with different embedding algorithms. Specifically, the secret bits are embedded by subtracting one from one part of coefficients, and adding one to the other part of coefficients.

III. The steganography

A. 2D-Haar-wavlet Transform

Wavelet transform has the capability to offer some information on frequency-time domain concurrently. In this transform, time domain is passed through high-pass and low-pass filters to extract high and low frequencies respectively. This process is repeated for a number of times and each time a section of the signal is drawn out. DWT analysis splits signal into two classes (i.e. Approximation and Detail) by signal decomposition for different frequency bands and scales. DWT employs two function sets: scaling and wavelet which associate with low and high pass filters orderly. Decomposition follows the manner of dividing time separability. Meanly, only half of the samples in a signal are sufficient to represent the whole signal, doubling the frequency separability.

Haar wavelet operates on data by calculating the sums and differences of adjacent elements. This wavelet operates first on adjacent horizontal elements and then on adjacent vertical elements. One important feature of the Haar wavelet transform is that the transform is equal to its inverse. Each transform computes the data energy in relocated to the top left hand corner.

Figure 1 shows the image Lena after one Haar wavelet transform

Fig. 1: 2D Haar Wavelet Transform Example

After each transform is performed the size of the square which contains the most important information is reduced by a factor of 4 as seen in Figure 2.

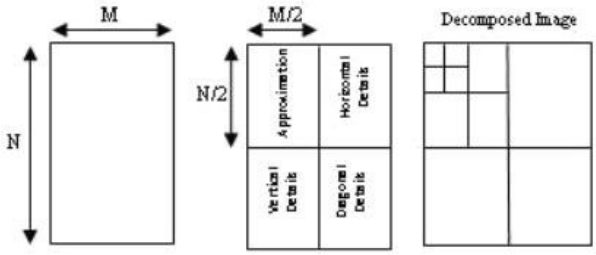


Fig. 2: Detailed 2D Haar Wavelet Transform

B. AES encryption technique

The cipher takes a plaintext block size of 128 bits, or 16 bytes. The key length can be 16, 24, or 32 bytes (128, 192, or 256 bits). Data block of 4 columns of 4 bytes is state key expanded to array of words. Ordering of bytes within a matrix is by column. The cipher consists of rounds, where the number of rounds depends on the key length: 10 rounds for a 16-byte key, 12 rounds for a 24-byte key, and so on, each transformation takes one or more 4 X 4 matrices as input and produces a 4 X 4 matrix as output (the cipher text).The key expansion function generates N + 1 round keys, each of which is a distinct 4 X 4 matrix. Each round key serves as one of the inputs to the Add Round Key transformation in each round.

C. LSB substitution technique

The wide used technique of hiding messages into the image without affecting the whole image is the LSB technique that uses the least significant bit of each pixel to embed the message into it. Figure 3 shows how the pixel can be illustrated from the most important bit to the least important bit. The basic concept of LSB substitution is to embed the message in the least important bit (least significant bit) of the pixel the equation for the LSB is

$$X_i^{'} = X_i - X_i mod 2^k + m_i \tag{1}$$

in this equation the $X_i^{'}$ is the i^{th} pixel value of the stego image, X_i is the i^{th} pixel of the image (cover image) and m_i denotes the decimal value of the i^{th} block in confidential data. K denotes the number of LSB places. we will use or substitute from the pixel. In the process of extracting the message from the image pixel is to copy the least significant bit directly, and this process is showed in the following equation (2):

$$m_i = x_i^{'} mod 2^k \tag{2}$$

This technique is easy and fast but it have drawback if the message have a great size, it will affect the image and it can be noticeable.

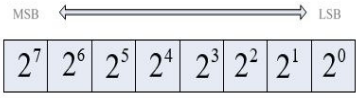


Fig. 3: The places of the bits in the pixel

IV. Proposed Approach

In this research the proposed approach includes two main process: Embedding process and Extracting process.

A. Embedding process

As seen in Algorithm 1, we started first by reading the cover image C and the message to be embedded M. The cipher message S is extracted by applying the AES encryption technique on message M. After performing the AES technique on the message, it is ready to be concealed in the image. Suppose that the 8-bit gray level cover image of size M_CxN_C as :

$$C = \{x_{i,j} | 1 \leqslant i \leqslant M_c, 1 \leqslant j \leqslant N_c, x_{i,j} \in \{1, 2, 3..., 255\}\} \tag{3}$$

S is the n-bit secret message represented as:

$$S = \{S_i | 1 \leqslant i \leqslant n, S_i \in \{0, 1\}\} \tag{4}$$

From Algorithm 1, the embedding steps will be as follows:

1) Apply the DWT on the cover image and get the four sub-bands obtained are denoted LL, HL, LH and HH.
2) Get three bits iteratively from the cipher message and distribute only one bit from the three bits in every one LSB of the three sub bands HH, HL and LH respectively, if the cipher message require more LSB, our algorithm will move cursor to the second LSB of each pixel in the three sub bands. This improves the capacity and permits high capacity with no effect on the pixel value. As the embedding process is done on the sub bands that don't contain details of the image.
3) Perform the inverse DWT on the result of step 2, by performing this step the resulted matrix is H, some pixels of H are not integers ranging from 1-255 due to LSB substitution. So we obtain (K-matrix) that contains the possible non-integer situation (0.0, 0.25, 0.5 and 0.75).
4) Round the matrix H to obtain the stego image E, in order to reconstruct the secret message we will use the k-matrix for reconstruction.
5) Send the Stego image to the receiver with the k-matrix in the description file or tag with the total number of message bits too.

Algorithm 1 Proposed Approach

1: input=gray scale cover image C and message to be embedded M
2: output= StegoImage
3: $S =$ Encrypt M using AES technique
4: $IfSize(S) >$ total first three bit of LSB in C
5: Then get another cover Image AC and go to step 4
6: D = Apply the DWT on C
7: Embed S in the coefficient of D
8: Inverse D
9: obtain (K-matrix) that contains the possible non-integer situation (0.0, 0.25, 0.5 and 0.75)
10: Calculate inverse 2D-DWT on each block to get the stego image.

B. Extraction process

In order to extract the original image the following steps are followed:

1) Extract the k-matrix of the file tag of E.

$$K = \{K_{i,j} | 1 \leqslant M_F \leqslant n, 1 \leqslant j \leqslant N_F, K_{i,j} \in \{00, 01, 10, 11\}\} \tag{5}$$

Transform all elements of k into (0.0, 0.25, 0.5 and 0.75).
2) Obtain the H by performing DWT which is calculated as H=E+K-matrix.
3) Obtain the total number of the message from the file tag of E, extract 3 bits iteratively from the LSB of the three sub bands HHH, HHL and HLH. After completing the first LSB and there are more bits of the message is still not extracted (bits extracted message size of E != 0) Get the remaining bits from the second LSB in the same way as the last step. Extracting the two LSB of the remaining bits, example [A,B]
4) After extracting the whole bits perform the inverse AES on the encrypted bits to obtain the original message.

V. Results and Discussion

The proposed approach is applied on 512x512 8-bit grayscale images Jet, Boat, Baboon and Lena. The messages

are generated randomly with size upto maximum hiding capacity. To measure the quality, a parameter is developed to compute the quality of the image this parameter is called PSNR and defined as follows:

$$PSNR = 10\log_{10}(255^2/MSE) \tag{6}$$

The root mean square error (RMSE) has been used as a standard statistical metric to measure model performance in meteorology, air quality, and climate research studies. The mean absolute error (MAE) is another useful measure widely used in model evaluations, denotes the mean error and it's the deference between the original image and the stego image as seen in Figure 5, and the equations to compute the MAE and MSE are:

$$MSE = \frac{1}{MXN}\sum(a-b)^2 \tag{7}$$

Where a denotes the pixel in the original image and b denotes the pixel in the stego image, with high PSNR means a high quality, this means that while the dB is low this means that the image has been modified or there is a noise or distortion.

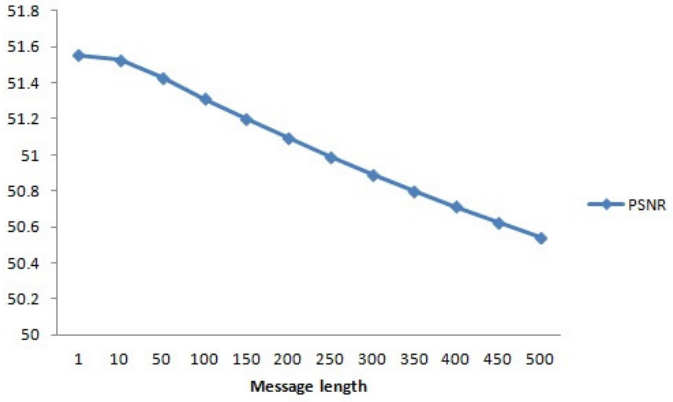


Fig. 4: PSNR Results

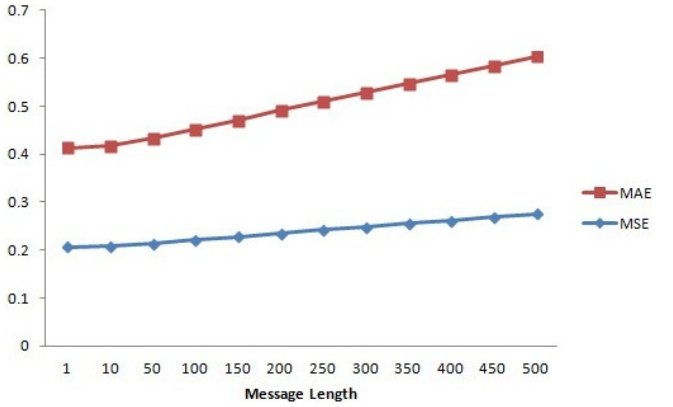


Fig. 5: MSE and MAE Results

Verification of the results has been done for various length of message (LM) and calculates error for all different messages the proposed technique works fine show in Figure 5. When compare our technique with scheme [5]. It gets results better than other method. Result of PSNR for other method between 75, 80 but PSNR of our method smaller than it as seen in Figure 4.

VI. Conclusion

There are demands of the algorithms of steganography to progress with the capacity or with quality, so with this method we aimed at the capacity demand to store as high as possible messages in the image with reduced effect of the quality of the image, and made it even harder to get the message by encrypting it before storing the message in the cover image, with this method the message is secured and hidden, and the cover image can take much more message size to hide. As the main goal of steganography is to hide and secure the message.

References

[1] S. Katzenbeisser and F. Petitcolas, *Information hiding techniques for steganography and digital watermarking*. Artech house, 2000.
[2] A. Nissar and A. Mir, "Classification of steganalysis techniques: A study," *Digital Signal Processing*, vol. 20, no. 6, pp. 1758–1770, 2010.
[3] S. Williams, "Cryptography and network security: Principles and practices," 2006.
[4] M. M. Amin, M. Salleh, S. Ibrahim, M. R. Katmin, and M. Shamsuddin, "Information hiding using steganography," in *Telecommunication Technology, 2003. NCTT 2003 Proceedings. 4th National Conference on*. IEEE, 2003, pp. 21–25.
[5] N. Bhardwaj and S. Agarwal, "A new technique for extracting image information beyond visibility," *International Journal of Information and Computation Technology*, vol. 3, pp. 539–548, 2013.
[6] T. Gao and Z. Chen, "A new image encryption algorithm based on hyper-chaos," *Physics Letters A*, vol. 372, no. 4, pp. 394–400, 2008.
[7] C.-C. Chang, T.-S. Chen, and L.-Z. Chung, "A steganographic method based upon jpeg and quantization table modification," *Information Sciences*, vol. 141, no. 1, pp. 123–138, 2002.
[8] T. Bhattacharya, N. Dey, and S. Chaudhuri, "A session based multiple image hiding technique using dwt and dct," *arXiv preprint arXiv:1208.0950*, 2012.
[9] P.-Y. Chen, H.-J. Lin *et al.*, "A dwt based approach for image steganography," *International Journal of Applied Science and Engineering*, vol. 4, no. 3, pp. 275–290, 2006.
[10] C.-L. Liu and S.-R. Liao, "High-performance jpeg steganography using complementary embedding strategy," *Pattern Recognition*, vol. 41, no. 9, pp. 2945–2955, 2008.
[11] C.-K. Chan and L.-M. Cheng, "Hiding data in images by simple lsb substitution," *Pattern recognition*, vol. 37, no. 3, pp. 469–474, 2004.
[12] K. Wong, X. Qi, and K. Tanaka, "A dct-based mod4 steganographic method," *Signal Processing*, vol. 87, no. 6, pp. 1251–1263, 2007.
[13] M. Juneja and P. S. Sandhu, "A new approach for information security using an improved steganography technique," *Journal of Information Processing Systems*, vol. 9, no. 3, pp. 405–424, 2013.
[14] S. Hemalatha, U. D. Acharya, A. Renuka, and P. R. Kamath, "A secure and high capacity image steganography technique," *Signal & Image Processing*, vol. 4, no. 1, p. 83, 2013.
[15] S. A. Laskar and K. Hemachandran, "High capacity data hiding using lsb steganography and encryption," *International Journal of Database Management Systems*, vol. 4, no. 6, p. 57, 2012.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 645–651
DOI: 10.15439/2016F516
ACSIS, Vol. 8. ISSN 2300-5963

Optimizing the parameters of Sugeno based adaptive neuro fuzzy using artificial bee colony: A Case study on predicting the wind speed

Fatma Helmy Ismail *, Mohamed Abdel Aziz †,§ Aboul Ella Hassanien‡,
* Faculty of Computer Science Misr International University, Egypt
† Faculty of Sceince, Zagazig University, Egypt
‡ Faculty of Computers and Information, Cairo University, Egypt
§ Faculty of Computer science, Nahda University, Beni Suef, Egypt.

***Abstract*—This paper presents an approach based on Artificial Bee Colony (ABC) to optimize the parameters of membership functions of Sugeno based Adaptive Neuro-Fuzzy Inference System (ANFIS). The optimization is achieved by Artificial Bee Colony (ABC) for the sake of achieving minimum Root Mean Square Error of ANFIS structure. The proposed ANFIS-ABC model is used to build a system for predicting the wind speed. To ensure the accuracy of the model, a different number of membership functions has been used. The experimental results indicates that the best accuracy achieved is 98% with ten membership functions and least value of RMSE which is 0.39.**

***Index Terms*—Wind Speed Prediction, Adaptive Network Based Fuzzy Inference System (ANFIS), Artificial Bee Colony (ABC), Swarm Intelligence, Root Mean Square Error (RMSE).**

I. Introduction

ADAPTIVE Neuro-Fuzzy Inference Systems (ANFIS)[1] can be used for energy planning. Its learning techniques is to adjust the parameters of FIS membership functions that best represent the given input/output data. An adaptive neuro fuzzy control [2] is applied to optimize the use of wind energy in smart grids. In [3] the ANFIS have have applied for wind power prediction. Also, the authors in [4] predicted the wind speed using soft computing models formulated on a back propagation neural network (BPNN) and an adaptive neurofuzzy inference system (ANFIS). The adaptive neuro-fuzzy inference system (ANFIS) has been applied to estimate optimal power coefficient value of the wind turbines by [5]. In [6] the fuzzy modeling techniques and artificial neural networks have applied to estimate annual energy output of a wind turbine. the authors in [7] have demonstrated an online fuzzy neural network controller for output maximization in a wind energy conversion system. An on-line training recurrent fuzzy neural network (RFNN) controller for wind generation system with a high-performance model reference adaptive system (MRAS) observer for the sensorless control of an induction generator (IG) have presented [8]. In [9,10] ANFIS have applied for wind speed profiling and for wind power prediction. In [11] the wind speed is predicted using fuzzy logic and artificial neural network.

The artificial bee colony (ABC) algorithm is relatively a new swarm intelligence based optimizer [12][13]. Some good properties of ABC has been revealed in[14][16]. Especially, the number of controlling parameters in ABC is less than that of other population-based algorithms, which makes it easier to be implemented. Moreover, the optimization performance of ABC is comparable and sometimes superior to the state-of-the-art meta-heuristics.That is why ABC has aroused interest and has been successfully applied to different kinds of optimization problems [17][19].

This paper presents an application of ANFIS-ABC to predict the wind speed. Where the ABC algorithm is applied to search the optimal parameters of ANFIS structure. The best parameters of membership functions are adjusted again using ABC.

Section II presents an outline of ANFIS structure. Section III introduces the proposed ANFIS-ABC structure. Section IV demonstrates the experimental results. Section V shows the implementation and results of building a wind speed prediction system. Section VI introduces future work and conclusion.

II. ANFIS

ANFIS is an adaptive fuzzy inference system [1]. The architecture of ANFIS is shown in Fig (1). The first two stages of the fuzzy inference process are fuzzifying the inputs and applying the fuzzy operator. The output of Sugeno membership functions are either linear or constant. The rule in a Sugeno fuzzy model has two main components, the antecedent and the consequent parts and has the form
If x_1 is A_{i1} and x_2 is A_{i2}, then y_i is f_i(x)
where x_1, x_2 are the input variables to the ANFIS. A_{i1},...,A_{in} are the linguistic variables of input membership function for the ith rule ($i = 1,2,...,n$) and y_i is the consequent part of ith rule. The fuzzy set A_{ij} at layer one uses a Gaussian membership function for each input variable and it has the form shown in Eq(1)

$$A_{ij}(x) = e^{-(\frac{x_j - m_{ij}}{\sigma_{ij}})^2} \quad (1)$$

where m_{ij} and σ_{ij} are the center and the width of the fuzzy set A_{ij} respectively. The parameters of this layer are the *antecedent parameters*. The output of the fuzzy inference system with n rules is calculated by weighting the real

values of consequent parts of all rules with the corresponding membership grade is shown in Eq(2,3,4)

$$\hat{y} = \sum_{i=1}^{n} (\bar{w}_i f_i) = \frac{w_i}{\sum_{i=1}^{n} (w_i)} \tag{2}$$

where

$$w_i = \prod_{j=1}^{n} A_{ij}(x_i) \tag{3}$$

and

$$y_i = f_i(x) = a_i x_1 + b_i x_2 + c_i \tag{4}$$

where a_i, b_i and c_i are the set of *consequent parameters*.

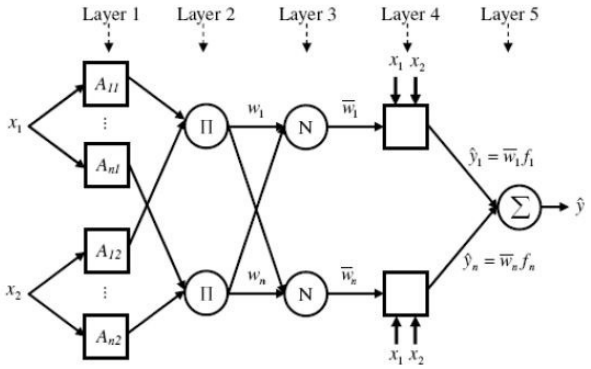


Fig. 1: Sugeno-Type Fuzzy Inference

III. Proposed ANFIS-ABC

In Artificial Bee Colony (ABC), the colony contains three groups: employed bees, onlookers and scouts. The first half of the colony consists of employed bees and the second half contains onlookers. Employed bees search for food sources and share the information about these sources to recruit onlookers. The food sources found by all employed bees are selected and exploited by onlookers according to the probability proportional to the quality of food sources. Scouts are generated from a few employed bees, which abandon their food sources through a predetermined number of cycles called *limit* and search new ones. The position of each food source is a possible solution to an optimization problem, and its fitness is the profitability of that food source. The number of food sources equals to the number of employed bees. Initially, a population containing SN solutions is generated randomly. SN is the number of food sources, which is half of the population size (NP). Let X_i= x_{i1},x_{i2},...,x_{iD} represents the ith food source in the population, where D is the number of optimization parameters. The population is subject to repeated cycles, C=1,2,..., maximum cycle number (MCN), of the search processes of the employed bees, onlookers and scouts. in ABC, the fitness function is defined as Eq(5)

$$fit(X_i) = \begin{cases} \frac{1}{1+f(X_i)} & if \;\; f(X_i) \geq 0 \\ 1 + abs(f(X_i)) & if \;\; f(X_i) < 0 \end{cases} \tag{5}$$

where $f(X_i)$ is the value of objective function of X_i, $fit(X_i)$ is the fitness value of X_i. The probability of a food source being selected by an onlooker can be represented by Eq(6)

$$p(X_i) = \frac{fit(X_i)}{\sum_{n=1}^{SN} fit(X_n)} \tag{6}$$

Afetr discovering or selecting the food source X_i by an employed bee or an onlooker, they exploit a neighboring food source V_i. V_i is determined by changing only one parameter of X_i, where $v_{ij} \neq x_{ij}$, while the rest of V_i keep the same value as X_i, v_{ij} is generated by Eq (7)

$$v_{ij} = x_{ij} + \phi_{ij}(x_{ij} - x_{kj}) \tag{7}$$

where $k \in$ 1,2,..,SN is a random chosen inde and k must be different from from i, $j \in$ 1,2,...,D , ϕ_{ij} is a random number between [-1 , 1]. After determining a new candidate food source in the neighborhood of its currently associated food source using Eq (7) by an employed bee or onlooker , a greedy selection method is used to distinguish between the new food source and the old one. if the abandoned food source is X_i, a scout produces a new food source according to Eq (8)

$$v_{ij} = x_{minj} + rand(0,1)(x_{maxj} - x_{minj}) \tag{8}$$

where x_{minj} and x_{maxj} are the lower and upper bounds of the variable x_{ij}, respectively.

ABC is a possible technique to optimize the parameters of ANFIS. In ANFIS-ABC, ANFIS parameters are considered as one food source that represent a possible solution, and parameters that affect the ANFIS training can be taken as the dimensions of each food source.

Two parameters in ANFIS are to be optimized which are the linguistic hedges p that affect the membership function values and the consequent parameters k that accilerates the performances value.

Denote X as a food source, and let p denotes the set of linguistic hedges of each input where $p \in \{verylow, low, medium, ...\}$. Each linguistic hedge p is presented by gaussian membership function $\mu^{gaussian}$ whose parameters are center m and the width σ as shown in Eq (1). The number of linguistic hedges per input equals the number of ANFIS rules R. Then an algebraic representation of the antecedent parameters is shown in Eq (9).

$$X = \left\{ \mu_{jr}^{gaussian} | r \in R; j \in J \right\} \tag{9}$$

where J is a set of inputs and R is a set of rules that forms ANFIS-ABC.

Let k be denoted as the consequent parameters of each food source. The consequent parameters k of each rule output presents the parameters of a linear membership function μ^{linear} whose the number of parameters equals J+1. Then an algebraic representation of the length of the dimension formed by k is shown in Eq(10).

$$X = \left\{ \mu_{r}^{linear} | r \in R \right\} \tag{10}$$

Because the dimensions of one food source in the ANFIS-ABC has two parameters, then an algebraic representation of that one food source is presented in Eq(11).

$$X = \left\{ (\mu_{jr}^{gaussian}, \mu_{r}^{linear}) | j \in J; r \in R \right\} \quad (11)$$

Fig (2) shows The coding of antecedent and consequent parameters in each food source. The same as ABC, the ANFIS-

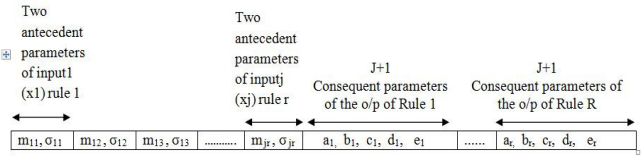


Fig. 2: The coding of ANFIS-ABC source

ABC consists of a number of food sources. Then an algebraic representation of the dimension of each food source in the ANFIS-ABC is represented in Eq (12).

$$X_i = \left\{ (\mu_{ijr}^{gaussian}, \mu_{ir}^{linear}) | j \in J; r \in R \right\} \quad (12)$$

where i=1,2,...N and N is the number of food sources. ANFIS-ABC requires an objective function to minimize which is the root mean square error (RMSE) of ANFIS structure. Eq(13) shows the RMSE calculation.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{s} (\hat{y}_i - y_i)^2}{s}} \quad (13)$$

where y_i is the observed value for the ith observation and $\hat{y}_i$ is the predicted output from fuzzy model, and s is the number of training data pairs. Fig (3) shows the whole ANFIS-ABC model. Fig (4) shows the antecedent part of ANFIS-ABC

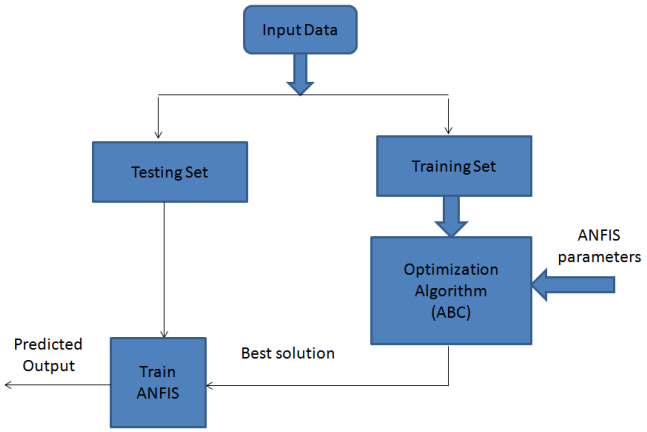


Fig. 3: ANFIS-ABC Structure

parameters and the process of generating a firing strength for each output rule. Fig (5) shows the consequent part of ANFIS-ABC and generating the final crisp output. Fig (6) shows how the objective function (RMSE) of the ANFIS-ABC model is calculated and then converted into fitness value. Fig (7) shows

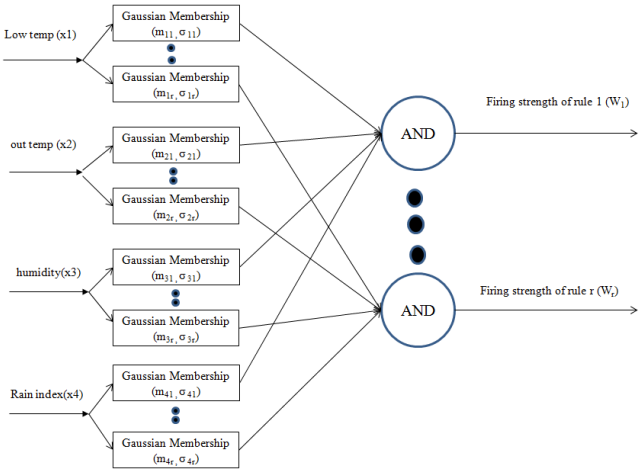


Fig. 4: Antecedent Part of ANFIS-ABC

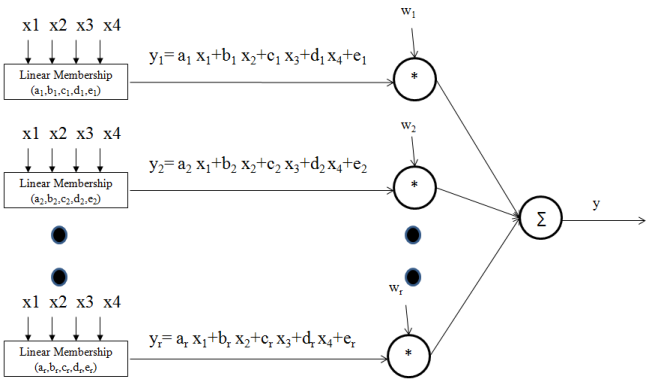


Fig. 5: Consequent Part of ANFIS-ABC

the proposed ANFIS-ABC structure and The main steps of ANFIS-ABC are outlines below as in algorithm 1:

IV. Experimental Results

Seventy percent of the data set is used in training ANFIS-ABC model. The parameters initialization and the achieved results are discussed in this section.

A. Data set description

Each instance of the data set consists of four inputs and only one output. The four inputs are the low temperature, the out temperature, the humidity and the rain index and are represented as ($In1$, $In2$, $In3$ and $In4$). The output value indicates the wind speed. Only 70% of the dataset is used in training the model which constitutes 2128 records out of 3040 records.

B. ANFIS parameters

The ANFIS model parameters are shown in table I. The ANFIS-ABC model was trained four times with different number of rules. In this training procedure, the ANFIS-ABC algorithm is used to optimise both the antecedent parameters

Algorithm 1: Pseudocode of the ANFIS-ABC algorithm

1: set the training dataset to be 70% of the whole data.
2: Initialize ANFIS structure parameters: J, R, Type of Input/Output membership functions.
3: Present the values of ABC parameters: D, SN, MCN, $limit$, $cycle$=1.
4: Form the ANFIS-ABC food sources using Eq (12) and Initialize The ANFIS-ABC food sources using eq (8).
5: ANFIS-evalute(foodsources).
6: **repeat**
7: Produce new solutions food sources for the employed bees using Eq (7) and ANFIS-evalute(foodsources), then apply greedy selection process employed bees'phase.
8: calculate the probability values for food sources using Eq(6).
9: Produce new food sources for the onlookers from the food source X_i selecetd depending on $p(X_i)$ using Eq (7) and ANFIS-evalute(foodsources), then apply the greedy selection process onlookers' phase.
10: Determine the abandoned food source for the scout, if exists, and replace it with a new randomly produced solution using Eq(8) scout's phase.
11: Memorize the best food source achieved so far.
12: $cycle$=$cycle$+1.
13: **until** $cycle$=Maximum Cycle Number (MCN)

Algorithm 2: ANFIS-evalute

1: Build ANFIS structure
2: Input Fuzzification.
3: Calculate firing strength.
4: Calculate summed crisp output.
5: Calculate RMSE using Eq(13).

TABLE I: ANFIS Parameters

ANFIS Parameters	
Number of Crisp Inputs (J)	4
Input Membership Functions Type	Gaussian
Number of optimized Parameters of Gaussian Membership Function	2
Output Membership Function Type	Linear
Number of Optimized Parameters of Linear Membership Function	J+1
Number of Output Rules (R)	3 or 4 or 5 or 10
Number of Fuzzy Sets per Input	3 or 4 or 5 or 10

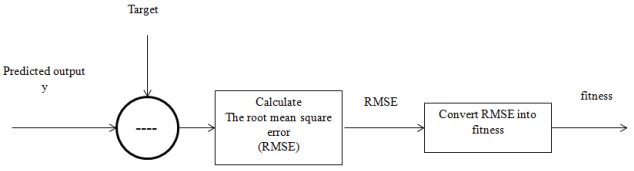


Fig. 6: Evaluating the output of ANFIS-ABC

TABLE II: ABC Parameters

ABC Parameters	
Number of Optimization Parameters (D)	J * R * 2 + R (J+1)
Number of Food Sources (SN)	120
Number of Employed Bees (Ebees)	60
Number of Onlookers (Onbees)	60
limit (L)	round(0.6*D*Ebees)
Maximum Cycle Number (MCN)	300

and the consequent parameters. A Gaussian membership function is used in the input layer and linear membership function is used in the output layer. The number of variables in each food source will depends on the number of parameters of membership function for each input and output. The consequent part of each z_i Suegno ANFIS rule will take the form k_1*$In1$+ k_2*$In2$ + k_3*$In3$ + k_4*$In4$ + k_5. where k_1, k_2, k_3, k_4 and k_5 are the consequent parameters of each rule output.

C. ABC parameters

the ABC parameters are initialized as shown in table II

D. Training ANFIS-ABC system with different number of membership functions

The system has been trained using different number of rules, hence different number of membership functions per input/output. The objective of the training is to obtain the parameters of membership functions that achieve least mean square error value. The results are shown in table III. It is worth to note that the increase in the number of membership functions does not improve the value of RMSE. The case of using four membership functions shows a slight improvement.

V. Wind Speed Prediction System

The optimized ANFIS parameters obtained by ABC are used to build and test a system for wind speed prediction. Thirty percentage of the data is used for testing the prediction

TABLE III: Results of Training ANFIS-ABC model

Training ANFIS-ABC Model	
Number of Membership Functions	RMSE
3	0.530
4	0.440
5	0.410
10	0.390

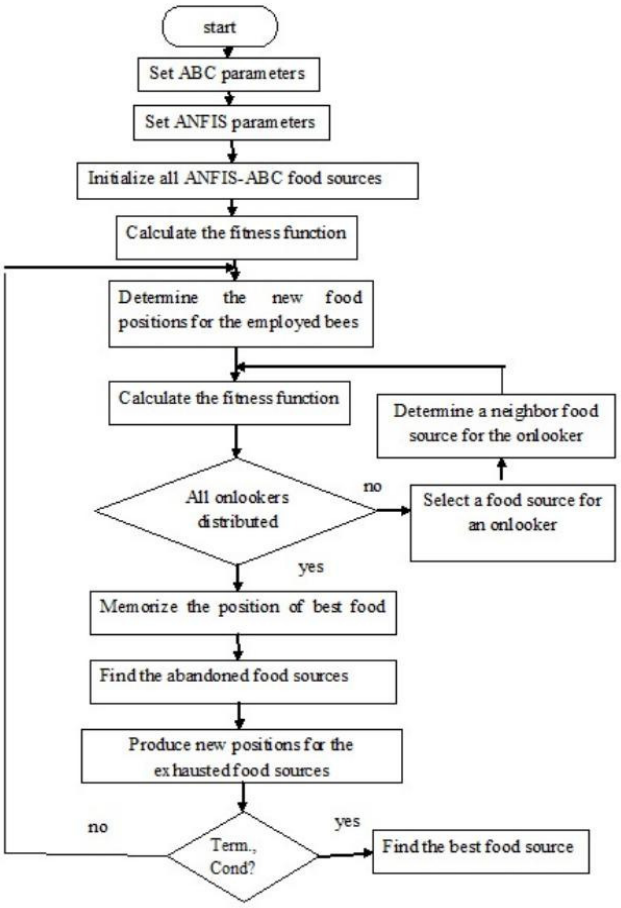


Fig. 7: Proposed ANFIS-ABC structure

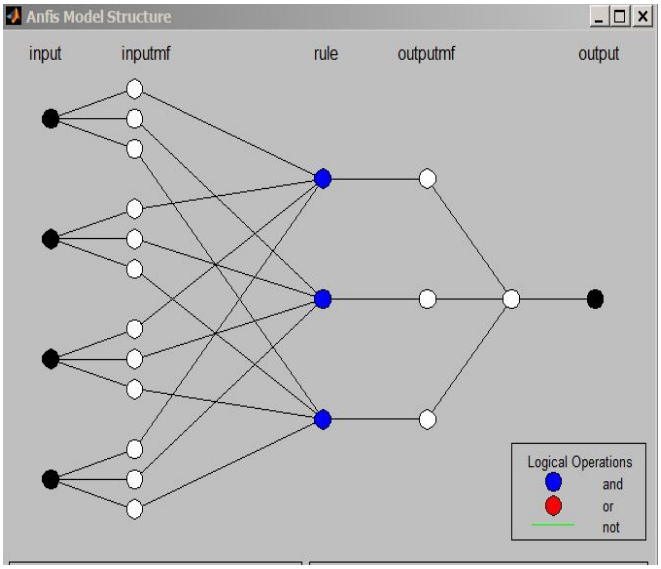


Fig. 8: ANFIS-ABC Structure with three rules

Fig. 9: Wind Prediction with three rules

system. Different number of membership functions are used along with 4 input parameters (low temeperature, out temperature, humidity and rain index). The accuracy of the resulting system is discussed below.

A. Wind Speed Prediction with 3 Membership Functions

Three membership functions means three linguistic variables low,medium and high for the measurements of the values of each input parameter. Fig 8 shows the ANFIS(ABC) structure and Fig 9 shows the generated system which can be used to predict the wind speed (output) by enter four input values corresponding to the input parameters. The average testing error is 0.29 which indicates accuracy with 71%. Fig (4,5) show the ANFIS-ABC structure and output with three rules.

B. Wind Speed Prediction with 4 Membership Functions

Four membership functions means four linguistic variables very low, low, medium and high for the measurements of the values of each input parameter. Fig 10 shows the ANFIS(ABC) structure and Fig 11 shows the generated system which can be used to predict the wind speed (output) by enter four input values corresponding to the input parameters. The average testing error is 0.30 which indicates accuracy with 70%.

C. Wind Speed Prediction with 5 Membership Functions

Five membership functions for each input means five linguistic variables very low, low, medium, high and very high for the measurements of the values of each input parameter. Fig 11 shows the ANFIS(ABC) structure and Fig 12 shows the generated system which can be used to predict the wind speed (output) by enter five input values corresponding to the input parameters. The average testing error is 0.29 which indicates accuracy with 71%.

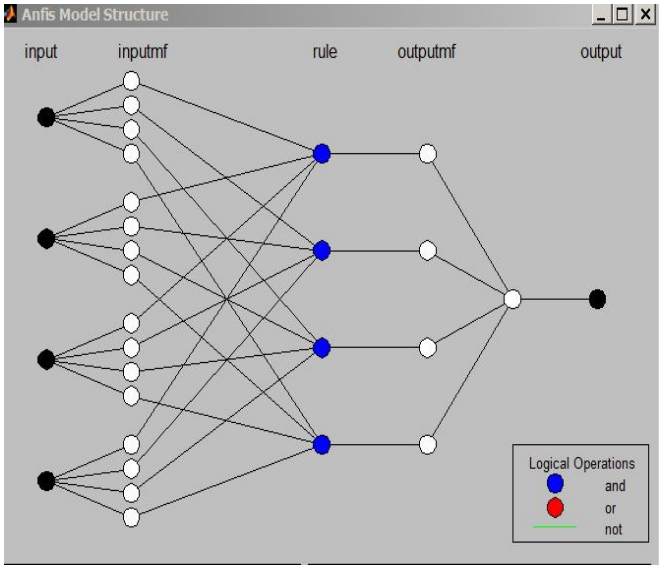


Fig. 10: ANFIS-ABC Structure with four rules

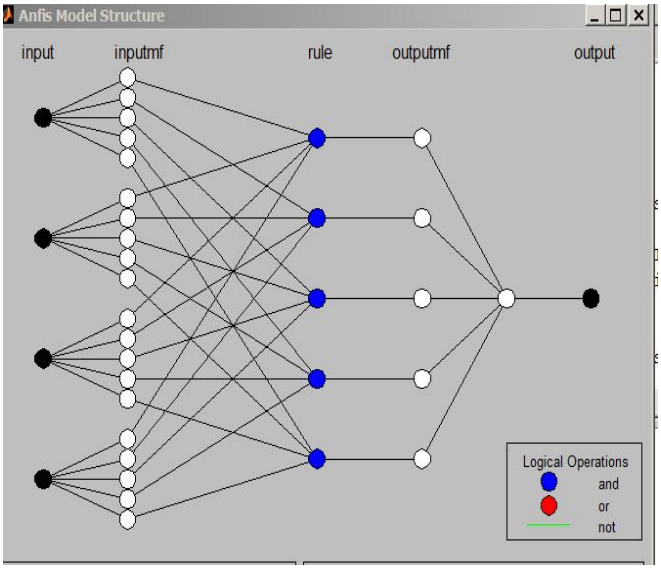


Fig. 12: ANFIS-ABC Structure with five rules

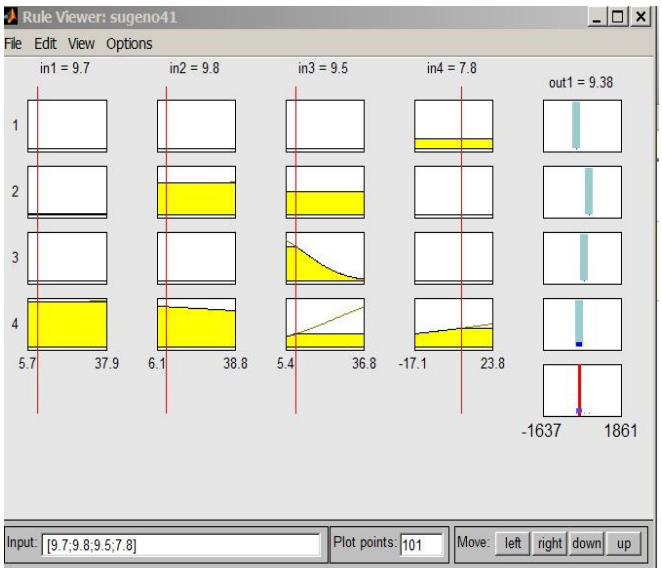


Fig. 11: Wind Prediction with four rules

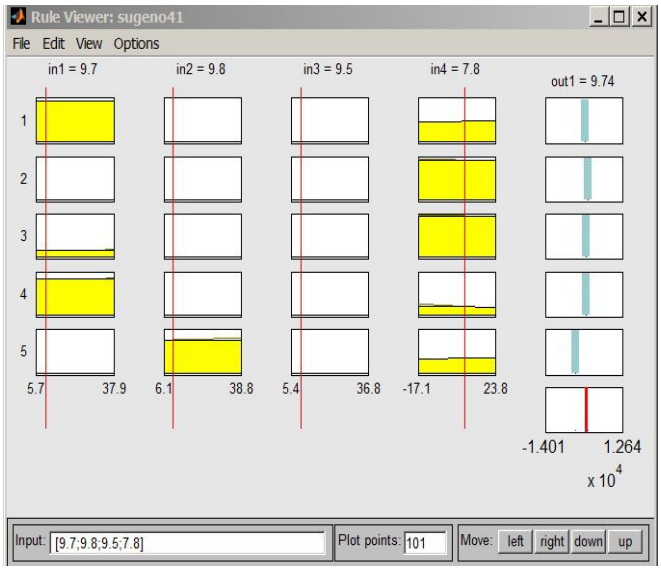


Fig. 13: Wind Prediction with five rules

D. Wind Speed Prediction with 10 Membership Functions

Ten membership functions for each input means Ten different classes for the measurements of the values of each input parameter. Fig 14 shows the generated system which can be used to predict the wind speed (output) by enter four input values corresponding to the input parameters. The average testing error is 0.29 which indicates accuracy with 71%. from the above discussion, It is deduced that the best achieved accuracy is 70% with four membership functions. The increase of membership functions does not improve the accuracy significantly. Table IV shows the summary of achieved results.

VI. Conclusion and Future Works

A new approach for optimising the Sugeno adaptive neuro-fuzzy inference system (ANFIS) in prediction problems has been proposed in this paper. The ABC technique is integrated into the process of ANFIS in order to achieve the optimal solution for ANFIS. This was achieved by simultaneously optimising the ANFIS performance based on a criteria which is enhancing the accuracy based on lower error rate. The experimental results indicated that ANFIS-ABC provides a promised accuracy in prediction problems. However, an algorithm that can result in a complete balance of accuracy and interpretability would be more adaptable for real applications.

TABLE IV: Results of Testing ANFIS-ABC

Results of Testing ANFIS-ABC			
Num of Membership Functions	MAE	MAPE%	Accuracy%
3	0.43	2.6%	97.4%
4	0.33	2.4%	97.6%
5	0.3	1.9%	98.1%
10	0.29	1.8%	98.2%

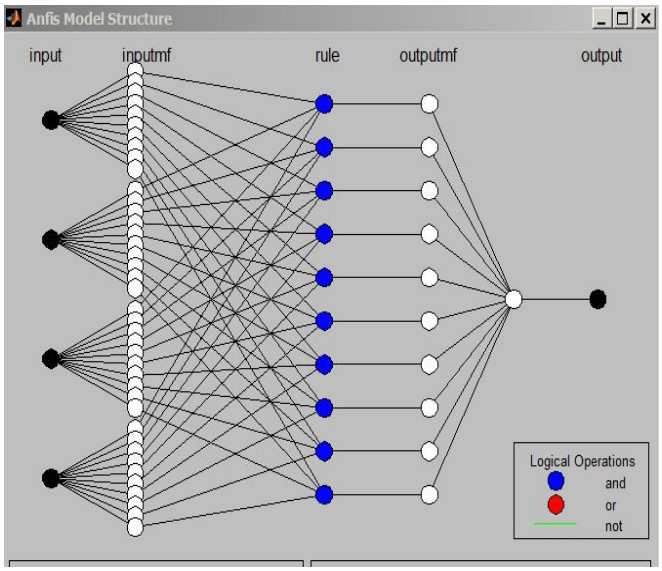


Fig. 14: ANFIS-ABC Structure with ten rules

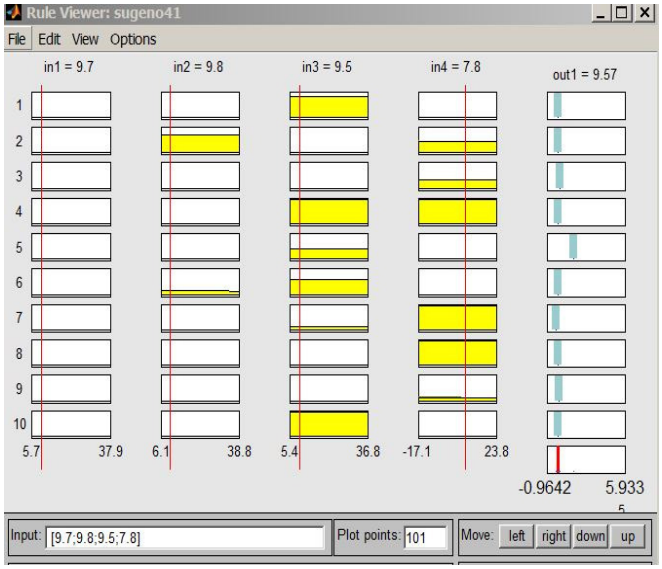


Fig. 15: Wind Prediction with ten rules

Thus, problems based on this approach are the subject of further work by integrating other swarm algorithms.

REFERENCES

[1] Jang J. S. R. Adaptive network based fuzzy inference systems. IEEE Transactions on systems man and cybernetics 1993, p. 665-685.
[2] Capovilla CE, Casella IRS, Filho AJS, Azcue-Puma JL, Jacomini RV, Ruppert E. A wind energy generator for smart grid applications using wireless-coded neuro-fuzzy power control. Comput Math Appl 2014;68:2112-23.
[3] De Giorgi MG, Ficarella A, Tarantino M. Error analysis of short term wind power prediction models. Appl Energy 2011;88:1298-311.
[4] Haque AU, Mandal P, Kaye ME, Meng J, Chang J, Senjyu T. A new strategy for predicting short-term wind speed using soft computing models. Renew Sustain Energy Rev 2012;16:4563-73.
[5] Petković D, Žarko C, Nikolić V. Adaptive neuro-fuzzy approach for wind turbine power coefficient estimation. Renew Sustain Energy Rev 2013;28:191-5.
[6] Jafarian M, Ranjbar AM. Fuzzy modeling techniques and artificial neural networks to estimate annual energy output of a wind turbine. Renew Energy 2010;35:2008-14.
[7] Lin WM, Hong CM, Cheng FS. Fuzzy neural network output maximization control for sensorless wind energy conversion system. Energy 2010;35:592-601.
[8] Lin WM, Hong CM, Cheng FS. Design of intelligent controllers for wind generation system with sensorless maximum wind energy control. Energy Convers Manag 2011;52:1086-96.
[9] Mohandes M, Rehman S, Rahman SM. Estimation of wind speed profile using adaptive neuro-fuzzy inference system (ANFIS). Appl Energy 2011;88:4024-32.
[10] De Giorgi MG, Ficarella A, Tarantino M. Error analysis of short term wind power prediction models. Appl Energy 2011;88:1298-311.
[11] Monfared M, Rastegar H, Kojabadi HM. A new strategy for wind speed forecasting using artificial intelligent methods. Renew Energy 2009;34:845-8.
[12] Karaboga, D.; Akay, B. A survey: algorithms simulating bee swarm intelligence. Artif. Intell. Rev. 2009, 31, 61–85.
[13] D. Karaboga, B. Basturk, A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm, Journal of Global Optimization 39(2007) 171–459.
[14] Karaboga, D.; Basturk, B. On the performance of artificial bee colony (ABC) algorithm. Appl. Soft Comput. 2008, 8, 687–697.
[15] Karaboga, D.; Akay, B. A comparative study of artificial bee colony algorithm. Appl. Math. Comput. 2009, 214, 108–132.
[16] Karaboga, D.; Basturk, B. A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm. J. Glob. Optim. 2007, 39, 459–471.
[17] Kang, F.; Li, J.; Xu, Q. Structural inverse analysis by hybrid simplex artificial bee colony algorithms. Comput. Struct. 2009, 87, 861–870.
[18] Sonmez, M. Discrete optimum design of truss structures using artificial bee colony algorithm. Struct. Multidiscip. Optim. 2011, 43, 85–97.
[19] Samanta, S.; Chakraborty, S. Parametric optimization of some non-traditional machining processes using artificial bee colony algorithm. Eng. Appl. Artif. Intell. 2011, 24, 946–957.

9th Computer Aspects of Numerical Algorithms

NUMERICAL algorithms are widely used by scientists engaged in various areas. There is a special need of highly efficient and easy-to-use scalable tools for solving large scale problems. The workshop is devoted to numerical algorithms with the particular attention to the latest scientific trends in this area and to problems related to implementation of libraries of efficient numerical algorithms. The goal of the workshop is meeting of researchers from various institutes and exchanging of their experience, and integrations of scientific centers.

TOPICS

- Parallel numerical algorithms
- Novel data formats for dense and sparse matrices
- Libraries for numerical computations
- Numerical algorithms testing and benchmarking
- Analysis of rounding errors of numerical algorithms
- Languages, tools and environments for programming numerical algorithms
- Numerical algorithms on coprocesors (GPU, Intel Xeon Phi, etc.)
- Paradigms of programming numerical algorithms
- Contemporary computer architectures
- Heterogeneous numerical algorithms
- Applications of numerical algorithms in science and technology

EVENT CHAIRS

- **Bylina, Beata,** Maria Curie-Sklodowska University, Poland
- **Bylina, Jaroslaw,** Maria Curie-Sklodowska University, Poland
- **Stpiczyński, Przemysław,** Maria Curie-Sklodowska University, Poland

PROGRAM COMMITTEE

- **Amodio, Pierluigi,** Università di Bari, Italy
- **Anastassi, Zacharias,** Qatar University, Qatar
- **Banaś, Krzysztof,** AGH University of Science and Technology, Poland
- **Brugnano, Luigi,** Universita' di Firenze, Italy
- **Czachorski, Tadeusz,** IITiS
- **Filote, Constantin**
- **Fourneau, Jean-Michel**
- **Gansterer, Wilfried,** University of Vienna, Austria
- **Georgiev, Krassimir,** IICT - BAS, Bulgaria
- **Gravvanis, George,** Democritus University of Thrace, Greece
- **Kozielski, Stanislaw**
- **Kucaba-Pietal, Anna,** Politechnika Rzeszowska, Poland
- **Lirkov, Ivan,** Institute of Information and Communication Technologies, Bulgarian Academy of Sciences, Bulgaria
- **Maksimov, Vyacheslav,** Institute of Mathematics and Mechanics, Russia
- **Marowka, Ami,** Bar-Ilan University, Israel
- **Mycka, Jerzy,** UMCS
- **Petcu, Dana,** West University of Timisoara, Romania
- **Satco, Bianca-Renata,** Stefan cel Mare University of Suceava, Romania
- **Sedukhin, Stanislav,** The University of Aizu, Japan
- **Sergeichuk, Vladimir,** Institute of Mathematics of NAS of Ukraine, Ukraine
- **Shishkina, Olga,** Max Planck Institute for Dynamics and Self-Organization, Germany
- **Srinivasan, Natesan,** Indian Institute of Technology, India
- **Szajowski, Krzysztof,** Institute of Mathematics and Computer Science, Poland
- **Tudruj, Marek,** Inst. of Comp. Science Polish Academy of Sciences/Polish-Japanese Institute of Information Technology, Poland
- **Tůma, Miroslav,** Academy of Sciences of the Czech Republic, Czech Republic
- **Ustimenko, Vasyl,** Marie Curie-Sklodowska University, Poland
- **Vazhenin, Alexander,** University of Aizu, Japan
- **Wyrzykowski, Roman,** Czestochowa University of Technology, Poland

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 655–664
DOI: 10.15439/2016F436
ACSIS, Vol. 8. ISSN 2300-5963

Parallelizing nested loops on the Intel Xeon Phi on the example of the dense WZ factorization

Jarosław Bylina, Beata Bylina
Marie Curie-Skłodowska University,
Institute of Mathematics,
Pl. M. Curie-Skłodowskiej 5,
20-031 Lublin, Poland
Email: {jaroslaw.bylina,beata.bylina}@umcs.pl

***Abstract*—In this article we evaluate some strategies of parallelizing nested loops on Intel Xeon Phi on the example of the WZ factorization for dense matrices. We employ both parallelism and vectorization to accelerate nested loops on manycore coprocessor.**

For random dense square matrices with the dominant diagonal we report the execution time and the performance of the nested loops. Numerical experiments show that the vectorization that is efficiently exploiting SIMD vector units do not always improve the application performance on the coprocessor.

I. Introduction

MANYCORE computers with shared memory are used to solve the computational science problems. One of the machines with manycore architecture is Intel Xeon Phi [8], [9], which is based on Intel's Many Integrated Core (MIC). Intel Xeon Phi has got over 60 cores, hardware threading capabilities and wide vector units (VPU).

To implement parallel programs on manycore systems with shared memory, in particular on Intel Xeon Phi, programmers can use the OpenMP standard [12] as for the traditional multicore processors. The programming model provides a set of directives to explicitly define parallel regions in applications. The compiler translates these directives. One of its most interesting features in the language is the support for the nested parallelism.

In the scientific applications, loops are an important source of the parallelism — nested loops, in particular. Parallelizing nested loops require the programmer to make a decision about applying some strategies of the parallelization and the vectorization.

The research of the parallelization of nested loops have been undertaken by different scientists.

In the work [10], the authors study five different models for the nested parallel loops execution on shared-memory manyprocessors and show a simulation-based performance comparison of different techniques using a real application. The possibility to take advantage of the parallelism in nested parallel loops with the use of good scheduling and synchronization algorithms is described.

An automatic mechanism to dynamically detect the best way to exploit the parallelism when having nested parallel loops is presented in the study [5]. This mechanism takes into account the number of threads, the size of the problem, the number of iterations in a loop and it was implemented inside the IBM XL runtime library. That paper examined (among others) an LU kernel, which decomposes the matrix $\mathbf{A}$ into the matrices: $\mathbf{L}$ (a lower triangular matrix) and $\mathbf{U}$ (an upper triangular matrix).

An algorithm for finding good distributions of threads to tasks is provided and the implementation of the nested parallelism in OpenMP is discussed in the paper [1].

The focus of [7] was to investigate the possibility of dynamically choosing, at runtime, the loop which utilizes the available threads the best.

One of the direct methods of solving a dense linear system is to factorize the matrix into some simpler matrices — it is its decomposition into factor matrices of a simpler structure or of some specific properties — and then solving simpler linear systems. The most known factorization is the LU factorization (mentioned above). Another form of the factorization is the WZ factorization. In the work [2] we investigated four strategies of parallelizing nested loops on multicore architectures on the example of the WZ factorization [3], [6], [11]. We dealt with the following parallelism strategies for nested loops: *outer*, *inner*, *nested* and *split*.

For random dense square matrices with the dominant diagonal we reported the execution time, the performance, the speedup of the WZ factorization for these four strategies of parallelizing nested loops and we investigated the accuracy of such solutions. The *outer* and *split* approaches achieved the best speedup.

The goal of this paper is to study parallelized nested loops on Intel Xeon Phi. We research only two strategies, namely: *outer* and *split* due to their best results for multicore architectures. The efficient parallelizing of a nested loop is very difficult on Intel Xeon Phi because we must employ both a large number of threads and wide vector units available in Intel Xeon Phi. For the thread-level parallelism we use OpenMP [12], [4].

The OpenMP standard supports the loop parallelism. For the OpenMP standard, it is done by the utilization of the directive `#pragma omp parallel for`, which provides a shortcut for specifying a parallel region that contains a single `#pragma omp for`. We scale nested loops to a large number of threads and choose a good load balancing. The division of the work among threads is controlled with the

978-83-60810-90-3/$25.00

`schedule` clause.

We provide enough work for the coprocessor and we also try to efficiently exploit 512-bit vectors on Intel Xeon Phi using adequate pragmas.

The paper deals with the following issues: In Section II it describes the main characteristics of the Intel Xeon Phi architecture. Section III provides some information about some strategies of parallelizing nested loops and their application to the original WZ factorization. Section IV presents the results of our experiments. The time, the speedup, the performance of the WZ factorization for different strategies on Intel Xeon Phi are analyzed. Section V is a summary of our experiments.

II. Intel Xeon Phi architecture

Intel Xeon Phi [9] coprocessors are one of the state-of-the-art architectures which goal is to execute parallel codes. Intel Xeon Phi is a manycore coprocessor created on the basis of the Intel MIC (Many Integrated Cores) technology. The first generation the Intel MIC architecture is based on the Knights Corner chips. Many redesigned Intel CPU cores are connected by a bi-directional 512-bit ring bus. The cores are enriched with 64-bit instructions and the L1 and L2 cache memories. Each of the cores contains a vector processing unit (VPU), which together with 32 512-bit vector registers allows to process many data with the use of one instruction (SIMD instructions).

A single Intel Xeon Phi has 61 cores of 1,238 GHz frequency and it serves 244 threads and communicates through PCI-Express 2.0.

Intel Xeon Phi enhances the performance of applications written in the C/C++ and Fortran languages. The Intel company offers a set of programming tools assisting programming processes such as compilers, debuggers, libraries, that allow creating parallel applications (e. g. Pthread, OpenMP, Intel Cilk plus, MPI). Intel Xeon Phi is able to use standard parallel programming models such as OpenMP.

The Intel Xeon Phi coprocessors can work in two executing modes: the native mode or the offload mode (for one Intel Xeon Phi).

In the native mode the task is executed directly by a coprocessor, which makes it a separate computing node. The compilation of the source code for the accelerator architecture demands so-called cross-compiling, which produces a file executable on Intel Xeon Phi. The native application can be started by hand on the coprocessor or by the *micnativeloadex* tool which automatically copies the program together with necessary files and then starts it.

III. Nested loops parallelization

An application with nested loops can be performed in parallel in different ways depending on compilers, hardware and run-time system support available. Nested loops require a programmer to take a decision concerning details of the parallelism. Nested loops can have a few levels of the parallelism. The outermost loop contains other loops. Next, each of these loops may also consist of loops. It is a reason for the increased complexity of the implementation.

Frequently, the parallelization is applied to the outer loop levels and the vectorization to the inner levels. If you are certain that the vectorization is a safe alternative (gives the same results as the non-vectorized code) in a particular loop where the compiler itself does not vectorize, using `#pragma simd` often provides the best and the most predictable benefits.

In this work we deal with the following parallelization strategies for the nested loops:

1) *outer*;
2) *split*.

All variables used in a parallel region are by default `shared`; in each strategy we declare explicitly all variables as `private` or `shared` for all directives respectively. Using the `private` clause, we specify that each thread has its own copy of variables.

To ensure a good load balancing for all threads we use the `schedule` clause, which specifies how the iterations of the loop are assigned to the threads. In the directive `#pragma omp parallel for` in the clause `schedule` we set values `static` or `dynamic`. We research the impact of this clause on the efficiency.

A. Outer

Outer — the simplest parallelization strategy of nested loops is the parallel execution of the outermost loop (not counting the loop which cannot be parallelized). This approach gives good results if the number of iterations in a loop is big and the iteration's granularity is coarse enough.

Figure 1 presents a listing of the outer strategy for the WZ factorization. The outermost k-loop cannot be parallelized. However, we can parallelize the i-loop. In this simple parallelization strategy the loop is divided equally between threads using both the static and dynamic scheduler.

Figure 2 also presents a listing of the outer strategy for the WZ factorization with the use of the vectorization. Again the outermost k-loop cannot be parallelized, however, we can parallelize the i-loop. The loop is divided equally between threads. The inner loop is vectorized. The compiler is unable to automatically vectorize the inner loop due to the vector dependences. To vectorize this code we use `#pragma simd`.

B. Split

The second (and final) strategy consists in the division of the i-loop into two separate loops and we denote it by *split*. Figure 3 shows a listing of the *split* strategy for the WZ factorization. The first loop is parallelized. The second loop is a nested loop and we execute only its outer loop in parallel.

Figure 4 also shows a listing of the *split* strategy for the WZ factorization. Here, the first loop is parallelized. The second loop is a nested loop and we execute its outer loop in parallel and we vectorize it at the same time. For the OpenMP standard, it is done by the utilization of the directive `#pragma omp parallel for simd`, which provides a shortcut for

```
for (k=1; k<=n/2-1; k++) {
  k2=n-k+1;
  det=a[k2,k]*a[k,k2]-a[k,k]*a[k2,k2];
#pragma omp parallel for default(none) private(i,j) shared(w,k,k2,n,a,det) \
schedule(static/dynamic)
  for (i=k+1; i<=(k2-1); i++) {
    w[i,k]=(a[k2,k] * a[i,k2] - a[k2,k] * a[i,k])/det;
    w[i,k2]=(a[k,k2] * a[i,k]-a[k,k] * a[i,k2])/det;
    for (j=k+1; j<=k2-1; j++)
        a[i,j]=a[i,j] - w[i,k]*a[k,j] - w[i,k2] * a[k2,j];
  }//for i
}//for k
```

Fig. 1. A fragment of the WZ factorization algorithm — the outer strategy with the static or dynamic scheduler without the vectorization

```
for (k=1; k<=n/2-1; k++) {
    k2=n-k+1;
    det=a[k2,k]*a[k,k2]-a[k,k]*a[k2,k2];
#pragma omp parallel for default(none) private(i,j) shared(w,k,k2,n,a,det) \
schedule(static/dynamic)
    for (i=k+1; i<=(k2-1); i++) {
      w[i,k]=(a[k2,k] * a[i,k2] - a[k2,k] * a[i,k])/det;
      w[i,k2]=(a[k,k2] * a[i,k]-a[k,k] * a[i,k2])/det;
#pragma simd
      for (j=k+1; j<=k2-1; j++)
        a[i,j]=a[i,j] - w[i,k]*a[k,j] - w[i,k2] * a[k2,j];
    }//for i
}//for k
```

Fig. 2. A fragment of the WZ factorization algorithm — the outer strategy with the static or dynamic scheduler with the vectorization

```
for (k=1; k<=n/2-1; k++) {
  k2=n-k+1;
  det=a[k2,k]*a[k,k2]-a[k,k]*a[k2,k2];
#pragma omp parallel for  default(none) private(i) shared(k2,k,w,n,a,det) \
schedule(static/dynamic)
   for (i=k+1; i<=(k2-1); i++) {
     w[i,k]=(a[k2,k]*a[i,k2]-a[k2,k2]*a[i,k])/det;
     w[i,k2]=(a[k,k2]*a[i,k]-a[k,k]*a[i,k2])/det;
   }
#pragma omp parallel for  default(none) shared(k2,k,a,w) private(i,j) \
schedule(static/dynamic)
   for (i=k+1; i<=(k2-1); i++)
     for (j=k+1; j<=k2-1; j++)
       a[i,j]=a[i,j]-w[i,k]*a[k,j]-w[i,k2]*a[k2,j];
}//for k
```

Fig. 3. A fragment of the WZ factorization algorithm — the split strategy with the static or dynamic scheduler without the vectorization

```
for (k=1; k<=n/2-1; k++) {
  k2=n-k+1;
  det=a[k2,k]*a[k,k2]-a[k,k]*a[k2,k2];
#pragma omp parallel for  default(none) private(i) shared(k2,k,w,n,a,det) \
schedule(static/dynamic)
  for (i=k+1; i<=(k2-1); i++) {
    w[i,k]=(a[k2,k]*a[i,k2]-a[k2,k2]*a[i,k])/det;
    w[i,k2]=(a[k,k2]*a[i,k]-a[k,k]*a[i,k2])/det;
  }
#pragma omp parallel for simd default(none) shared(k2,k,a,w) private(i,j) \
schedule(static/dynamic)
  for (i=k+1; i<=(k2-1); i++) {
    for (j=k+1; j<=k2-1; j++)
      a[i,j]=a[i,j]-w[i,k]*a[k,j]-w[i,k2]*a[k2,j];
  }
}//for k
```

Fig. 4. A fragment of the WZ factorization algorithm — the split strategy with the static or dynamic scheduler with the parallelization of the first loop and with the parallelization and the vectorization of the second outer loop

specifying a parallel region that contains a single `#pragma omp for simd`.

Figure 5 shows a listing of another version of the *split* strategy for WZ factorization. The first loop is parallelized. The second loop is a nested loop and we execute the outer loop in parallel and we vectorize the inner loop using `#pragma simd`.

IV. Numerical experiments

In this section we show how we tested the time and the performance of the parallelized nested loops for Intel Xeon Phi. Our intention was to investigate different nested loops parallelization strategies for nested loops on manycore architectures. We examined 10 versions of the parallelized nested loops:

1) *outer*:
   - *static* (Figure 1),
   - *dynamic* (Figure 1),
   - *static+simd* (Figure 2),
   - *dynamic+simd* (Figure 2);
2) *split*:
   - *static* (Figure 3),
   - *dynamic* (Figure 3),
   - *static+forsimd* (Figure 4),
   - *dynamic+forsimd* (Figure 4)
   - *static+simd* (Figure 5),
   - *dynamic+simd* (Figure 5).

The input matrices were generated (by the authors). They were random, dense, square matrices with a dominant diagonal of even sizes (1000, 2000, ..., 12000).

A. Test environment

The tests were carried out using a computing node of the following parameters:

- Platform: Intel Server Chassis R2000WTXXX, Intel Server Board S2600WT2.
- CPU: 2×Intel Xeon E5-2670 v3 (2x12 cores, 2.3 GHz).
- Memory: 128 GB DDR4 2133MT/s (8×Crucial CT16G4RFD4213).
- Network card: FDR InfiniBand ConnectX-3 Mellanox AXX1FDRIBIOM (FDR 56GT/S).
- Coprocessor: Intel Xeon Phi Coprocessor 7120P.
- Software: Intel Parallel Studio XE 2016 Cluster Edition for Linux (Intel C++ Compiler, Intel Math Kernel Library, Intel OpenMP).

The algorithms were implemented with the use of the C language and with the use of the double precision. Our codes were compiled by INTEL C Compiler (icc) with the optimization flag `-O3` and with the cross-compiling option `-mmic`. Additionally, all algorithms were linked with the OpenMP library. To run the native executable on the co-processor, the `micnativeloadex` command was used. We set the number of threads using the environment variable `OMP_NUM_THREADS`.

B. The run-time

All the processing times are reported in seconds. The time was measured with the OpenMP function `open_get_wtime()`. They were tested in the double precision.

In Figures 6 and 7 we have compared the average running time of the four versions of the *outer* strategy on Intel Xeon Phi.

Figure 6 shows the dependence of the time on the number of threads for the matrix of the size 12000 on Intel Xeon Phi.

Figure 7 shows the dependence of the time on the matrix size for 240 threads on Intel Xeon Phi.

In Figures 8 and 9 we have compared the average running time of the six versions *split* strategy on Intel Xeon Phi.

```
for (k=1; k<=n/2-1; k++) {
  k2=n-k+1;
  det=a[k2,k]*a[k,k2]-a[k,k]*a[k2,k2];
#pragma omp parallel for  default(none) private(i) shared(k2,k,w,n,a,det) \
schedule(static/dynamic)
   for (i=k+1; i<=(k2-1); i++) {
     w[i,k]=(a[k2,k]*a[i,k2]-a[k2,k2]*a[i,k])/det;
     w[i,k2]=(a[k,k2]*a[i,k]-a[k,k]*a[i,k2])/det;
    }
#pragma omp parallel for default(none) shared(k2,k,a,w) private(i,j) \
schedule(static/dynamic)
   for (i=k+1; i<=(k2-1); i++) {
#pragma simd
     for (j=k+1; j<=k2-1; j++)
       a[i,j]=a[i,j]-w[i,k]*a[k,j]-w[i,k2]*a[k2,j];
   }
}//for k
```

Fig. 5. A fragment of the WZ factorization algorithm — the split strategy with the static or dynamic scheduler with the parallelization of the first loop and the second loop and with the vectorization of the second outer loop and with vectorization of the inner loop

Figure 8 shows the dependence of the time on the number of threads for the matrix of the size 12000 on Intel Xeon Phi.

Figure 9 shows the dependence of the time on the matrix size for 240 threads on Intel Xeon Phi .

Figure 10 shows the dependence of the time on the number of threads for the matrix of the size 12000 on Intel Xeon Phi comparing the best of the *outer* and *split* strategies.

Figure 11 shows the dependence of the time on the matrix size for 240 threads on Intel Xeon Phi comparing the best of the *outer* and *split* strategies.

Using obtained results we conclude that:

- The *outer* strategy gives better results without the vectorization.
- For the *split* strategy the best results were obtained for the *static+forsimd* version (Figure 4).
- The *split* strategy achieves better execution time than the *outer* strategy.
- The worse execution time was achieved for the *outer* strategy in the *static+simd* version (Figure 2).
- The choice of the scheduler usually makes small differences in the execution time.
- If the size of the matrix is increased, then the runtime is increased too and it may become more profitable to use a big number of threads.
- Our approaches (both *outer* and *split* strategies) are scalable to a large number of threads.

C. The performance

Figures 12 and 13 compare the performance (in Gflops) results obtained for both the *outer* and the *split* strategies — in double precision on Intel Xeon Phi. The performance is based on the number of floating-point operations in the WZ factorization, namely:

$$\sum_{k=1}^{\frac{n}{2}-1}\left(3+\sum_{i=k+1}^{n-k}\left(8+\sum_{j=k+1}^{n-k}4\right)\right)=\frac{4n^3-7n-18}{6}.$$

Figure 12 shows the dependence of the performance (of the best algorithms of both strategies) on the size of the matrix for 240 threads.

Figure 13 shows the dependence of the performance (of the best algorithms of both strategies) on the number of threads for the matrix size of 12000.

We can see that the best performance (about 11 Gflops) is achieved by the *split* strategy for the matrix of the size 12000 for 240 threads, and worst (less than 2 Gflops) is for the *outer* strategy for the smallest sizes. The performance increases fast with the growth of the number of threads.

V. Conclusion

In this paper we examined several practical aspects of the nested parallel loop execution on Intel Xeon Phi. Our approach exploits the thread-level and SIMD parallelism of the Intel Xeon Phi coprocessor. We used different strategies for executing nested parallel loops on the examples of the WZ factorization.

Both the *outer* and the *split* algorithms exploited the available number of threads. The *split* strategy achieves the best performance. The performance of 11 Gflops was achieved for 240 threads on Intel Xeon Phi. The implementation of the *split* strategy presented in this paper achieves high performance results, which has a direct impact on the solution of linear systems. Using the split strategy can help programmers with loop parallelization in multithreading environments.

This paper is another example of the successful use of OpenMP for solving scientific applications on Intel Xeon Phi.

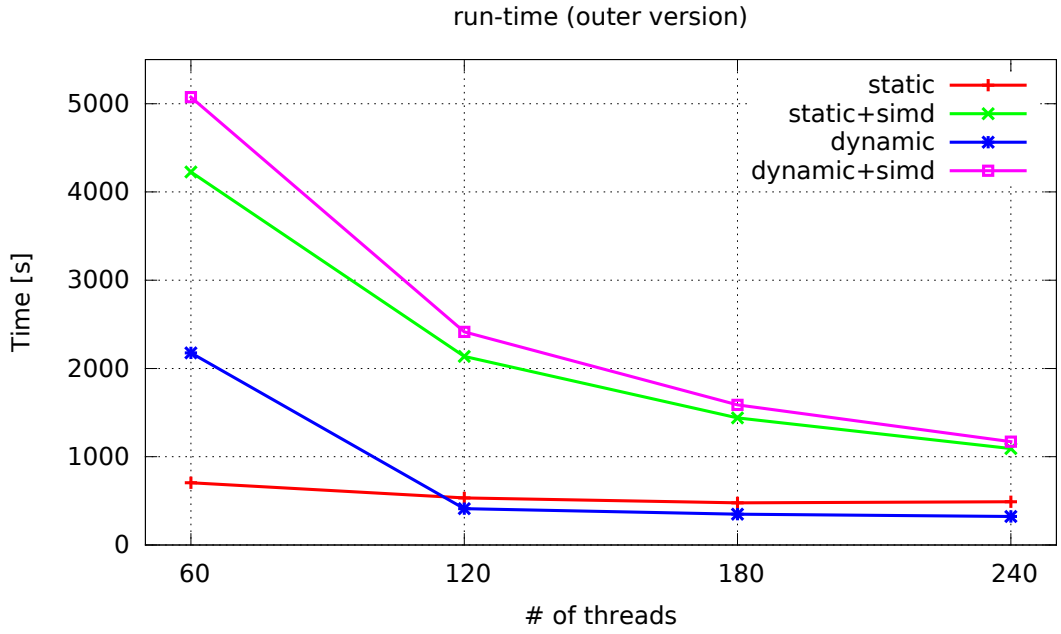


Fig. 6. The average running time of the WZ matrix decomposition as a function of the number of threads — for the algorithms using the double precision on Intel Xeon Phi for the matrix of the size 12000.

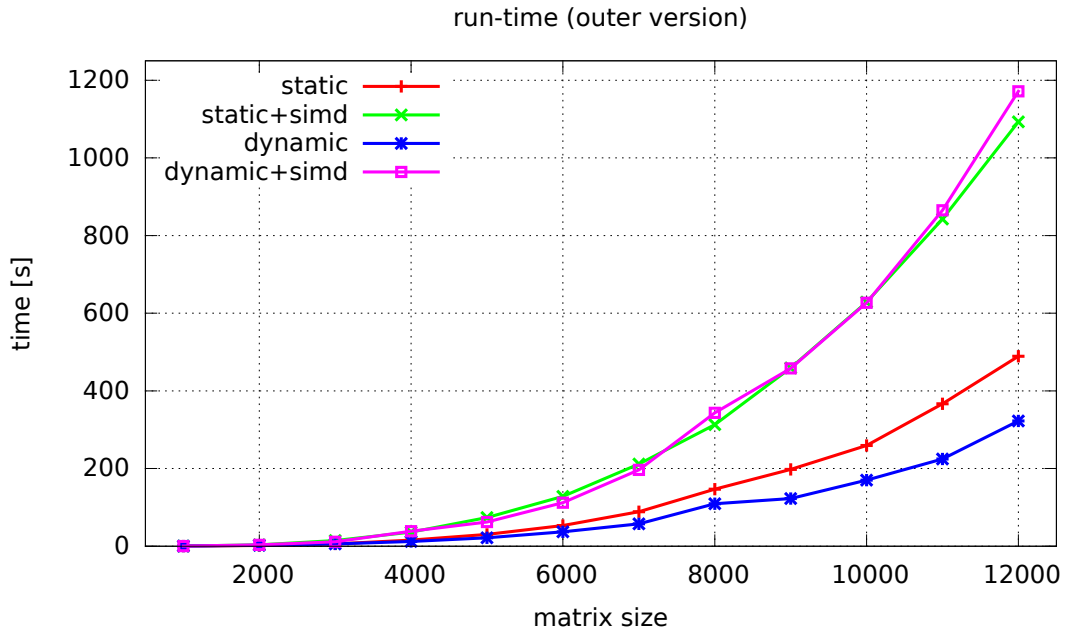


Fig. 7. The average running time of the WZ matrix decomposition as a function of the matrix size — for 240 threads on Intel Xeon Phi

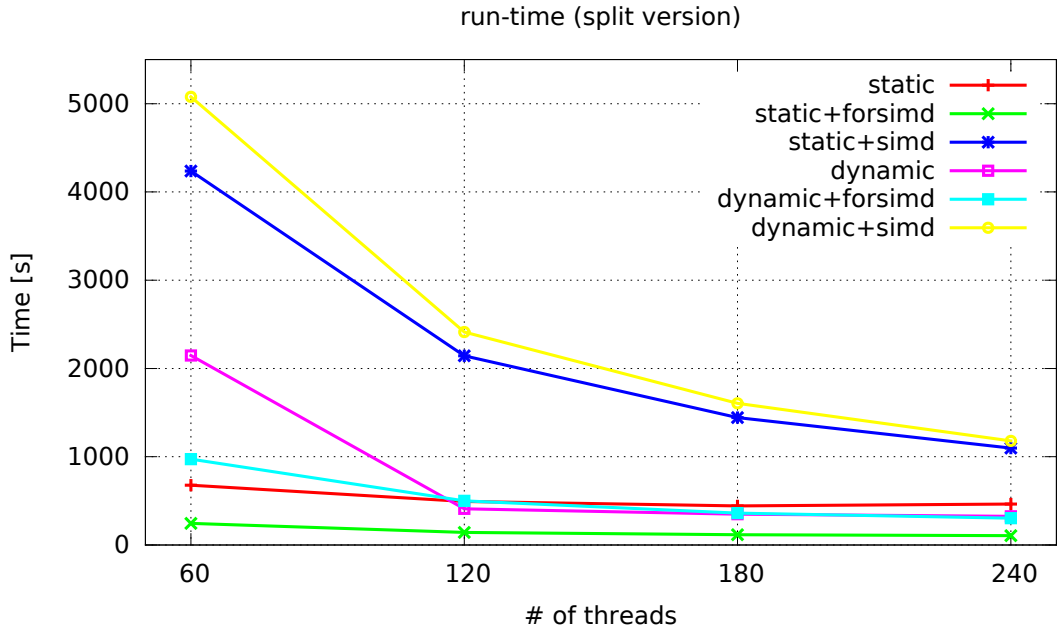


Fig. 8. The average running time of the WZ matrix decomposition as a function of the number of threads — for the *split* strategy using the double precision on Intel Xeon Phi for the matrix of the size 12000.

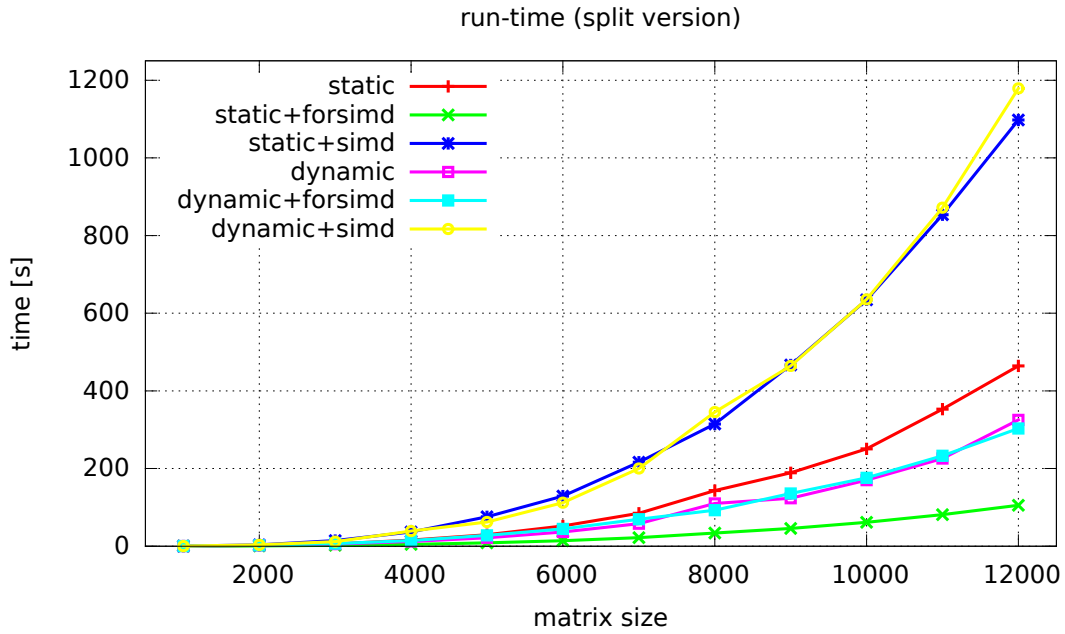


Fig. 9. The average running time of the WZ matrix decomposition as a function of the matrix size — for the *split* strategy for 240 threads on Intel Xeon Phi

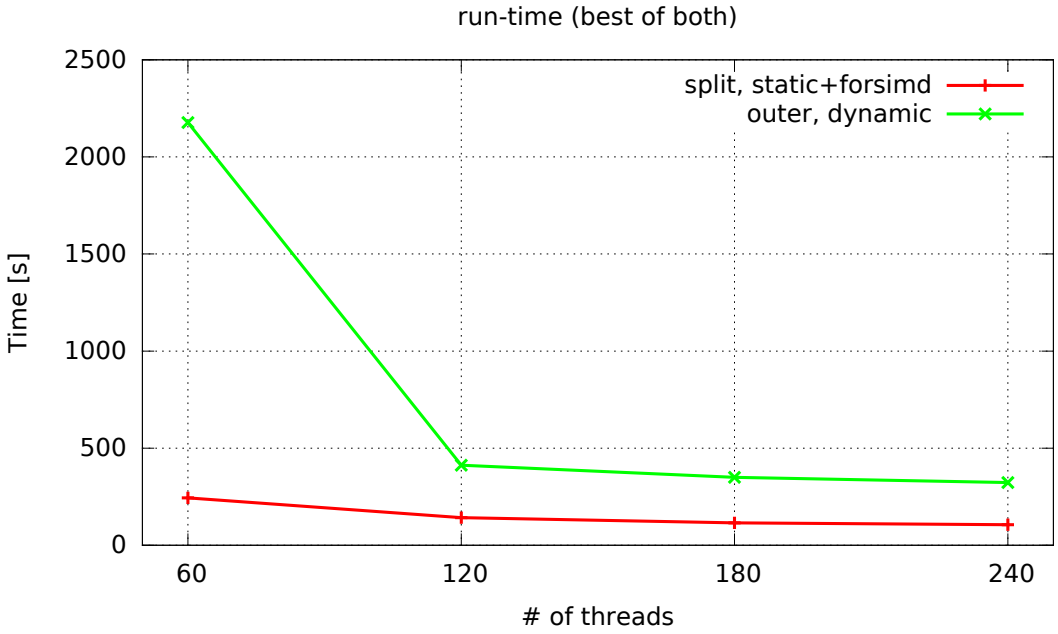


Fig. 10. The average running time of the WZ matrix decomposition as a function of the number of threads — the best of the *split* and *outer* strategies for the matrix of the size 12000

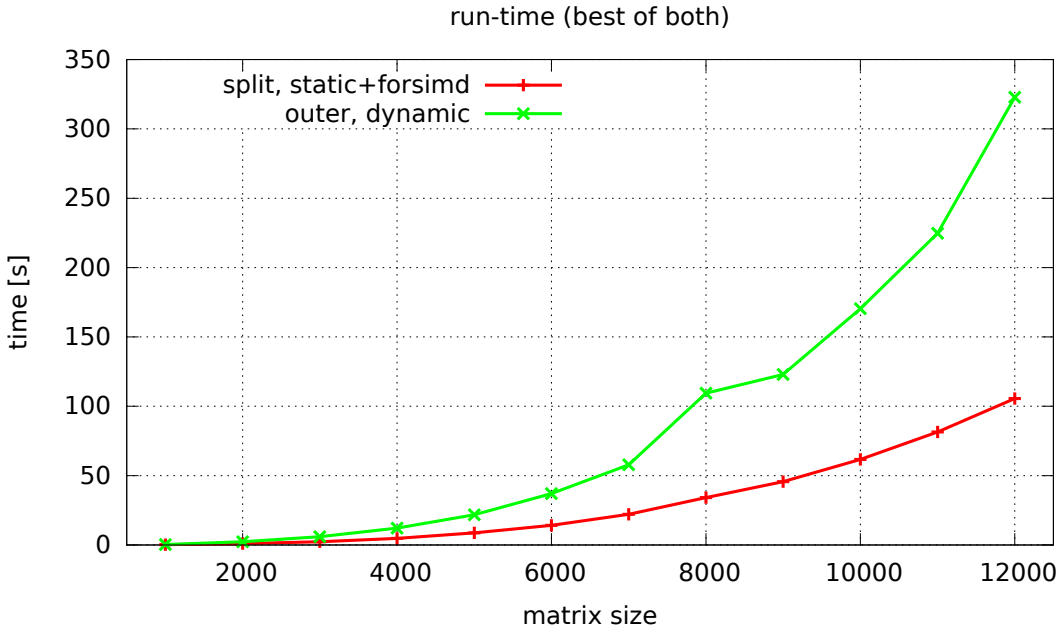


Fig. 11. The average running time of the WZ matrix decomposition as a function of the matrix size — the best of the *split* and *outer* strategies for 240 threads

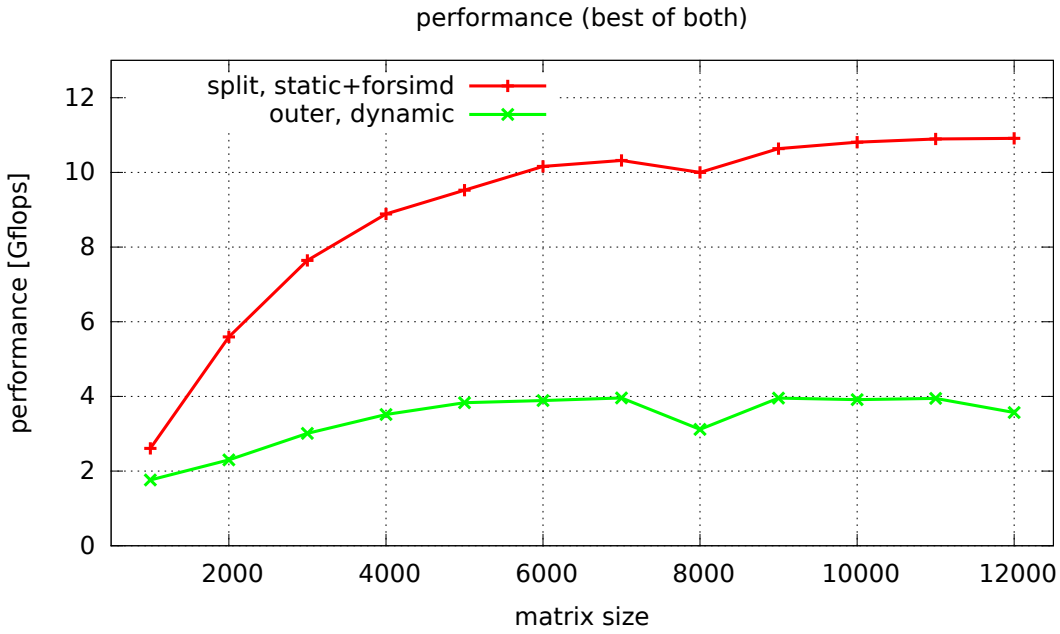


Fig. 12. The performance results for nested loops — for 240 threads for the best of both strategies as a function of the matrix size

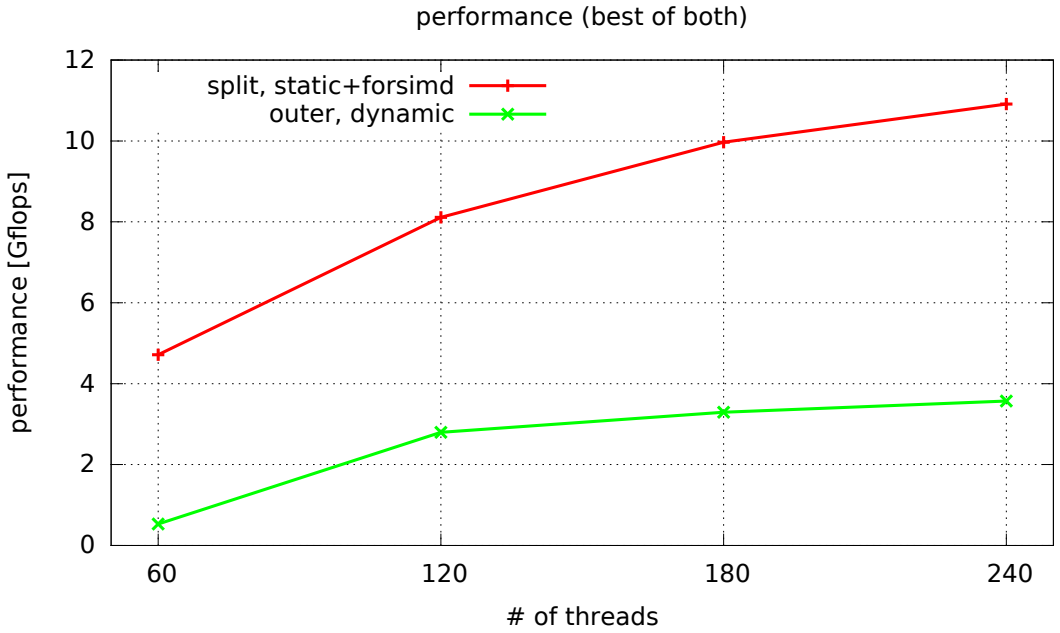


Fig. 13. The performance results for nested loops — for the matrix of the size 12000 for the best of both strategies as the function of number of threads

This paper is also example that the vectorization is not always the only key to achieving high performance on Intel Xeon Phi architecture.

The achieved results provide the basis for the further research on optimization of nested loops using the loop tiling technique in such a way that a data element used once is reused as soon as possible. Another field of further research is the use of thread affinity strategies which allow improving the performance and the scalability of our solution.

REFERENCES

[1] R. Blikberg, T. Sørevik, "Load balancing and OpenMP implementation of nested parallelism", *Parallel Computing* 31, Elsevier, 2005, pp. 984–998.

[2] B. Bylina, J. Bylina, "Strategies of parallelizing nested loops on the multicore architectures on the example of the WZ factorization for the dense matrices", Proceedings of the Federated Conference on Computer Science and Information Systems, Annals of Computer Science and Information Systems 5, 2015.

[3] S. Chandra Sekhara Rao, "Existence and uniqueness of WZ factorization", *Parallel Computing* 23, (1997), pp. 1129–1139.

[4] T. Cramer, D. Schmidl, M. Klemm, D. Mey, "OpenMP programming on Intel Xeon Phi coprocessors: An early performance comparison, Proc. of the Many-core Applications Research Community Symposium at RWTH Aachen University, (2012), pp. 38–44.

[5] A. Duran, R. Silvera, J. Corbalan, J. Labarta, "Runtime adjustment of parallel nested loops", *Proceedings of the 5th international conference on OpenMP Applications and Tools: shared Memory Parallel Programming with OpenMP*, Houston, 2004, pp. 137–147.

[6] D. J. Evans, M. Hatzopoulos, "The parallel solution of linear system", *Int. J. Comp. Math.* **7** (1979), pp. 227–238.

[7] A. Jackson, O. Agathokleous, "Dynamic Loop Parallelisation", arXiv: 1205.2367v1, 10 May 2012.

[8] J. Jeffers,J. Reinders, "Intel Xeon Phi Coprocessor High Performance Programming", Morgan Kaufmann Publishers Inc, 2013.

[9] R. Rahman, "Intel Xeon Phi Coprocessor Architecture and Tools: The Guide for Application Developers",Apress , Berkely, USA, 2013.

[10] A. Sadun, W. W. Hwu: "Executing nested parallel loops on shared-memeory multiprocessors", *Proceedings of the 21st Annual International Conference on Parallel Processing*, 1992.

[11] P. Yalamov, D. J. Evans: "The WZ matrix factorization method", *Parallel Computing* 21, 1995, pp. 1111–1120.

[12] OpenMP, http://openmp.org/wp/, April 2015.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 665–668
DOI: 10.15439/2016F345
ACSIS, Vol. 8. ISSN 2300-5963

Data Structures for Markov Chain Transition Matrices on Intel Xeon Phi

Beata Bylina, Joanna Potiopa
Department of Computer Science, Maria Curie-Skłodowska University,
Plac M. Curie-Skłodowskiej 1, 20-031 Lublin, Poland
Email: {beatas, joannap}@hektor.umcs.lublin.pl

***Abstract*—We employ Intel Xeon Phi as a high-performance coprocessor to solve Markov chains. Matrices arising from Markov models are very sparse with short rows. In this paper, the authors research two storage formats of Markov chain transition matrices on Intel Xeon Phi. In this work CSR and HYB (modification ELL) formats for such matrices are studied. Numerical experiments results for transition matrices of Markov chains from wireless networks and call-center models show that HYB format in offload version is more effective than CSR format. The obtained performance for HYB format is even `1.45` times better in comparison to multi-threaded CPU (dual Intel Xeon E5-2670) with the use of the CSR format (SpMV from the MKL library on CPU).**

I. Introduction

MARKOV chains are a tool for modeling various natural complex systems as well as computer systems and networks. Lately, they have been used to model wireless networks [3], [5], [6] and they often appear in computational biology [10] as well as in modeling call-centers [9].

In Markov modeling the models are very large because of exponential explosion of the states number, which happens due to the fact that complex systems usually consist of a certain number of subsystems and the states' space size of the whole complex system is usually exponentially dependent on the number of subsystems.

Any Markov chain can be described in terms of linear algebra with the use of a square matrix. A transition rate matrix $\mathbf{Q}$ (describing a Markov chain which models a system or a phenomenon) has some particular properties. It is a huge one and very sparse (with short rows). Sparse matrices are stored in special data structures and special algorithms are used to process these structures optimally. We can find descriptions of many such storage schemes in the literature (e.g. [14], [4], [2]).

The problem of efficiency of the sparse matrix-vector multiplication operation (SpMV) on Intel Xeon Phi was considered in [13], [12], [8]. In paper [13], the performance of the Intel Xeon Phi coprocessor for SpMV is investigated. One of the studied aspects in this work is CSR format for the sparse matrices. The authors showed that this format is not suited for Intel Xeon Phi for very sparse matrix (with short rows) in particular. The use of OpenMP based parallelization on Intel MIC (Intel Many Integrated Core Architecture) was evaluated in [8]. An efficient implementation of SpMV on the Intel Xeon Phi coprocessor by using a specialized ELLPACK-based format with load balancing is described in [12]. This implementation outperforms the implementation using CSR format even for matrices with very short rows.

The aim of this work is to shorten the computation time for the transition matrices from Markovian models of complex systems on Intel Xeon Phi by way of application of two data structures to store sparse matrices. Namely, CSR format will be used as in the works of [13], [8] and HYB format (as ELLPACK modification) similar as in the work [12]. HYB format was analyzed on GPU [7] and was much more effective than CSR for Markov chains.

Sparse matrix-vector multiplication (SpMV) operation on Intel Xeon Phi is implemented for these formats using the thread-level and the SIMD parallelism. Next, these SpMV's implementations are employed to the explicit fourth-order Runge-Kutta method. The numerical experiments were conducted for two groups of transition rate matrices, namely for a model of a call-center and a model of a wireless network on Intel Xeon Phi. A comparative analysis is also done with the CSR format from the Intel MKL library (Intel Math Kernel Library) on multithread CPU (dual Intel Xeon E5-2695).

The structure of the article is the following. Section II gives characteristics of the sparse matrix storage, chiefly CSR and HYB formats. Sections III and IV contain a description of the explicit fourth-order Runge-Kutta method and some details of its implementation on Intel Xeon Phi in particulary sparse matrix-vector multiplication operation (SpMV). Section V analyzes Intel Xeon Phi's performance on this kernel for two data structures (CSR and HYB). Section VI presents conclusions.

II. Storage of a sparse matrix

In the literature [14], a lot of ways which represent sparse matrices and enable their effective storage and processing have been suggested.

One of the formats to store any sparse matrices is *Compressed Sparse Row* (CSR). The operations on matrices stored in this format are part of Intel MKL library in version on Intel MIC architecture [11]. In this format, the information about $\mathbf{A}$ matrix, where $\mathbf{A}$ is a sparse matrix of $m \times n$ size and nz nonzero elements, is stored in three one-dimentional arrays:

- $data[\cdot]$, of nz size stores values of nonzero elements (in increasing order of row indices);
- $col[\cdot]$, of nz size stores column indices of nonzero elements (in order conforming to the data array content);

978-83-60810-90-3/$25.00

- $ptr[\cdot]$, of $m+1$ size, stores indices of beginnings of successive rows in $data$ array — that is $data[ptr[i]]$ is the first nonzero element of i-th row in $data$ array, similar, $col[ptr[i]]$ is the column number of this element.

Hybrid format (HYB) comprises of two other formats of sparse matrix storage in the memory disregarding its weaknesses and simultaneously making use of its advantages: COO format (*Coordinate format*) and ELLPACK package format (ELL).

ELLPACK [1] package format (ELL) is a sparse matrix format which is helpful in vector architecture. It is useful for matrices in which the number of the elements in almost every row is the same, especially when many rows reach maximum length in a given matrix or approach it, it becomes useless when the number of elements in a row is dispersed, e.g. when there are many rows which are longer than the mean. In this format the sparse matrix is stored in two two-dimensional arrays.

In HYB format the matrix is stored in two two-dimensional arrays (ELL) and three one-dimensional arrays (COO):

- $ell\_data[\cdot]$ stores values of nonzero elements as two-dimensional rectangular array of $M \times MNNZ$ size, where M is the number of rows in the matrix and $MNNZ$ denotes the mode of number of nonzero elements in a row. The rows with fewer nonzero elements than $MNNZ$ are aligned to the left and filled with zero (meaningless) values in the remaining part, while the longer rows are cut off.
- $ell\_indices[\cdot]$ stores column indices of a matrix elements placed appropriately in $ell\_data[\cdot]$. The size and the structure of this array is the same as $ell\_data[\cdot]$.
- $coo\_data[\cdot]$ stores nonzero elements values, which were cut off from $ell\_data[\cdot]$.
- $coo\_col[\cdot]$ stores column indices of nonzero elements from $coo\_data[\cdot]$ (in the same order as $coo\_data[\cdot]$).
- $coo\_row[\cdot]$ stores row indices of nonzero elements from $coo\_data[\cdot]$ (in the same order as $coo\_data[\cdot]$).

III. Parallel Runge-Kutta algorithm

General form explicit fourth-order Runge-Kutta method in parallel version is presented as Algorithm 1.

The implementation of the Algorithm 1 contains our implementation of SpMV operation and vector addition. We use the OpenMP standard and the `for` directives to parallelize all operations. We use a `static` scheduler for the distribution of the matrix rows and the values of vector. The SpMV operation is a simple task to assign a row block to a single thread in a parallel execution. The idea of vectorization is to process all the nonzero elements in row at once. Since the Intel Xeon Phi architecture has 32 512-bit registers, the matrices should have at least 8 values in each row to fully utilize the register. For one-row block we use a pragma compilator `#pragma simd` which enforces vectorization of the inner loops. This vectorization is not effective because our matrices have got short rows — shorter than 8 elements (see table I).

Algorithm 1 The parallel algorithm which determines the transient probabilities vector, where $*$ operation denotes parallel sparse matrix-vector multiplication and $+$ operation denotes parallelized and vectorized vector addition

Require: Q^T — transition rate matrix, pi_0 — initial probability vector, h — step, t — time
Ensure: vector of transient probabilities pi_t in the time t

$lk \leftarrow t/h$
$pi_t \leftarrow pi_0$
for $k = 1$ to lk **do**
  $k_1 \leftarrow Q^T * pi_t$
  $k_2 \leftarrow Q^T * (pi_t + \frac{h}{2} k_1)$
  $k_3 \leftarrow Q^T * (pi_t + \frac{h}{2} k_2)$
  $k_4 \leftarrow Q^T * (pi_t + h k_3)$
  $pi_t = pi_t + \frac{h}{6} \cdot (k_1 + 2k_2 + 2k_3 + k_4)$
end for
return pi_t

IV. Numerical experiment

In this section we tested the time, the speedup and the performance of the explicit fourth-order Runge-Kutta method (RK4). The programs were implemented in C++ language and three implementations of this algorithm were created:

- MKL-CSR version — it is a version using parallelism and vectorization offered by the function of the Intel MKL library in the version of Intel MIC architecture, where the sparse matrix was stored in CSR format.
- CSR version — it is a version, where the sparse matrix was stored in CSR format; all vector and matrix operations were implemented by the authors.
- HYB version — it is a version, where the sparse matrix was stored in HYB format; all vector and matrix operations were implemented by the authors.

The impact of the program execution mode (native and offload) and the various numbers of threads were tested.

For each version, the program was compiled by using the Intel C++ compiler (`icc`) with a compiler flag `-O3`, which resulted in automatic computing vectorization, `-openmp`, `-mmic` (enabling the cross compilation needed for a native execution on Intel Xeon Phi). Additionally, during development, we used the flag `-vec-report2` to verify whether the RK4 kernel was successfully vectorized. In every case, alignment of the memory data was used as vectorization support; the data were aligned with 64 bytes limit, which was recommended by the documentation. `-mkl` option was also used to allow introducion of parallelism in MKL-CSR version. Intel MKL library was applied to measure the elapsed time.

In the table I, the properties of matrices used during the test are given: WF1, WF2 describe wireless networks, CC1, CC2 describe the call-center.

The matrices we tested are very sparse, however, the pattern varies in dependence on the model (table I). L_i, $1 \leq i \leq n$, denoted the number of nonzero elements per row. CC matries

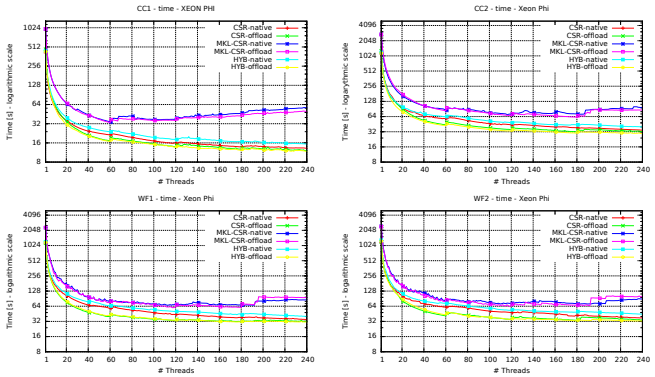


Fig. 1. Runtime of explicit fourth-order Runge-Kutta method on Intel Xeon Phi for CC1, CC2, WF1 and WF2 matrices

have similar number of nonzero elements per row (3—6). WF matrices have different number of elements per row.

TABLE I
THE PROPERTIES OF THE TESTED MATRICES

No	Name	n	nz	$\frac{nz}{n}$	$\min L_i$	$\max L_i$
1.	CC1	335421	1996701	5.95	3	6
2.	CC2	937728	5588932	5.56	3	6
3.	WF1	962336	4434326	4.61	1	12
4.	WF2	1034273	4660479	4.51	1	11

All tested matrices have very short rows; the mean number of elements in a row is between 4.51 and 5.95 elements. The input arrays size is long enough and we provide enough work for each thread.

The tests were carried out using computing node of the following parameters:

- Platform: Intel Server Chassis R2000WTXXX, Intel Server Board S2600WT2
- CPU: 2xIntel Xeon E5-2670 v3 (2x12 cores, 2.3 GHz)
- Memory: 128 GB DDR4 2133MT/s (8xCrucial CT16G4RFD4213)
- Network card: FDR InfiniBand ConnectX-3 Mellanox AXX1FDRIBIOM (FDR 56GT/S)
- Coprocessor: Intel Xeon Phi Coprocessor 7120P (16GB, 1.238 GHz, 61 cores)
- Software: Intel Parallel Studio XE 2016 Cluster Edition for Linux (Intel C++ Compiler, Intel Math Kernel Library, Intel OpenMP)

V. RESULTS

In this section we evaluate the time, the speedup and the performance of our approach to the problem of the different ways of sparse matrices storage, two execution modes for various numbers of threads.

A. Time

Fig. 1 presents execution time of RK4 algorithm. It is obvious that independently of the models and matrices size, the version using MKL library routines is the slowest. In case of our implementations offload versions are the fastest but there is no clear difference between CSR and HYB formats. Our implementations (in offload version) always perform about two times faster than MKL-CSR version, for a similar number of threads (except CC1 matrix, when execution time for 240 threads in HYB-offload version and CSR-offload is even four times faster than MKL-CSR-offload version).

For every solution, the time for the first 60 threads decreases most rapidly (with 1 thread per core activated). The gain with the use of a large number of threads per core is meaningless.

We obtain the best time for each matrix for HYB-offload version. Basing on the received charts, it seems that for MKL-CSR version the size matrix is essential; for CC1 matrix execution time increases with 60 threads and for the remaining matrices with 3 times bigger size, the time increases with 180 threads. In case of our implementations, execution time is even independent of the matrix size and model.

B. Speedup

Fig. 2 shows the speedup of RK4 method on Intel Xeon Phi with respect to a sequential version running on one thread of Intel Xeon Phi. The speedup for CC1 matrix gives a bit different charts in comparison with other matrices. It is due to a small matrix size in relation to others. For the number of threads from 1 to 60, HYB-native and CSR-native implementations give the lowest speedup, for other version the results are similar.

For more than 60 threads we can see that the speedup of the MKL-CSR version decreases. Moreover, for over 140 threads it gives the poorest results. With 60 to 240 threads (2-4 threads per core) we can see that CSR-offload and HYB-offload perform the best (with minimal superiority of the first implementation).

CC1 matrix achieves maximum speedup (which is 40) for CSR-offload version with 240 threads. The other matrices (CC1, WF1, WF2) have similar sizes and their speedup charts look similar. The lowest speedup is obtained for the CSR-native and HYB-native versions with 1 to 180 threads (1-3 threads per core).

The remaining implementations give similar results. The difference appears when we start over 180 threads (4 threads per core). HYB-offload and CSR-offload have the best speedup while MKL-CSR (native and offload version) clearly decrease. The best achieved speedup is 44 for CC2 matrix in MKL-CSR-offload version with 180 threads. For WF matrices the lowest speedup is 39 for WF1 and almost 40 for WF2 in MKL-CSR-offload version.

C. Performance

Fig. 3 presents the performance of RK4. Each of the figure's bars shows maximum performance obtained for a given stored matrix in a given format and for the program executed in the given mode. In comparison we also present the performance of RK4 method on CPU where all matrix and vector operations were realized with the use of the kernels from MKL library (denoted MKL-CSR-CPU).

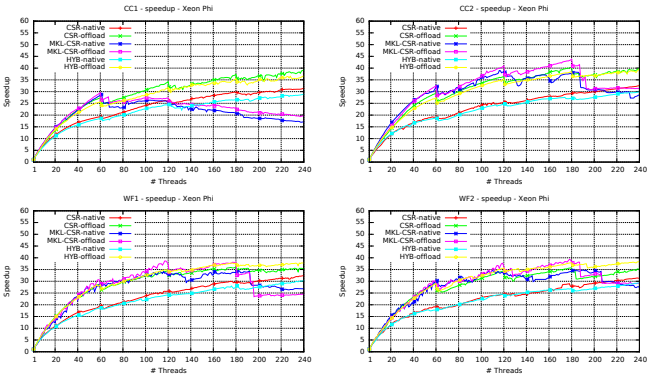


Fig. 2. Speedup of explicit fourth-order Runge-Kutta method on Intel Xeon Phi with respect to a sequential version running on one thread Intel Xeon Phi for CC1, CC2, WF1 and WF2 matrices

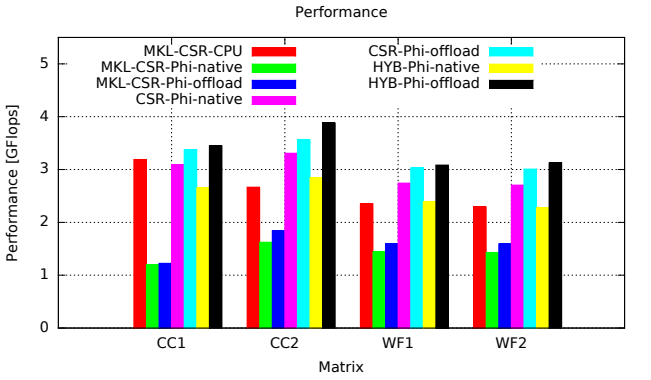


Fig. 3. Performance of explicit fourth-order Runge-Kutta method on Intel Xeon Phi

The results analysis shows a clear difference between the performance of our implementations as opposed to MKL-CSR version on Intel Xeon Phi. We can also notice that independently of the matrix size and storage format (MKL-CSR, CSR, HYB) we obtained better performance when starting the application in offload version. The biggest differences between native and offload modes occurred for HYB format. In case of our implementations we achieved the best results for every matrix in HYB-Phi-offload version. Moreover it was from 1.9 to 2.8 times more efficient than MKL-CSR-Phi-offload version due to the fact that our implementations is less general and enables better control over multithreading and vectorization.

The matrices generated for call-center model achieve better performance than the matrix from wireless network models. Despite the size, is connected with slightly higher matrix density and more regular pattern (tab. I).

Our implementations also have better performance in relation to MKL-CSR-CPU; with small matrix CC1 the differences become insignificant, but for the bigger matrices our approach is clearly favourable even up to 50% for HYB-Phi-offload version.

VI. Conclusion

In this article we investigated the use of two sparse matrices format storage in context of Markov chain problems for accelerating on the Intel Xeon Phi. Our approach exploits the thread-level parallelism and vectorization for SpMV operation and the thread-level parallelism and vectorization for the vector addition.

Based on the conducted experiments, we can clearly state that the Intel MKL library for Intel MIC architecture performed worse than our own CSR and HYB implementations.

Our implementations significantly outperform the optimized implementation routine SpMV from Intel MKL library using the CSR format on Intel Xeon Phi. The CSR and HYB versions are scalable to a large number of threads and they use all the cores on Intel Xeon Phi. We achieve the best performance for HYB format offload version.

Our implementation can still be improved. In future works we will employ thread affinity strategies which allowe to improve the performance and scalability of our approach.

References

[1] ELLPACK, 2014. http://www.cs.purdue.edu/ellpack.

[2] N. Bell and M. Garland. Efficient sparse matrix-vector multiplication on CUDA. Technical report, NVIDIA, 2008. Tech. Report No. NVR-2008-004.

[3] G. Bianchi. Performance analysis of the IEEE 802.11 distributed coordination function. *IEEE Journal on Selected Areas in Communications*, 18(3):535–547, March 2000.

[4] B. Bylina, J. Bylina, and M. Karwacki. Computational aspects of GPU-accelerated sparse matrix-vector multiplication for solving Markov models. *Theoretical and Applied Informatics*, 23(2):127–145, 2011.

[5] J. Bylina and B. Bylina. A Markovian queuing model of a WLAN node. *Communications in Computer and Information Science*, 160:80–86, 2011.

[6] J. Bylina, B. Bylina, and M. Karwacki. A Markovian model of a network of two wireless devices. *Communications in Computer and Information Science*, 291:411–420, 2012.

[7] Jarosław Bylina, Beata Bylina, and Marek Karwacki. An efficient representation on GPU for transition rate matrices for Markov chains. In *Parallel Processing and Applied Mathematics*, pages 663–672. Springer, 2013.

[8] Tim Cramer, Dirk Schmidl, Michael Klemm, and Dieter an Mey. OpenMP programming on Intel Xeon Phi coprocessors: An early performance comparison. In *Proc. of the Many-core Applications Research Community Symposium at RWTH Aachen University*, pages 38–44, 2012.

[9] N. Gans, G. Koole, and A. Mandelbaum. Telephone call centers: tutorial. *Review and Research Prospects, Manufact. and Service Oper. Manag.*, 5:79–141, 2003.

[10] N. A. Hamilton, K. Burrage, and A. Bustamam. Fast parallel markov clustering in bioinformatics using massively parallel computing on gpu with cuda and ellpack-r sparse format. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 9(3):679–692, 2012.

[11] Intel. Intel Math Kernel Library (MKL). http://software.intel.com/en-us/intel-mkl, 2014.

[12] Xing Liu, Mikhail Smelyanskiy, Edmond Chow, and Pradeep Dubey. Efficient sparse matrix-vector multiplication on x86-based many-core processors. In *Proceedings of the 27th International ACM Conference on International Conference on Supercomputing*, ICS '13, pages 273–282, New York, NY, USA, 2013. ACM.

[13] Erik Saule, Kamer Kaya, and Ümit V. Çatalyürek. Performance evaluation of sparse matrix multiplication kernels on Intel Xeon Phi. *CoRR*, abs/1302.1078, 2013.

[14] W. J. Stewart. *An Introduction to the Numerical Solution of Markov Chains*. Princeton University Press, Princeton, NJ, 1994.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 669–672
DOI: 10.15439/2016F88
ACSIS, Vol. 8. ISSN 2300-5963

Block Subspace Projection PCG Method for Solution of Natural Vibration Problem in Structural Analysis

Sergiy Fialko
Tadeusz Kościuszko Cracow University of Technology
ul. Warszawska 24 St., 31-155 Kraków, Poland
Email: sergiy.fialko@gmail.com

Filip Żegleń
Tadeusz Kościuszko Cracow University of Technology
ul. Warszawska 24 St., 31-155 Kraków, Poland
Email: filipzeglen@hotmail.com

***Abstract*—The block subspace projection preconditioned conjugate gradient method for analysis of natural vibration frequencies and modes applying to large problems of structural mechanics is proposed. It is oriented at the usage in finite element analysis software operated on multi-core desktop computers with restricted amount of core memory as an alternative approach to widespread block Lanczos method and subspace iteration method. We focused our attention on achievement of high computational stability and parallelization of proposed algorithm. The solution of real-life large problems confirms the reliability of proposed approach.**

I. Introduction

THE block Lanczos method as well as various versions of subspace iteration method widely is used for extraction of natural vibration frequencies and modes in modern engineering software applying to the problems of structural mechanics

$$\mathbf{K}\mathbf{v}_i - \lambda_i \mathbf{M}\mathbf{v}_i = 0 \ , \tag{1}$$

where **K** and **M** are the sparse symmetric stiffness and mass matrices arising when the finite element method is applied to problems of structural mechanics, $\{\lambda_i, \mathbf{v}_i\}$ – eigenpair for i-th mode, $i \in [1, n]$, $n << N$, n – number of required eigenmodes, N – dimension of problem (1).

The Lanczos method as well as subspace iteration (SI) approach produces the inverse iterations

$$\mathbf{K}\mathbf{v}_i^{k+1} = \mathbf{M}\mathbf{v}_i^k \ , \tag{2}$$

where $\mathbf{v}_i^k$ is an approximation of eigenvector $\mathbf{v}_i$ on iteration step k (k = 1, 2, … until converges). For large problems solved on desktop and laptop computers with restricted amount of core memory the lower triangular matrix **L** of factorized stiffness matrix $\mathbf{K} = \mathbf{L}\cdot\mathbf{D}\cdot\mathbf{L}^T$ is stored block-by-block on disk. Therefore, on each iteration step k mentioned above eigenvalue solvers must read twice the lower triangular matrix **L** from disk. Taking into account that size of matrix **L** for large design models (N = 3 000 000 – 6 000 000 equations) achieves 6 – 20 GB and more, performance of such eigenvalue solver drastically decreases.

Unlike mentioned Lanczos and SI approaches, the preconditioned conjugate gradient (PCG) method [8], [12] uses only RAM. However, for poorly conditioned problems, which are the most of problems of structural mechanics (see [7]), we have to construct an efficient preconditioning, because the conventional SSOR, symmetrical Gauss-Seidel, ICCG0 preconditioners result in unacceptable slow convergence. We found that the aggregation multilevel preconditioning [1], [2] and incomplete Cholesky factorization by value, realized in technique of sparse matrices [4], demonstrate a stable convergence for solution of linear equation sets (static analysis) as well as for extraction of natural vibrations frequencies and modes (modal analysis) for considered class of problems.

Achieving of stable convergence of the conjugate gradient method for solving the eigenvalue problem (1) is much more difficult than in solving systems of linear algebraic equations. Most likely, for this reason, in modern commercial FEA software mainly used eigenvalue solvers based on inverse matrix iteration (2) [9]. In present article, we propose the block subspace projection preconditioned conjugate gradient (BSPPCG) method for solution of problem (1).

II. Block Subspace Projection Preconditioned Conjugate Gradient Method

A. State of problem

To achieve the computational stability of PCG method at solution of poorly conditioned problems of structural mechanics, the aggregation multilevel preconditioning and shift technique have been used [1]. A little later, an aggregation multilevel preconditioning was replaced by an incomplete Cholesky factorization by value implemented in technique of sparse matrix [4].

Article [5] presents a block version of PCG method, where several vectors in block are iterated simultaneously. The Gram-Schmidt orthogonalization provides orthogonality of vectors in the block between themselves as well as their orthogonality to the eigenmodes, converged earlier. The shift technique is used for acceleration of convergence.

This work was not supported by any organization

A very interesting idea of local block PCG (LOBPCG) method was proposed in [10]. The approximations on the next iteration step are presented as:

$$\begin{cases} \mathbf{v}_i^{k+1} = \sum_{j=1}^{m} \alpha_j^k \mathbf{z}_j^k + \sum_{j=1}^{m} \tau_j^k \mathbf{v}_j^k + \sum_{j=1}^{m} \gamma_j^k \mathbf{p}_j^k \\ \mathbf{p}_i^{k+1} = \sum_{j=1}^{m} \alpha_j^k \mathbf{z}_j^k + \sum_{j=1}^{m} \gamma_j^k \mathbf{p}_j^k \end{cases}, \quad i \in [1, m], \quad (3)$$

where m is a dimension of subspace $span\{\mathbf{z}_1^k,\ldots, \mathbf{z}_m^k, \mathbf{v}_1^k,\ldots, \mathbf{v}_m^k, \mathbf{p}_1^k,\ldots, \mathbf{p}_m^k\}$, $\mathbf{z}_j^k = \mathbf{B}^{-1}\mathbf{r}_j^k$, $\mathbf{B}$ – preconditioning operator, $\mathbf{r}_j^k = \lambda_j^k \mathbf{M}\mathbf{v}_j^k - \mathbf{K}\mathbf{v}_j^k$ – residual vector, $\mathbf{p}_j^k$ – conjugate direction vector, $j = 1, \ldots, m$. The projection matrix $\mathbf{Q} = \{\mathbf{Z}\ \mathbf{V}\ \mathbf{P}\}$ has dimension $N \times 3{\cdot}m$ and consists of $N \times m$ submatrices $\mathbf{Z} = \{\mathbf{z}_1^k,\ldots, \mathbf{z}_m^k\}$, $\mathbf{V} = \{\mathbf{v}_1^k,\ldots, \mathbf{v}_m^k\}$ and $\mathbf{P} = \{\mathbf{p}_1^k,\ldots, \mathbf{p}_m^k\}$. In matrix form:

$$\begin{aligned} &\mathbf{V}^{k+1} = \mathbf{Q}\cdot\mathbf{q}, \quad \mathbf{q} = \{\mathbf{q}_1,\ldots,\mathbf{q}_m\}, \\ &\mathbf{q}_j = \{\boldsymbol{\alpha}_j^k\ \boldsymbol{\tau}_j^k\ \boldsymbol{\gamma}_j^k\}^T, \quad j \in [1, m] \end{aligned}, \quad (4)$$

where $\boldsymbol{\alpha}_j^k = \{\alpha_{1,j}^k, \ldots, \alpha_{m,j}^k\}$, $\boldsymbol{\tau}_j^k = \{\tau_{1,j}^k, \ldots, \tau_{m,j}^k\}$, $\boldsymbol{\gamma}_j^k = \{\gamma_{1,j}^k, \ldots, \gamma_{m,j}^k\}$ and iteration number k is omitted for matrices $\mathbf{Q}$, $\mathbf{q}$. Subscript j denotes a number of eigenmode in expansion (6).

Let us substitute (7) in (1) and multiple at left by $\mathbf{Q}^T$:

$$\mathbf{k}\mathbf{q} - \mathbf{m}\mathbf{q}\boldsymbol{\Lambda} = 0, \quad (5)$$

where $\mathbf{k} = \mathbf{Q}^T\mathbf{K}\mathbf{Q}$, $\mathbf{m} = \mathbf{Q}^T\mathbf{M}\mathbf{Q}$ and $\boldsymbol{\Lambda}$ is a diagonal matrix with approximations of eigenvalues on iteration step k. The approximation of eigenmodes $\mathbf{v}^{k+1}$ on the next iteration step follows from (4) after substitution of $\mathbf{q}$, which is obtained from solution of reduced eigenproblem (5). The conjugate direction vector is derived as

$$\mathbf{P}^{k+1} = \mathbf{Q}'^T\mathbf{q}', \quad (6)$$

where $\mathbf{Q}' = \{\mathbf{Z}\ \mathbf{P}\}$ and $\mathbf{q}' = \{\boldsymbol{\alpha}^k\ \boldsymbol{\gamma}^k\}^T$. Then, the Rayleigh quotient is used for approximation of eigenvalues on iteration step k+1 and evaluation of residual vectors $\mathbf{r}_j^k$ is produced after it.

The proposed approach in the form of [10] is little suitable to analysis of real-life problems of structural mechanics, since the columns in matrix $\mathbf{Q}$ become the linearly dependent as soon as the first eigenpair begins to converge. The authors of current article obtained the computational instability even for simple test problem – simply supported beam. The clear explanation of this fact is in [11]: the second expression in (3) is a linear combination between vectors $\mathbf{z}_j^k$ and $\mathbf{p}_j^k$. Therefore, the first expression in (3) is possible to rewrite as

$$\mathbf{v}_i^{k+1} = \sum_{j=1}^{m} \alpha_j^k \mathbf{z}_j^k + \sum_{j=1}^{m} \tau_j^k \mathbf{v}_j^k + \sum_{j=1}^{m} \beta_j^k \mathbf{v}_j^{k-1}, \quad i \in [1, m]. \quad (7)$$

For instance, let vector $\mathbf{v}_1$ begins to converge. Then, basis vectors $\mathbf{v}_1^k$ and $\mathbf{v}_1^{k-1}$ are almost linearly dependent, because $\mathbf{v}_1^{k-1}$, $\mathbf{v}_1^k$ as well as $\mathbf{v}_1^{k+1}$ tends to the same eigenvector $\mathbf{v}_1$ – an exact solution. In [11] for stabilization of computational process was suggested to remove the vectors $\mathbf{z}_j^k$, $\mathbf{v}_j^k$, $\mathbf{p}_j^k$ of subspace $span\{\mathbf{z}_1^k,\ldots, \mathbf{z}_m^k, \mathbf{v}_1^k,\ldots, \mathbf{v}_m^k, \mathbf{p}_1^k,\ldots, \mathbf{p}_m^k\}$, as soon as the corresponding vector $\mathbf{v}_j^k$ converges.

The drawbacks of such approach are:

- For some problems of structural mechanics it turns out that the iterative process is falling apart even before the error

$$err_i = \left\|\mathbf{r}_i^k\right\|_2 \Big/ \left[\lambda_i^k \left\|\left(\mathbf{v}_i^k\right)^T \mathbf{M}\mathbf{v}_i^k\right\|_2\right] \quad (8)$$

is still enough small ($err_j \le tol$) to recognize the convergence of vector $\mathbf{v}_j^k$.

- The dimension of subspace m must be not less than the number of required eigepairs n ($m \ge n$). Such an approach results in enormous computational efforts if it is required a large number of eigenpairs (n = 100 – 200 and more). In addition, it is required a significant amount of core memory for allocation of matrix $\mathbf{Q}$.

The proposed in given article eigenvalue block subspace projection PCG (BSPPCG) solver uses the idea (3) of LOBPCG, but possesses the high computational stability and requires essentially less computational efforts when large number of eigenpairs is required. That allows us to recommend this solver for use in FEA software intended for analysis of problems of structural mechanics on widespread desktop and laptop computers as well as on shared memory workstations.

B. Algorithm of BSPPCG method

1. Set $m \in [8, 32]$, $m\%np = 0$; prepare linearly independent start vectors $\mathbf{V}^0=\{\mathbf{v}_1^0,\ldots,\mathbf{v}_m^0\}$, $\mathbf{P}^0=\{\mathbf{p}_1^0,\ldots,\mathbf{p}_m^0\}$, $nconvmodes = 0$;
2. **do** $k = 1, 2, \ldots$, **until** $nconvmodes < n$.
3. **parallel loop for** $j = 1, \ldots, m$
   $(\mathbf{v}_j^k)^T{\cdot}\mathbf{M}{\cdot}\mathbf{v}_j^k = \mathbf{I}$ (normalization procedure)
   $\lambda_j^k = [(\mathbf{v}_j^k)^T\ \mathbf{K}\ \mathbf{v}_j^k]/[(\mathbf{v}_j^k)^T\ \mathbf{M}\ \mathbf{v}_j^k]$
   $\mathbf{r}_j^k = \lambda_j^k\mathbf{M}\mathbf{v}_j^k - \mathbf{K}\mathbf{v}_j^k$
   $\mathbf{B}\ \mathbf{z}_j^k = \mathbf{r}_j^k \rightarrow \mathbf{z}_j^k$
   end of parallel loop for
4. **do** j=1, m
   if($err_j \le tol$)
   $nconvmodes$++;
   put $\{\lambda_j, \mathbf{v}_j\}$ as final results,
   prepare new start vectors $\mathbf{v}_j^k$ and $\mathbf{p}_j^k$,
   orthogonalize $\mathbf{v}_j^k$ against converged modes
   and put to blocks $\mathbf{V}^k$, $\mathbf{P}^k$.
   $\lambda_j^k = [(\mathbf{v}_j^k)^T\ \mathbf{K}\ \mathbf{v}_j^k]/[\ [(\mathbf{v}_j^k)^T\ \mathbf{M}\ \mathbf{v}_j^k]$
   $\mathbf{r}_j^k = \lambda_j^k\mathbf{M}\mathbf{v}_j^k - \mathbf{K}\mathbf{v}_j^k$
   $\mathbf{B}\ \mathbf{z}_j^k = \mathbf{r}_j^k \rightarrow \mathbf{z}_j^k$, put $\mathbf{z}_j^k$ to $\mathbf{Z}^k$.

end if
end do
5. **paralel loop for** $s = 1, 3m$
loop for $p = s, 3m$ (s, p – columns of matrix $\mathbf{Q}$)
$\mathbf{m}_{sp} = \mathbf{Q}_p{}^T\mathbf{M}\mathbf{Q}_s$ – evaluation of matrix $\mathbf{m}$
end loop for
end of parallel loop for
6. **if**(Chol($\mathbf{m}$): $\mathbf{m} = \mathbf{l}\cdot\mathbf{l}^T$)
paralel loop for $s = 1, 3m$
loop for $p = s, 3m$ (s, p – columns of matrix $\mathbf{Q}$)
$\mathbf{k}_{sp} = \mathbf{Q}_p{}^T\mathbf{K}\mathbf{Q}_s$ – evaluation of matrix $\mathbf{k}$
end loop for
end of parallel loop for
else
Gram-Schmidt orthogonalization of all columns in $\mathbf{Q}$ and normalization $\mathbf{Q}^T\mathbf{M}\mathbf{Q} = \mathbf{I}$
paralel loop for $s = 1, 3m$
loop for $p = s, 3m$ (s, p – columns of matrix $\mathbf{Q}$)
$\mathbf{m}_{sp} = \mathbf{Q}_p{}^T\mathbf{M}\mathbf{Q}_s$ – evaluation of matrix $\mathbf{m}$
$\mathbf{k}_{sp} = \mathbf{Q}_p{}^T\mathbf{K}\mathbf{Q}_s$ – evaluation of matrix $\mathbf{k}$
end loop for
end of parallel loop for
Chol($\mathbf{m}$): $\mathbf{m} = \mathbf{l}\cdot\mathbf{l}^T$
end if
7. solve reduced eigenproblem $\mathbf{kq} - \mathbf{mq}\Lambda = 0$
8. obtain $\mathbf{V}^{k+1}$ and $\mathbf{P}^{k+1}$ using (4), (6).
9. parallel orthogonalization of $\mathbf{V}^{k+1}$ and $\mathbf{P}^{k+1}$ against converged eigenmodes.
end do

Algorithm 1. BSPPCG method

We accept the fixed dimension of subspace m, which is multiple to available number of threads np ($m\%np = 0$) for achievement a load balance between threads (point 1). Then, we prepare linearly independent start vectors $\mathbf{V}^0$ and set $\mathbf{p}_j^0 = \emptyset$, $j \in [1, m]$, where $\emptyset$ is a zero vector, and number of converged modes set to sero: $nconvmodes = 0$.

The iteration loop **do** $k = 1, 2, \ldots$ runs until $nconvmodes < n$, where n is the number of required modes (point 2).

In parallel loop (point 3) for each mode j we produce the normalization of $\mathbf{v}_j^k$, obtain the current approximation of eigenvalue λ_j^k, residual vector $\mathbf{r}_j^k$ and vector $\mathbf{z}_j^k$ from solution of linear equation set arising when preconditioning operator $\mathbf{B}$ is introduced for acceleration of convergence. On first iteration, the projection matrix $\mathbf{Q}$ contains only submatrices $\mathbf{Z}$ and $\mathbf{V}$ because submatrix $\mathbf{P}$ is zero. On all subsequent iterations, $\mathbf{Q}$ comprises $\mathbf{Z}$, $\mathbf{V}$ and $\mathbf{P}$ submatrices.

Iteration loop **do** j=1, m (point 4) checks of convergence. If convergence of j-th mode is achieved, we store the eigenpair $\{\lambda_j, \mathbf{v}_j\}$ to structure of data containing the final results, increment $nconvmodes$ and prepare the new linearly independent start vectors in addresses of vectors $\mathbf{v}_j^k$ and $\mathbf{p}_j^k$. In each starting vector $\mathbf{p}_j^k$ we put only one element equal to unit, all remaining elements are zero. Due to such an action, all new vectors $\mathbf{p}_j^k$ are linearly independent. All remaining elements of each vector $\mathbf{p}_j^k$ are zero. In such a way, we avoid the linear dependency between columns of projection matrix $\mathbf{Q}$. Then, we orthogonalize the new starting vectors $\mathbf{v}_j^k$ against converged modes and compute the λ_j^k, $\mathbf{r}_j^k$ and $\mathbf{z}_j^k$ corresponding to new starting vectors $\mathbf{v}_j^k$. Vectors $\mathbf{z}_j^k$, corresponding to new starting vectors, replace columns j in submatrix $\mathbf{Z}$, corresponding to converged vectors on current iteration step k.

The reduced matrix $\mathbf{m}$ is evaluated in **parallel loop for** $s = 1, 3m$. The sparse matrix $\mathbf{M}$ is multiplied by columns $\mathbf{Q}_s$ of matrix $\mathbf{Q}$ in parallel region: $\mathbf{w}_s = \mathbf{K}\cdot\mathbf{Q}_s$. The number of threads in team is np. In second loop (**loop for** $p = s, 3m$) we calculate the element $\mathbf{m}_{sp}$ as a dot product of column $\mathbf{Q}_p{}^T$ and previously obtained vector $\mathbf{w}_s$. Matrix $\mathbf{m}$ is symmetrical therefore p starts with s.

Chol($\mathbf{m}$) denotes the Cholesky factorization of matrix $\mathbf{m}$ and $\mathbf{l}$ is a lower triangular matrix. (point 6). We consider the matrix $\mathbf{m}$ as a weighted Gram matrix of subspace $span\{\mathbf{z}_1{}^k,\ldots,\mathbf{z}_m{}^k, \mathbf{v}_1{}^k,\ldots,\mathbf{v}_m{}^k, \mathbf{p}_1{}^k,\ldots,\mathbf{p}_m{}^k\}$. Therefore, if Cholesky factorization completes successfully, the columns of matrix $\mathbf{Q}$ are linearly independent and basis vectors are OK. Otherwise, if Cholesky factorization of $\mathbf{m}$ is failed, the columns of matrix $\mathbf{Q}$ are almost linearly dependent, and we produce the Gram-Schmidt orthogonalization of all columns in $\mathbf{Q}$ and normalization $\mathbf{Q}^T\mathbf{M}\mathbf{Q} = \mathbf{I}$. After this, we prepare reduced matrices $\mathbf{k}$, $\mathbf{m}$ using multithreaded parallelization and repeat Cholesky factorization of $\mathbf{m}$.

We apply procedures from LAPACK of Intel math kernel library (Intel MKL) [13] for solution of the generalized algebraic eigenproblem (5).

Submatrices $\mathbf{V}^{k+1}$ and $\mathbf{P}^{k+1}$ (point 8) is derived using (4), (6). We apply the multithreaded version of *dgemm* procedure from Intel MKL for multiplication of dense matrices.

The Gram-Schmidt orthogonalization provides orthogonality of vectors in the submatrices $\mathbf{V}^{k+1}$ and $\mathbf{P}^{k+1}$ to the converged eigenmodes (point 9). The columns in $\mathbf{V}^{k+1}$ as well as in $\mathbf{P}^{k+1}$ can be independently orthogonalized against converged eigenmodes, therefore, this algorithm can be easily parallelized. Also, orthogonalization, made in point 4, can be easily parallelized. Unlike these algorithms, the Gram-Schmidt orthogonalization procedure applied to all columns of matrix $\mathbf{Q}$ (point 6) has a strongly sequential nature and cannot be successfully parallelized.

We emphasize the fundamental differences between proposed method and LOBPCG.

1. BSPPCG method keeps constant the dimension of subspace m, which does not depend on number of required

eigenmodes. This allows us to reduce the amount of core memory and computing time and makes proposed approach applicable for solution of large problems on desktops and laptops.

2. As soon as the vectors converge, we immediately remove them from the block and replace with new starting vectors. In many cases, this allows us to keep a linear independence of the columns in matrix **Q**.

3. If in spite of everything, linear dependencies between base vectors still appears, we make the full reortogonalization of columns of the matrix **Q**.

4. We apply an efficient preconditioning for considered class of problems – incomplete Cholesky factorization by value developed in technique of sparse matrices [4].

III. Numerical Results

We consider example "stadium" taken from computational practice of SCAD Soft IT Company, developer of the SCAD FEA software, one of the most popular software used in the CIS countries for structural analysis and design, certified according to the regional norms.

We use computer with 16-core processor AMD Opteron 6276, 2.3/3.2 GHz, 64 GB DDR3 RAM, OS Windows Server 2008 R2 Enterprise SP1, 64 bit. The large amount of RAM allows us on application of incomplete Cholesky factorization for preparation of preconditioning with very small value of drop parameter $\psi = 10^{-16}$ [4] and keep all data in core memory. In addition, 16 processor cores provide the opportunity to explore the speed-up of method when we increase the number of cores. The tolerance is accepted as $tol = 10^{-3}$ – see (8).

The design model of stadium comprises 4 033 620 equations and consists of several types of finite elements: spatial frames, triangular and quadrilateral flat shell finite elements, elastic supports and rigid links One hundred eigenpairs are extracted (n = 100). The large number of almost multiple natural vibration frequencies occurs due to local vibration modes of bars in spatial trusses.

For accepted values of ψ, tol the number of iterations is 121 and number of reorthogonalizations, when Cholesky factoring of matrix **m** was failed, is 20. If there is at least one reorthogonalization, that means that LOBPCG method in version [10] would fail, since the columns of the matrix **Q**, which are basis vectors, are linearly dependent. Reorthogonalization of columns in matrix **Q** allows us to successfully continue a computation process. The shortest computational time is achieved on 16 threads.

Table I presents the total time of eigenvalue analysis of considered problem using proposed BSPPCG method, shifted block PCG (SBPCG) method [5] and shifted block Lanczos (SBLANC) method [3]. Method SBLANC solves this problem in core memory using PARFES [6], [7] – one from fastest for today sparse direct solvers on shared memory computers.

TABLE I
COMPARISON OF COMPUTATIONAL TIME FOR DIFFERENT METHODS. COMPUTER A, PROBLEM 1.

Method	Total time, s
BSPPCG	6 466
SBPCG	20 708
SBLANC (core mode)	6 228

The proposed BSPPCG demonstrates the solution time, which is slightly bigger than solution time of SBLANC method. The SBPCG method solves this problem considerably slower.

References

[1] S. Yu. Fialko, "Natural vibrations of complex bodies," *Int. Applied Mechanics,* vol. 40, no. 1, pp. 83 – 90, 2004, http://DOI: 10.1023/B:INAM.0000023814.13805.34.

[2] S. Fialko, "Aggregation Multilevel Iterative Solver for Analysis of Large-Scale Finite Element Problems of Structural Mechanics: Linear Statics and Natural Vibrations", in *PPAM 2001*, R. Wyrzykowski et al. (Eds.), *LNCS* 2328, Springer-Verlag Berlin Heidelberg, 2002, pp. 663–670, http://DOI: 10.1007/1-4020-5370-3_41.

[3] S. Yu. Fialko, E. Z. Kriksunov and V. S. Karpilovskyy, "A block Lanczos method with spectral transformations for natural vibrations and seismic analysis of large structures in SCAD software," in *Proc. CMM-2003 – Computer Methods in Mechanics*, Gliwice, Poland, 2003, pp. 129 —130.

[4] S. Yu. Fialko, "Iterative methods for solving large-scale problems of structural mechanics using multi-core computers," Archieves of Civil and Mechanical Engineering, vol. 14, pp. 190 – 203, 2014, http:// doi:10.1016/j.acme.2013.05.009.

[5] S. Yu. Fialko, F. Żegleń, "Block Preconditioned Conjugate Gradient Method for Extraction of Natural Vibration Frequencies in Structural Analysis", Proceedings of the FedCSIS. Łódź, 2015. IEEE Xplore Digital Library, pp. 655 – 662. DOI: 10.15439/2015F87. URL: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=7321505&tag=1.

[6] S. Yu. Fialko, "PARFES: A method for solving finite element linear equations on multi-core computers," Advances in Engineering software, vol. 40, no. 12, pp. 1256-1265, 2010, http:// doi:10.1016/j.advengsoft.2010.09.002.

[7] S. Yu. Fialko, "Parallel direct solver for solving systems of linear equations resulting from finite element method on multi-core desktops and workstations", Computers and Mathematics with Applications 70, pp. 2968–2987, 2015. doi:10.1016/j.camwa.2015.10.009

[8] C. K. Gan, P. D. Haynes and M. C. Payne, "Preconditioned conjugate gradient method for sparse generalized eigenvalue problem in electronic structure calculations," *Computer Physics Communications*, vol 134, nr. 1, pp. 33 – 40, 2001, http://DOI: 10.1016/S0010-4655(00)00188-0.

[9] V. Hernbadez, J. E. Roman, A. Tomas and V. Vidal, "A survey a software for sparse eigenvalue problems," Universitat Politecnica De Valencia, SLEPs technical report STR-6, 2009.

[10] A. V. Knyazev and K. Neymayr, "Efficient solution of symmetric eigenvalue problem using multigrid preconditioners in the locally optimal block conjugate gradient method," *Electronic Transactions on Numerical Analysis*, vol. 15, pp. 38 – 55, 2003.

[11] A. V. Knyazev, M. E. Argentati, I. Lashuk, E.E. Ovtchinnikov, "Block Locally Optimal Preconditioned Eigenvalue Xolvers (BLOPEX) in HYPRE and PETSC". URL: http://arxiv.org/pdf/0705.2626.pdf..

[12] Y. Saad, *Numerical methods for large eigenvalue problems, Revised edition, Classics in applied mathematics*. SIAM, 2011, http://dx.doi.org/10.1137/1.9781611970739.

[13] Intel Math Kernel Library Reference Manual. URL: https://software.intel.com/ru-ru/node/521001 (Last access: 20.06.2016).

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 673–678
DOI: 10.15439/2016F455
ACSIS, Vol. 8. ISSN 2300-5963

Error analysis for the first-order Gaussian recursive filter operator

Ardelio Galletti, Giulio Giunta
University of Naples "Parthenope"
Department of Science and Technology
Centro Direzionale, Isola C4, 80143, Naples, Italy
Email: {ardelio.galletti, giulio.giunta}@uniparthenope.it

***Abstract*—Nowadays, recursive filters (RFs) are frequently used in several research fields. More in particular, Gaussian RFs offer a more efficient way for computing approximate Gaussian filters and Gaussian-based convolutions. The use of such recursive filters introduces many sources of errors. Among them, here we consider the filter truncation error, that is the error due to the transition from the starting filter operator to the RF approximating it. Since input and output signals have infinite dimensions, the analysis of the related filter operator involves infinite matrices. In this paper, starting from a summary of the comprehensive mathematical background, we consider the case of the first-order Gaussian recursive filter. Then, taking into account the matrix form of the related operator, we perform the error analysis and provide theoretical results that estimate the filter truncation error.**

I. Introduction

IN RECENT years, recursive filters have been frequently used in several fields. For example Gaussian RFs are usually involved in image processing [1], [2] and are also implemented for solving three-dimensional variational analysis schemes in data assimilation [3]. Moreover, they have been recently constructed specifically for the electrocardiogram denoising [4], [5], [6]. The idea of recursive filters is to approximate a given filter, or for example the convolution with the impulse response of such a filter, in a more efficient way. More in particular, among RFs, the Gaussian RFs are very efficient implementations that approximate Gaussian-based convolutions. Gaussian RFs can be constructed in several ways but, in this work, we deal just with the kind derived by Deriche [7] and Young and van Vliet (see [8], and the references therein). As is well known, Gaussian RF based algorithms, applied to a signal with compact support, generate unbounded distortions in the output signal boundary entries (a detailed explanation is in [8]). This is known as edge effect and, to the aim of removing it, theoretical tools (named boundary conditions) and implementative improvements have been proposed in literature [8], [9]. Here we are not interested in edge effects and only focus on the error due to the the transition from the starting filter operator to the RF approximating it. We refer to this error as the filter truncation error. In this work, we are interested in studying the filter truncation error for the case of the first-order Gaussian recursive filter. The aim is to investigate on the quality of the approximation supplied by that filter. We underline that input and output signals have infinite dimensions, hence the analysis of the related filter operator will involve infinite matrices.

The paper is organized as follows. In Section 2, we give some mathematical preliminaries about the first-order Gaussian RF and also provide its matrix formulation. In Section 3, the analysis of the filter operator structure is carried out. In Section 4, we report the error analysis and provide an upper bound for the filter truncation error. Finally, conclusions in Section 5 close the paper.

II. Mathematical background

Let:

$$s^{(0)} = \{s_j^{(0)}\}_{j\in\mathbf{Z}} = (\dots, s_{-2}^{(0)}, s_{-1}^{(0)}, s_0^{(0)}, s_1^{(0)}, s_2^{(0)}, \dots)$$

be a input signal. $s^{(0)}$ can be thought of as a complex function defined on the set of integers, that is an element of the set of sequences of complex numbers $\mathbf{C}^{\mathbf{Z}}$. Let g denote the Gaussian function with zero mean and standard deviation σ. Let:

$$\delta_j = \begin{cases} 1 & \text{if } j = 0 \\ 0 & \text{if } j \neq 0 \end{cases} \tag{1}$$

be the *unit-sample*. The Gaussian filter is a filter whose impulse response to the unit-sample, i.e. the output of such a filter when the input is δ, is the Gaussian function g, or an approximation to it. Applying the Gaussian filter to the input $s^{(0)}$ gives rise to a response that can be simply expressed by the discrete Gaussian convolution:

$$s_j^{(g)} = \left(g * s^{(0)}\right)_j = \sum_{t=-\infty}^{+\infty} g_{j-t} s_t^{(0)}, \qquad \forall\, j \in \mathbf{Z}, \tag{2}$$

where:

$$g_t \equiv g(t) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{t^2}{2\sigma^2}\right). \tag{3}$$

The expression in (2) can be conveniently rewritten by changing the index t of the summation in $t - j$, and by making use of the symmetry $g_t = g_{-t}$. We have:

$$s_j^{(g)} = \sum_{t=-\infty}^{+\infty} g_t s_{j+t}^{(0)}, \qquad \forall\, j \in \mathbf{Z}. \tag{4}$$

The entries $s_j^{(g)}$ of $s^{(g)}$ can be efficiently approximated by means of Gaussian RFs and K-iterated Gaussian RFs.
A K-iterated n-order Gaussian RF filter computes the output

978-83-60810-90-3/$25.00

signal $s^{(K)}$, i.e. the K-iterate approximation of $s^{(g)}$, whose entries solve the infinite sequences of equations:

$$p_j^{(k)} = \beta s_j^{(k-1)} + \sum_{t=1}^{n} \alpha_t p_{j-t}^{(k)}, \qquad \forall\, j \in \mathbf{Z}, \tag{5}$$

$$s_j^{(k)} = \beta p_j^{(k)} + \sum_{t=1}^{n} \alpha_t s_{j+t}^{(k)}, \qquad \forall\, j \in \mathbf{Z}. \tag{6}$$

The filter iteration counter k goes from 1 to K (number of filter iterations). For $K = 1$, the filter merely becomes an n-order Gaussian RF filter. Equations in (5) and (6) are conveniently referred to as the advancing and backing filters, respectively: when a Gaussian RF is implemented as an algorithm, the index j must be treated in increasing order in the former and decreasing in the latter [8]. The values α_t and β are called *smoothing coefficients* and satisfy the constraint:

$$\beta = 1 - \sum_{t=1}^{n} \alpha_t.$$

In a general setting they depend on σ, n and K. In the following we consider just the first-order Gaussian RF with one-iteration only ($n = 1, K = 1$), for which equations (5) and (6) take the simplified form:

$$p_j = \beta s_j^{(0)} + \alpha p_{j-1}, \qquad \forall\, j \in \mathbf{Z}, \tag{7}$$

$$s_j = \beta p_j + \alpha s_{j+1}, \qquad \forall\, j \in \mathbf{Z}. \tag{8}$$

The smoothing coefficients are given by:

$$\alpha = 1 + E_\sigma - \sqrt{E_\sigma(E_\sigma + 2)} \tag{9}$$

and:

$$\beta = \sqrt{E_\sigma(E_\sigma + 2)} - E_\sigma, \tag{10}$$

with $E_\sigma = \sigma^{-2}$. The behaviour of α and β, as σ varies, is shown in Figure 1. Using Taylor expansion arguments, we can observe that, for small σ, it is:

$$\alpha = \frac{1}{2}\sigma^2 - \sigma^4 + \mathcal{O}\big(\sigma^6\big), \quad (\text{for} \quad \sigma \to 0),$$

while, for large σ, it is:

$$\alpha = 1 - \sigma^{-1}\sqrt{2} + \mathcal{O}\big(\sigma^{-2}\big), \quad (\text{for} \quad \sigma \to +\infty).$$

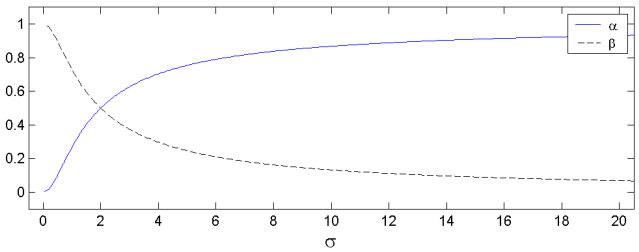


Fig. 1. Blue solid line: α. Black dashed line: β

By adapting the discussion in [3] to the case of the infinite dimension signals:

$$s^{(0)} = \big\{s_j^{(0)}\big\}_{j\in\mathbf{Z}}, \quad p = \big\{p_j\big\}_{j\in\mathbf{Z}} \quad \text{and} \quad s = \big\{s_j\big\}_{j\in\mathbf{Z}},$$

we can rewrite, in matrix form, the advancing filter (7) as:

$$Lp = s^{(0)}, \tag{11}$$

and the backing filter (8) as:

$$Us = p. \tag{12}$$

L and U are (bi)infinite matrices (matrices with infinite rows and columns, see [10] for notation and conventions), whose nonzero entries are:

$$L_{i,i} = \frac{1}{\beta}, \qquad L_{i,i-1} = -\frac{\alpha}{\beta}, \qquad \forall\, i \in \mathbf{Z}, \tag{13}$$

$$U_{i,i} = \frac{1}{\beta}, \qquad U_{i,i+1} = -\frac{\alpha}{\beta}, \qquad \forall\, i \in \mathbf{Z}. \tag{14}$$

Moreover L and U are both Toeplitz, bidiagonal matrices and U is the transpose of L, in the sense that:

$$U_{i,j} = L_{j,i}, \qquad \forall\, i, j \in \mathbf{Z}. \tag{15}$$

The products in (11) and (12) can be thought of as multiplications of infinite matrices. Given two infinite matrices:

$$A = \big\{A_{i,j}\big\}_{i,j\in\mathbf{Z}} \qquad \text{and} \qquad B = \big\{B_{i,j}\big\}_{i,j\in\mathbf{Z}},$$

the matrix product $C = AB$ has entries expressed as the series:

$$C_{i,j} = A_i B^j = \sum_{k=-\infty}^{+\infty} A_{i,k} B_{k,j},$$

where A_i and B^j are sequences denoting a row of A and a column of B, respectively. Now, substituting in equation (11) the expression of p given in (12), we deduce that the output s and the input $s^{(0)}$ of the first-order Gaussian RF satisfy:

$$Ds = s^{(0)}, \quad \text{with} \quad D = LU. \tag{16}$$

Then, to obtain the operator form for such a filter, we need just to invert the matrix D in (16). For infinite matrices, many definitions of inverse are given. Here, we mean the matrix A^{-1} as the inverse of an infinite matrix A, if and only if:

$$AA^{-1} = A^{-1}A = I,$$

where $I_{i,j} = \delta_{i-j}$, with δ_j as in (1). If D has inverse D^{-1}, from (16) it trivially results:

$$s = D^{-1}s^{(0)}. \tag{17}$$

Equation (17) proves that the infinite matrix:

$$F \stackrel{\text{def}}{=} D^{-1} = (LU)^{-1} \tag{18}$$

acts as the operator related to first-order Gaussian recursive filter. In the next section we will provide the structure of D and F.

III. FILTER OPERATOR

We need the following results: the first lemma provides a formula, equivalent to (7) and (8), which expresses the output entries s_j in terms of the input entries $s_j^{(0)}$, without using p_j values; the second lemma makes explicit the expression of the entries of the infinite matrix product LU.

Lemma III.1. *(Filter output entries representation) Let s, p_j and $s_j^{(0)}$ be as in* (7) *and* (8)*. If $\beta = 1-\alpha$ and $|\alpha| < 1$, then the output entries s_j are given by the series:*

$$s_j = \sum_{t=-\infty}^{+\infty} c_t s_{j+t}^{(0)}, \qquad \forall\, j \in \mathbf{Z}, \tag{19}$$

with:

$$c_t \equiv \frac{\beta}{1+\alpha}\alpha^{|t|}, \qquad \forall\, t \in \mathbf{Z}. \tag{20}$$

Proof. Let k be a positive integer. Combining the equation (7) in itself repeatedly, with indices j, $j-1, \ldots, j-(k-1)$, we obtain inductively:

$$p_j = \beta \sum_{m=0}^{k-1} \alpha^m s_{j-m}^{(0)} + \alpha^k p_{j-k}, \qquad \forall\, j \in \mathbf{Z},$$

and, since $|\alpha| < 1$, for $k \to +\infty$ it is:

$$p_j = \beta \sum_{m=0}^{+\infty} \alpha^m s_{j-m}^{(0)}, \qquad \forall\, j \in \mathbf{Z}, \tag{21}$$

Similarly, combining the equation (8) in itself, with indices j, $j+1, \ldots, j+(k-1)$, we get:

$$s_j = \beta \sum_{l=0}^{k-1} \alpha^l p_{j+l} + \alpha^k s_{j+k}, \qquad \forall\, j \in \mathbf{Z},$$

and, for $k \to +\infty$ it is:

$$s_j = \beta \sum_{l=0}^{+\infty} \alpha^l p_{j+l}, \qquad \forall\, j \in \mathbf{Z}. \tag{22}$$

Hence, using (21) with $j+l$ in (22), we obtain:

$$\begin{aligned} s_j &= \beta \sum_{l=0}^{+\infty} \alpha^l \left(\beta \sum_{m=0}^{+\infty} \alpha^m s_{j+l-m}^{(0)} \right) \\ &= \sum_{l=0}^{+\infty} \sum_{m=0}^{+\infty} \beta^2 \alpha^{l+m} s_{j+l-m}^{(0)}, \qquad \forall\, j \in \mathbf{Z}. \end{aligned} \tag{23}$$

Observing that, as l, m vary in $0, 1, \ldots$, $t = l - m$ varies in $\mathbf{Z}$, we can simplify the representation in (23) by collecting the coefficients of $s_{j+l-m}^{(0)} = s_{j+t}^{(0)}$, for each fixed t. Then, by means of a suitable change of indices in (23), we get:

$$s_j = \sum_{t=-\infty}^{+\infty} \left(\sum_{m=0}^{+\infty} \beta^2 \alpha^{|t|+2m} \right) s_{j+t}^{(0)}, \qquad \forall\, j \in \mathbf{Z}.$$

Finally, the thesis follows using $\beta = 1-\alpha$ and because of:

$$\sum_{m=0}^{+\infty} \beta^2 \alpha^{|t|+2m} = \beta^2 \alpha^{|t|} \sum_{m=0}^{+\infty} \alpha^{2m} = \frac{\beta^2}{1-\alpha^2}\alpha^{|t|} = c_t.$$

□

Lemma III.2. *(D structure) Let L and U be with entries as in* (13) *and* (14)*, respectively. Then $D = LU$ is a Toeplitz, tridiagonal, symmetric, infinite matrix, with entries:*

$$D_{i,j} = \begin{cases} \frac{1+\alpha^2}{\beta^2} & \text{if } \; j = i \\ -\frac{\alpha}{\beta^2} & \text{if } \; j = i \pm 1 \\ 0 & \text{if } \; j = i \pm m, \quad m \geq 2 \end{cases} \tag{24}$$

Proof. We need just to prove (24). Since $L_{i,k} = 0$ for $k < i-1$ and $k > i$, $\forall\, i, j \in \mathbf{Z}$ it is:

$$D_{i,j} = \sum_{k=-\infty}^{+\infty} L_{i,k} U_{k,j} = L_{i,i-1} U_{i-1,j} + L_{i,i} U_{i,j}.$$

Then recalling (13) and (15), it holds:

$$D_{i,j} = L_{i,i-1} L_{j,i-1} + L_{i,i} L_{j,i} = -\frac{\alpha}{\beta} L_{j,i-1} + \frac{1}{\beta} L_{j,i}. \tag{25}$$

Putting $j = i$ in (25), we obtain:

$$D_{i,i} = -\frac{\alpha}{\beta} L_{i,i-1} + \frac{1}{\beta} L_{i,i} = \left(-\frac{\alpha}{\beta}\right)^2 + \left(\frac{1}{\beta}\right)^2 = \frac{1+\alpha^2}{\beta^2}.$$

For $j = i-1$, it is:

$$D_{i,i-1} = -\frac{\alpha}{\beta} L_{i-1,i-1} + \frac{1}{\beta} L_{i-1,i} - \frac{\alpha}{\beta}\frac{1}{\beta} + \frac{1}{\beta} \cdot 0 = -\frac{\alpha}{\beta^2},$$

and for $j = i+1$, it is:

$$D_{i,i+1} = -\frac{\alpha}{\beta} L_{i+1,i-1} + \frac{1}{\beta} L_{i+1,i} = \frac{1}{\beta}\left(-\frac{\alpha}{\beta}\right) = -\frac{\alpha}{\beta^2}.$$

Finally, for $j = i \pm m$, with $m \geq 2$, we have:

$$L_{j,i-1} = L_{i\pm m,i-1} = 0$$

and:

$$L_{j,i} = L_{i\pm m,i} = 0.$$

Hence, from (25), we obtain:

$$D_{i,j} = D_{i,i+m} = L_{i,i-1} L_{i\pm m,i-1} + L_{i,i} L_{i\pm m,i} = 0,$$

and this completes the proof. □

Using the result in Lemma III.1 we can derive the structure of the operator $F = D^{-1}$. Notice that, from (17) and (18), it is:

$$s_j = \left(F s^{(0)}\right)_j = F_j s^{(0)} = \sum_{k=-\infty}^{+\infty} F_{j,k} s_k^{(0)}.$$

With the substitution $k = j + t$ we obtain the expression:

$$s_j = \sum_{t=-\infty}^{+\infty} F_{j,j+t} s_{j+t}^{(0)}, \tag{26}$$

which has the same form of the result in (19). This suggests that the F entries are given by the coefficients in (20). Starting from this remark, we deduce the form of F.

Theorem III.1. *(F structure) Let L and U be with entries as in* (13) *and* (14)*, respectively. Then $F = (LU)^{-1}$ is a Toeplitz, symmetric, infinite matrix, with entries:*

$$F_{i,j} = \frac{\beta}{1+\alpha}\alpha^{|j-i|}, \qquad \forall\, i, j \in \mathbf{Z}. \tag{27}$$

Proof. By comparing the series in (26) and (19), for each fixed $j \in \mathbf{Z}$, we get:

$$\sum_{t=-\infty}^{+\infty} F_{j,j+t} s^{(0)}_{j+t} = \sum_{t=-\infty}^{+\infty} c_t s^{(0)}_{j+t}.$$

Therefore, by taking for all $t \in \mathbf{Z}$, the input signal $s^{(0)}$ as the time shifted unit-sample with nonzero entry $s^{(0)}_{j+t}$, it follows:

$$F_{j,j+t} = c_t, \qquad \forall\, t \in \mathbf{Z}.$$

Then, recalling (20) we obtain the thesis:

$$F_{i,j} = F_{i,i+(j-i)} = c_{j-i} = \frac{\beta}{1+\alpha}\alpha^{|j-i|}, \qquad \forall\, i,j \in \mathbf{Z}.$$

□

Another proof that F actually acts as the inverse of D, can be obtained by verifying, by direct computation, that $DF = I$. To do this, observe that from Lemma III.2, we have:

$$(DF)_{i,j} = \sum_{k=-\infty}^{+\infty} D_{i,k} F_{k,j} = \sum_{k=i-1}^{i+1} D_{i,k} F_{k,j}.$$

So, for $j = i + m$, it is:

$$(DF)_{i,i+m} = D_{i,i-1}F_{i-i,i+m} + D_{i,i}F_{i,i+m} + D_{i,i+1}F_{i+1,i+m},$$

and by exploiting (24) and (27), we get:

$$(DF)_{i,i+m} = \frac{-\alpha\alpha^{|m+1|} + (1+\alpha^2)\alpha^{|m|} - \alpha\alpha^{|m-1|}}{\beta(1+\alpha)}. \quad (28)$$

For $j = i$, that is $m = 0$, (28) becomes:

$$(DF)_{i,i} = \frac{-\alpha^2 + (1+\alpha^2) - \alpha^2}{\beta(1+\alpha)} = \frac{1-\alpha^2}{\beta(1+\alpha)} = 1.$$

For $j > i$, that is $m \geq 1$, (28) becomes:

$$(DF)_{i,i+m} = \frac{-\alpha\alpha^{m+1} + (1+\alpha^2)\alpha^{m} - \alpha\alpha^{m-1}}{\beta(1+\alpha)} = \alpha^m \frac{-\alpha^2 + (1+\alpha^2) - 1}{\beta(1+\alpha)} = 0.$$

For $j < i$, that is $m \leq -1$, (28) becomes:

$$(DF)_{i,i+m} = \frac{-\alpha\alpha^{-m-1} + (1+\alpha^2)\alpha^{-m} - \alpha\alpha^{-m+1}}{\beta(1+\alpha)} = \alpha^{-m} \frac{-1 + (1+\alpha^2) - \alpha^2}{\beta(1+\alpha)} = 0.$$

IV. Error analysis

In this section we are interested in studying the error occurring when the Gaussian filter is substituted by the first-order Gaussian RF, namely the filter truncation error. Let $\|\cdot\|$ denote the sup-norm, defined for signals f as:

$$\|f\| = \sup_{k \in \mathbf{Z}} |f_k|,$$

and for infinite matrices A as:

$$\|A\| = \sup_{f \in \mathbf{C}^{\mathbf{Z}}, \|f\|=1} \|Af\| = \sup_{i \in \mathbf{Z}} \sum_{j \in \mathbf{Z}} |A_{i,j}|.$$

Let denote by:

$$\tau_j = s^{(g)}_j - s_j, \quad (29)$$

the difference between the output entries of the Gaussian filter and the first-order Gaussian RF. We refer to:

$$\tau = \left\| \{\tau_j\}_{j \in \mathbf{Z}} \right\| \quad (30)$$

as the filter truncation error (f.t.e.). Before giving an upper bound for τ, let us indicate by V the infinite Gaussian matrix, with entries:

$$V_{i,j} = g_{j-i}, \qquad \forall\, i,j \in \mathbf{Z}, \quad (31)$$

where g_t values are as in (3). Now, using the operator V, the equation (4) is compactly represented as:

$$s^{(g)} = V s^{(0)}. \quad (32)$$

Therefore, combining (17), (18), (29) and (32), we get:

$$\{\tau_j\}_{j \in \mathbf{Z}} = s^{(g)} - s = V s^{(0)} - F s^{(0)} = (V - F) s^{(0)}, \quad (33)$$

that is the f.t.e. is simply obtained as the product of the filter operators difference and the input signal. Starting from (33) we can proof the following main result.

Theorem IV.1. *(Filter truncation error) Assume that $\|s^{(0)}\| \leq S$ and let V be the same as in* (31)*. Then, for the f.t.e. defined in* (29) *and* (30)*, it holds that:*

$$\tau \leq \kappa \cdot S, \quad (34)$$

with:

$$\kappa = \|V - F\| = \sum_{t=-\infty}^{+\infty} |g_t - c_t| \quad (35)$$

and g_t and c_t as in (3) *and* (20)*, respectively.*

Proof. Immediate by construction. The proof of (34) is as follows. From (30) and (33) it is:

$$\tau = \|(V - F)s^{(0)}\| \leq \|V - F\| \cdot \|s^{(0)}\| = \kappa \cdot S.$$

To complete the proof, we need just to prove (35). From (20) and (27) we deduce that $\forall\, i,t \in \mathbf{Z}$ it is:

$$F_{i,i+t} = \frac{\beta}{1+\alpha}\alpha^{|(i+t)-i|} = \frac{\beta}{1+\alpha}\alpha^{|t|} = c_t.$$

Then, changing the summation index j in $i + t$, and by using (31), we get the thesis:

$$\begin{aligned} \kappa = \|V - F\| &= \sup_{i \in \mathbf{Z}} \sum_{j \in \mathbf{Z}} |V_{i,j} - F_{i,j}| \\ &= \sup_{i \in \mathbf{Z}} \sum_{t \in \mathbf{Z}} |V_{i,i+t} - F_{i,i+t}| = \sup_{i \in \mathbf{Z}} \sum_{t \in \mathbf{Z}} |g_t - c_t| \\ &= \sum_{t \in \mathbf{Z}} |g_t - c_t| = \sum_{t=-\infty}^{+\infty} |g_t - c_t|. \end{aligned} \quad (36)$$

□

The previous result proves that, without considering the order of magnitude S of the input signal, the factor κ behaves

like a physical limit in the accuracy provided by the first-order Gaussian RF in approximating the Gaussian convolution. Then, for investigating on the f.t.e., and completing the error analysis, we can limit our discussion to the behaviour of κ. Recalling (3), (9), (10) and (20), we deduce that coefficients g_t and c_t depend on σ and t, making κ dependent just on σ. We remark that if $g_t - c_t$ was of constant sign, for example $g_t - c_t \geq 0,\ \forall\, t \in \mathbf{Z}$, then we would simply have:

$$\kappa = \sum_{t=-\infty}^{+\infty} g_t - \sum_{t=-\infty}^{+\infty} c_t = \sum_{t=-\infty}^{+\infty} g_t - 1 \leq \frac{1}{\sigma\sqrt{2\pi}}, \qquad (37)$$

where the last inequality arises from:

$$\sum_{t=-\infty}^{+\infty} g_t \leq 1 + \frac{1}{\sigma\sqrt{2\pi}}.$$

This bound can be easily proved exploiting the monotonicity properties of the Gaussian function:

$$g_t \leq \int_{t-1}^{t} g(x)dx, \qquad \forall t = 1, 2, \ldots,$$

and the symmetry $g_t = g_{-t}$. Indeed, we have:

$$\sum_{t=-\infty}^{+\infty} g_t = g_0 + 2\sum_{t=1}^{+\infty} g_t \leq \frac{1}{\sigma\sqrt{2\pi}} + 2\sum_{t=1}^{+\infty}\int_{t-1}^{t} g(x)dx$$
$$= \frac{1}{\sigma\sqrt{2\pi}} + 2\int_{0}^{+\infty} g(x)dx = \frac{1}{\sigma\sqrt{2\pi}} + 1.$$

However, in general, the coefficients $g_t - c_t$ change sign as t varies. An example of this fact is in Figure 2, where $g_t - c_t$ values are obtained for $\sigma = 100$. Consequently, (37) cannot

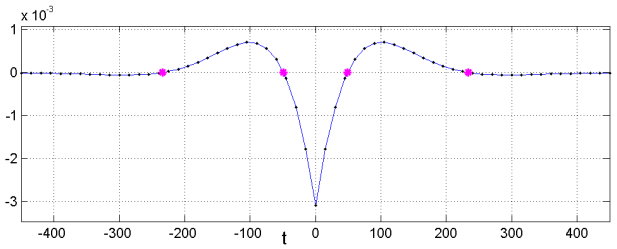


Fig. 2. Behaviour of $g_t - c_t$ for $\sigma = 100$ and $t = -450, \ldots, 450$

be proved and κ is not bounded by $1/(\sigma\sqrt{2\pi})$. In fact, the behaviour of κ, as σ varies in $[0.05, 50\,000]$, is shown in Figure 3. The figure highlights that κ takes its minimum values $\kappa_{min} = 0.05$ for $\sigma \approx 0.47$ and, except for values of σ in a small interval ($[0.37, 0.60]$), is always greater than 0.25. Moreover, we observe the following asymptotic behaviours:

- for small values of σ, κ seems to be unbounded and to increase like $1/\sigma$. This trend is consistent with (37) and is easily proved. Observing that $\alpha > 0$ implies:

$$c_0 = \frac{\beta}{1+\alpha} = \frac{1-\alpha}{1+\alpha} < 1,$$

for $\sigma < 1/\sqrt{2\pi} = 0.398$, it is:

$$\frac{1}{\sigma\sqrt{2\pi}} - 1 \leq g_0 - c_0 = |g_0 - c_0| \leq \kappa;$$

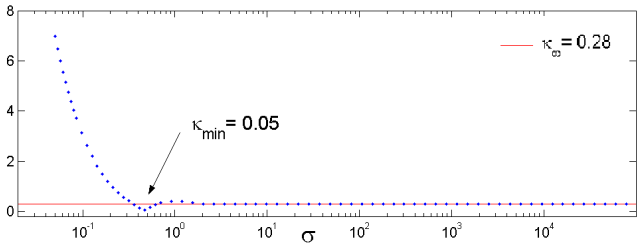


Fig. 3. Behaviour of κ for 75 values of σ increasing exponentially in the interval $[0.05, 50\,000]$

- for σ large enough, ($\sigma > 3.4$), κ becomes indeed constant and takes nearly the asymptotic value:

$$\kappa_\infty = \lim_{\sigma\to\infty} \kappa \approx 0.28.$$

This is the empirical evidence that (37) does not hold true for all σ. The value κ_∞ can be found observing that, for large σ, κ can be accurately approximated by the integral:

$$\int_{-\infty}^{+\infty} |g(t) - c(t)|dt,$$

where g is the Gaussian function, and c is the function:

$$c(t) = \frac{\beta}{1+\alpha}\alpha^{|t|}, \qquad \forall\, t \in \mathbf{R}.$$

To compute the integral, we need to establish when $g - c$ changes sign. Figure 2 shows that, for $\sigma = 100$, $g - c$ has 4 zeros (two pairs: $\pm t_1$, $\pm t_2$) and that changes sign 4 times. That is what actually happens for each large enough σ. Indeed, using the approximations:

$$\alpha \approx 1 - \frac{\sqrt{2}}{\sigma},\ \beta \approx \frac{\sqrt{2}}{\sigma},\ \frac{1}{1+\alpha} \approx \frac{1}{2},\ 1 - \frac{\sqrt{2}}{\sigma} \approx \exp\left(-\frac{\sqrt{2}}{\sigma}\right),$$

we obtain:

$$c(t) \approx \frac{\sqrt{2}}{\sigma}\frac{1}{2}\left(1 - \frac{\sqrt{2}}{\sigma}\right)^{|t|} \approx \frac{1}{\sigma\sqrt{2}}\exp\left(-\sqrt{2}\frac{|t|}{\sigma}\right) = \tilde{c}(t).$$

Solving $g(t) = \tilde{c}(t)$ for $x = \frac{|t|}{\sigma}$, we get:

$$\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right) = \frac{1}{\sigma\sqrt{2}}\exp\left(-\sqrt{2}x\right),$$

from which, taking the logarithms:

$$-x^2 - \ln\pi = -2x\sqrt{2},$$

and so:

$$x_1 = \sqrt{2} - \sqrt{2 - \ln\pi} = 0.489, \quad \text{and} \quad t_1 = x_1\sigma,$$

and:

$$x_2 = \sqrt{2} + \sqrt{2 - \ln\pi} = 2.339, \quad \text{and} \quad t_2 = x_2\sigma.$$

Finally, the value κ_∞ is achieved exploiting the symmetry and the sign of $|c-g|$. We have:

$$\kappa_\infty = 2\int_0^{+\infty} |g(t)-\tilde{c}(t)|dt = 2\int_0^{x_1\sigma} \big(\tilde{c}(t)-g(t)\big)dt+$$
$$+2\int_{x_1\sigma}^{x_2\sigma} \big(g(t)-\tilde{c}(t)\big)dt + 2\int_{x_2\sigma}^{+\infty} \big(\tilde{c}(t)-g(t)\big)dt = 0.28,$$

where, after changing the variable of integration t in $x = t/\sigma$ and specifying computations, one can see that the values of the integrals are no longer dependent on σ. In conclusion, our analysis has pointed out that, except for a small subinterval of σ values, the bound κ of the filter truncation error is never significantly small. Then we can state that the first-order Gaussian RF, when used in a single iteration, does not offer a good approximation of the Gaussian convolution.

V. Conclusions

In this work, we have given mathematical preliminaries and definitions about Gaussian RFs by focusing on the first-order Gaussian RF. For this filter we have studied the related matrix operator, by providing a complete description of its structure. Then, we have studied the associated filter truncation error and, exploiting the structure operator, we have given a theoretical error upper bound. This result has been used to investigate about the quality of the approximation supplied by that filter and has allowed to conclude that, in general, this filter is not very accurate.

References

[1] van Vliet, L.J., Young, I.T., Verbeek, P.W.. - *Recursive Gaussian derivative filters*. The 14 th International Conference on Pattern Recognition, pp. 509-514, DOI: 10.1109/ICPR.1998.711192, 1998.

[2] Young, I.T., van Vliet L.J.. - *Recursive implementation of the Gaussian filter*. Signal Processing 44, pp 139-151, 1995.

[3] Cuomo, S., Farina, R., Galletti, A., Marcellino, L.. -*An error estimate of Gaussian recursive filter in 3Dvar problem* Federated Conference on Computer Science and Information Systems, FedCSIS 2014, art. no. 6933068, pp. 587-595, 2014. DOI: 10.15439/2014F279

[4] Cuomo, S., De Pietro, G., Farina, R., Galletti, A., Sannino, G.. - *A novel O(n) numerical scheme for ECG signal denoising* Procedia Computer Science, 51 (1), pp. 775-784, 2015. DOI: 10.1016/j.procs.2015.05.198

[5] Cuomo, S., De Pietro, G., Farina, R., Galletti, A., Sannino, G.. - *A framework for ECG denoising for mobile devices* PETRA 2015 ACM. ISBN 978-1-4503-3452-5/15/07, DOI: 10.1145/2769493.2769560, 2015.

[6] Cuomo, S., De Pietro, G., Farina, R., Galletti, A., Sannino, G.. - *A revised scheme for real time ECG Signal denoising based on recursive FILTERING*, Biomedical Signal Processing and Control, 27, pp. 134-144, 2016. DOI: 10.1016/j.bspc.2016.02.007

[7] R. Deriche - *Recursively implementating the Gaussian and its derivatives*. INRIA Research Report RR-1893, 1993, pp.24.

[8] Cuomo, S., Farina, R., Galletti, A., Marcellino, L.. - *A K-iterated scheme for the first-order Gaussian Recursive Filter with boundary conditions* Federated Conference on Computer Science and Information Systems, FedCSIS 2015, pp.641-647, 2015. DOI: 10.15439/2015F286

[9] Triggs, B., Sdika M.. - *Boundary conditions for Young-van Vliet recursive filtering*. IEEE Transactions on Signal Processing, 54 (6 I), pp. 2365-2367, 2006.

[10] de Boor, C., Jia, R.-q., Pinkus, A.. - *Structure of invertible (bi)infinite totally positive matrices* Linear Algebra and Its Applications, 47 (C), pp. 41-55, 1982. DOI: 10.1016/0024-3795(82)90225-7

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 679–683
DOI: 10.15439/2016F344
ACSIS, Vol. 8. ISSN 2300-5963

Acceleration of image reconstruction in 3D Electrical Capacitance Tomography in heterogeneous, multi-GPU system using sparse matrix computations and Finite Element Method

Paweł Kapusta*,Michał Majchrowicz†,Dominik Sankowski‡ and Lidia Jackowska-Strumiłło§
Lodz University of Technology
Institute of Applied Computer Science
ul. Stefanowskiego 18/22, Łódź, Poland
* Email: pawel.kapusta@p.lodz.pl † Email: mmajchr@iis.p.lodz.pl ‡ Email: dsan@iis.p.lodz.pl § Email: lidia_js@iis.p.lodz.pl

***Abstract*—3D Electrical Capacitance Tomography provides a lot of challenging computational issues that have been reported in the past by many researchers. Image reconstruction using deterministic methods requires execution of many basic operations of linear algebra. Due to significant sizes of matrices used in ECT for image reconstruction and the fact that best image quality is achieved by using algorithms of which significant part is FEM and which are hard to parallelize or distribute. In order to solve these issues a new set of algorithms had to be developed.**

I. Introduction

ELECTRICAL Capacitance Tomography (ECT) is a relatively new imaging technique that can be used for non-invasive visualization in industrial applications in 2D, 3D and even 4D dynamic mode. ECT is performing the task of imaging of materials with a contrast in dielectric permittivity by measuring capacitance from a set of electrodes (Fig. 1). Among other non-invasive imaging techniques, ECT is characterized by much higher temporal resolution than Magnetic Resonance Imaging, Computed Tomography etc.

Unfortunately to achieve best image quality in 3D image reconstruction complex algorithms have to be used, especially ones that use large sensitivity matrices, Finite Element Method as well as neural networks approach [3].

In this article the authors have focused on accelerating non-linear image reconstruction algorithms, that are based on Finite Element Method and use sparse matrices to store data. We show that it is indeed possible to parallelize such algorithm and achieve significant speed-up, as well as develop them in such a way, to be able to use them in a distributed, heterogeneous computational system.

A. Image reconstruction in ECT

The scheme of image synthesis in Electrical Capacitance Tomography is called image reconstruction. It is based on solving the so called inverse problem, in which the spatial distribution of electric permittivity from the measured values of capacitance C is approximated. We can distinguish two types of image reconstruction algorithms. Firstly there are linear algorithms, which, because of higher temporal resolution, are used for monitoring fast-varying industrial process applications, like oil-gas flows in pipelines [1] or gravitational flows and discharging of silo [9] and non-linear algorithms, which allow reconstructing images with higher quality. Afterwards, reconstructed images can be analysed using either state of the art algorithmic approach, such as fuzzy-logic based classification [1] or by using a novel method of applying crowdsourcing [2], in order to determine, for example, flow characteristics.

II. Non-linear reconstruction algorithms

Non-linear three-dimensional image reconstruction in 3D capacitance tomography is a complex numerical problem, saturated with linear algebra transformations. During this iterative calculation process a set of parameters is determined, that is necessary for proper reconstruction of three-dimensional tomographic image optimization. The general idea of the algorithm is presented in Figure 2. One of the three key stages of the iterative process of reconstruction is a forward problem involving setting up a simulated vector based on a given spatial distribution of dielectric permittivity. The accuracy of the forward problem solution has a significant impact on the quality and speed of image reconstruction, and depends on the method of its determination. Most often forward problem is determined numerically using the Finite Element Method (FEM) based on a numerical model of a capacitance sensor. The authors have focused primarily on developing methods for accelerating the

Fig. 1. Object and 3D reconstruction obtained using Electrical Capacitance Tomography

calculations using algorithms developed specifically for use with sparse matrices (CULA library, CUSP). This made it possible to develop proprietary paralel computing algorithms (as a set of functions and procedures), dedicated to specific processing of tomographic data. Developed methods allow reconstructing three-dimensional images by using relatively fast methods of solving sparse matrix equations (AMG method - Algebraic Multi Grid, the Jacobi method and the Conjugate Gradient algorithm), which are computed on graphic processors.

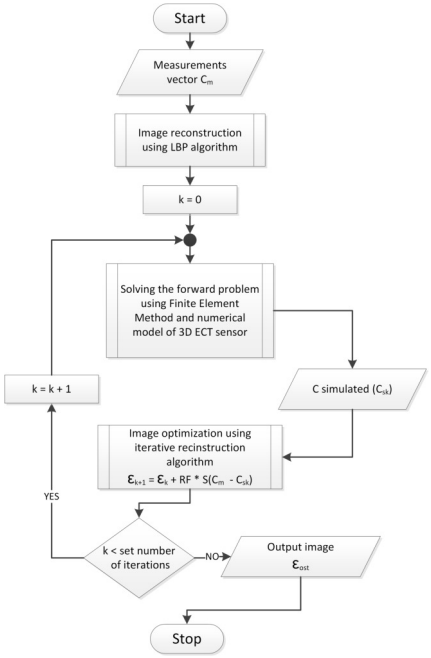


Fig. 2. General non-linear reconstruction algorithm in ECT

A. Finite Element Method

Image reconstruction algorithms using the Finite Element Method are often used in 3D Electrical Capacitance Tomography because of the possibility of obtaining a more accurate solution to the forward problem than linear algorithms, which in turn can improve spatial resolution of resulting 3D images. A major drawback of this method, however, is its large computational complexity. The authors have developed a number of proprietary software algorithms, which are designed to significantly reduce the time of image reconstruction using the Finite Element Method, through the implementation of parallel computing for sparse matrices and calculations in a heterogeneous and distributed environments.

Main idea of the developed algorithm is to obtain the solution (electric field distribution) given in the form of equation:

$$\boldsymbol{\varphi} = \boldsymbol{Y}^{-1}\boldsymbol{F} \tag{1}$$

where:

$\boldsymbol{\varphi}$ - is a sought distribution of the electric field - represented by the spatial distribution of nodal potential - partial solution of the forward problem in capacitance tomography;

$\boldsymbol{Y}$ - is a transformation matrix, built according to the geometric dependencies of sensor model mesh and Neumann boundary conditions;

$\boldsymbol{F}$ - is the extortion vector, defining the given Dirichlet boundary conditions

The first step of the algorithm is to pre-process the input data and store it as a set of sparse matrices. Then, in order to obtain $\boldsymbol{Y}$, matrix decomposition is performed as described by the equation:

$$\boldsymbol{Y} = \boldsymbol{A}^{\boldsymbol{T}}\boldsymbol{B}\boldsymbol{A} \tag{2}$$

where:

$\boldsymbol{A}$ - shape functions gradients matrix

$\boldsymbol{B}$ - matrix of normalized mesh volumes

$\boldsymbol{A}^{\mathrm{T}}$ - transposed shape functions gradients matrix

The next step of the algorithm is processing of the input data matrix and selecting the rows corresponding to each electrode - known potential in nodes describing the electrode. Then the matrix is preconditioned using either Jacobi or Algebraic Multi-Grid method. This makes it possible to solve the equation (1) using a Conjugate Gradients Method. Once this is done the resulting matrix is supplemented with data from the Dirichlet boundary conditions. The last step of the algorithm is to determine the vector of simulated measurements using Gauss' law described by the formula:

$$C_{eg} = \frac{\iiint_{\Omega} \varepsilon\,(x,y,z)\; grad[\varphi(x,y,z)]d\Omega}{\varphi_e - \varphi_g} \tag{3}$$

where:

C_{eg} – Capacitance between electrodes e and g

ε(x,y,z) – distribution of electric permitivity

φ(x,y,z) – distributon of potential

φ_e – electric potential on electrode e

φ_g – electric potential on electrode g

x,y,z – cartesian coordinates

B. Computations using sparse matrices

The operation of multiplying three matrices, represented by the formula (2), is an integral part of the Finite Element Method for 3D ECT. This action, however, is characterized by high computational complexity. Moreover, the stiffness matrices $\boldsymbol{Y}$, generated by numerical models of 3D ECT sensors, are too large to fit entirely in RAM of graphics cards. However, the number of non-zero elements is relatively small in relation to the dimensions. Thus, it is possible to treat them as sparse matrices to reduce memory usage.

There are many formats for storing sparse matrices. Among them the most common formats are CSC (Compressed Sparse Column) and CSR (Compressed Sparse Row). The authors decided to use CSR format because it allows, in most cases, for optimal access to the data stored in GPU memory. Reading and writing data is usually done in a row-major manner, which is optimal for most architectures of CPUs and GPUs. Saving sparse matrix in the CSR format is, for the same reason, not optimal for multiplication, as there needs to be a way of quickly accessing the columns of the matrix without causing

uncoalesced reads/writes from GPU memory. This situation arises when data must be read in a manner that does not comply with optimal memory access for specific hardware and cannot be obtained in one transaction. The impact of this phenomenon on the speed of computations is highly dependent on the hardware architecture of the GPU, however, it is always significant.

In order to significantly reduce this problem the authors have introduced a hybrid format, called Hybrid Compressed Sparse Row-Column (H-CSRC), comprised of both the records of CSR and CSC. Depending on the needs, data can be accessed in either row or column-major manner, while minimizing memory operation and maintaining compatibility with other algorithms.

Multiplication of three sparse matrices has been implemented as a single operation. This approach allows for optimal use of local and private memory on the GPU, in order to increase the speed of calculations. In this algorithm, it is necessary to use the local memory, shared by thread groups, to minimize the number of global memory accesses.

In 3D ECT it is particularly important to optimize each of the algorithms for the speed of execution. Hence the authors have developed a special version of three matrix multiplication algorithm, which takes into account all the specific properties of the matrix calculations in the 3D ECT, as defined by equation (2). There are three main properties of the input data, specific to the ECT, that allow for further optimizations:

- Items in the matrix $\boldsymbol{B}$ have a non-zero values on the main diagonal only. In addition, they repeat in sets of three, which is due to the specificity of the input data.
- The output array is symmetrical along its main diagonal, which, using proper element indexing, can reduce the number of operations almost by half.
- Due to the nature of the calculations, the amount of output elements and their position, does not change during the execution of the program, assuming the immutability of input data distribution. Hence this can be determined before the execution of the program and put into the algorithm as a map of elements to instantly skip the input matrix elements, which are known a priori to not produce results.

C. Parallelization

The first variant of image reconstruction algorithm using the Finite Element Method is the reconstruction in the local system. As it constitutes a platform for further modifications it was necessary to design and implement a solution, that would be also applicable in multi-GPU [4], as well as distributed systems. To ensure efficient 3D image reconstruction the proposed algorithm includes data caching solutions. This issue is particularly important in the case of heterogeneous systems. In most 3D ECT systems measurement data is collected with higher frequency than it can be reconstructed. Moreover, because of the asynchronous nature of the developed solution, based on the commissioning of tasks to local GPUs using CUDA technology, as well as remote computing nodes, delays can accumulate, therefore there is a need for their elimination by buffering systems. All the algorithms have been designed, implemented and optimized from the start as a solution suited to multi-GPU and distributed systems. Due to the specific nature of the computations the most optimal solution is to start a separate thread for each GPU in the system, that are synchronised when reading the results.

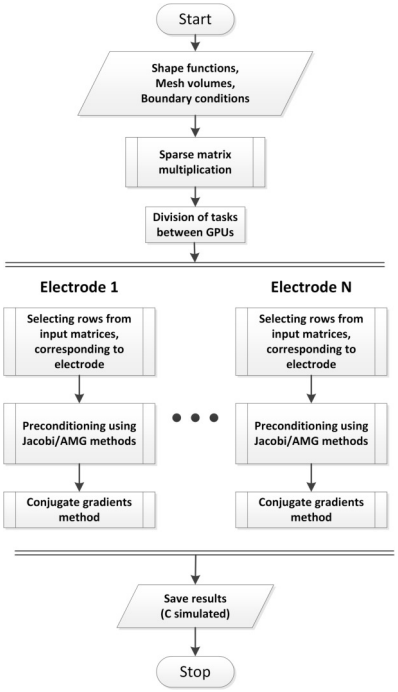


Fig. 3. Algorithm for calculating solution to forward problem using multiple GPUs

All the algorithms, developed by the authors, were designed to function in systems with multiple GPUs. Thus a natural direction for obtaining a further acceleration of computations is to perform non-linear 3D image reconstruction in a distributed system, which tends to have a higher degree of heterogeneity than the local systems.

The main idea of the developed algorithm is, that each GPU inside the compute calculates the solution to forward problem for one or more electrodes, by assigning to each GPU calculations specific to the selected electrode of the system, where all the GPU compute the result of a single image reconstruction (Fig. 3). This approach allows for more precise control of tasks allocation. This in turn enables its use in distributed systems with a high degree of heterogeneity. This solution also allows for potential reduction in overall system response time, since all nodes perform calculation for a single output image. Therefore, the task of caching is greatly simplified, and the image can be displayed on the screen without introducing delays larger than the reconstruction time for a single image. The disadvantage of this solution is, however, that it limits the scalability of a distributed system, since the total number of graphics processors cannot be greater than the number of electrodes.

TABLE I
COMPONENTS OF TEST SYSTEMS

Processor	HPC Hal: Intel i7-930 (4 cores, 8 threads) HPC Dave: Intel i7-920 (4 cores, 8 threads)
RAM	HPC Hal: 12 GB (6x2 GB) DDR3 1833 MHz HPC Dave: 8 GB (4x2 GB) DDR3 1833 MHz
GPU accelerators	HPC Hal: NVIDIA Tesla S1070 + Tesla C2070 HPC Dave: 2x GTX 570
Operating systems	Windows 7 64-bit

III. RESULTS

The research conducted on ECT algorithms [6] has shown that, although, dynamic development of GPU computing performance and its recent application for image reconstruction in ECT has significantly improved calculations time, in modern systems a single GPU is not enough to perform many tasks [7]. As a result multiple GPUs have to be used to accelerate calculations [5]. Thus, the authors are proposing a distributed, multi-node, multi-GPU heterogeneous system with a software layer that will allow use of multiple computers with fast GPUs to perform calculations across network connection [5]. The developed system, based on the proprietary KISDC networking platform [8], is designed to fully exploit parallel performance of all devices that the nodes are equipped with. Such architecture is very scalable and makes it possible to increase computation performance by adding new network nodes. Reconstruction algorithm verification tests were conducted using real measurement data, recorded during the research under Ministry of Education grant number 4664/B/T02/2010/38, using semi-industrial installation. Due to the nature of calculations using the graphics processors, the stability of execution times is lower than for algorithms executed on the CPU. Therefore, all of the results shown in this paper represent the worst case scenario - the lowest number of reconstructed images per second, achieved during testing.

All the results achieved with GPUs were compared with the performance of algorithms executed on the CPU, implemented using optimized BLAS libraries (Basic Linear Algebra Subprograms), compiled with Intel compiler and optimized for the tested CPU architecture.

A. Non-linear algorithms - local system

Reconstruction tests using non-linear algorithms and multiple graphics processors at the same time were carried out using NVIDIA Tesla C2070 card and NVIDIA Tesla S1070-400 computing server, which has four graphics cores (Table I). Verification of developed solutions in this case was performed for 1, 2, and 4 GPUs. The division of tasks between the graphics processors was done by creating a new thread for each GPU. As a result, it was possible to separate the control flow of the application from computations, thus allowing for asynchronous commission of tasks to the GPUs. Tests were performed for 10 iterations of non-linear reconstruction algorithm. The test results are presented as the number of images obtained per second.

TABLE II
RESULTS OF NON-LINEAR IMAGE RECONSTRUCTION [IMAGES/SECOND]

Elements in image vector	4 GPUs	2 GPUs	GPU	CPU (BLAS)
8488	0.035	0.024	0.015	0.003
20499	0.021	0.012	0.007	0.002
60896	0.007	0.004	0.002	0.001
87172	0.005	0.003	0.002	<0.001
157264	0.003	0.002	0.001	<0.001

All the tests were performed using the developed task division algorithm, by assigning each GPU calculations for a specific set of electrodes (Fig. 3). Moreover, verification was conducted using an optimized version of this algorithm, which enables asymmetric division of compute jobs between the units. Based on the known efficiency parameters, each GPU was assigned with solving the forward problem for appropriate number of electrodes. For example, by using two GPUs - Tesla C2070 card and a single GPU from Tesla S1070 accelerator, the C2070 card computes solution for 18 electrodes, and the S1070 GPU for the remaining 14. The results for this configuration are shown in Table II and in Figure 4.

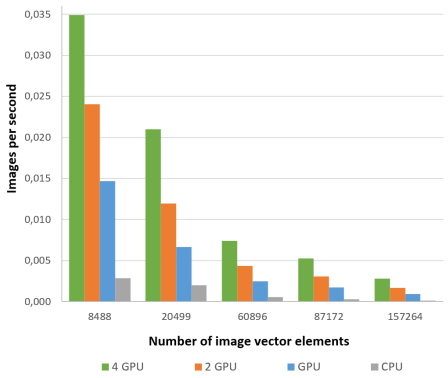


Fig. 4. Speed of non-linear reconstruction

As can be seen from Table II and Fig. 4, the use of multiple GPUs allowed to accelerate a non-linear image reconstruction by up to seventeen times for the two graphics processors and 28-fold in the case of four GPU compared to calculations using traditional algorithms executed on a CPU.

B. Non-linear algorithms - distributed system

Image reconstruction tests in distributed environment were performed using two HPC nodes code named: Dave and Hal. Their full specification is shown in Table I. In this case verification was performed for a total of 6 GPUs - four in HPC Hal (Tesla C2070 + Tesla S1070) and two in HPC Dave (2x Nvidia GTX 570). As it was in the case with computations in the local system all the tests were performed using task allocation based on number of electrodes. The data was sent between the nodes using KISDC networking layer, developed specifically by the authors for use in distributed

TABLE III
RESULTS OF DISTRIBUTED NON-LINEAR IMAGE RECONSTRUCTION [IMAGES/SECOND]

Elements in image vector	Local system 4 GPUs	Distributed system 4+2 GPUs	Speed-up
8488	0.0349	0.0381	1.09
20499	0.0210	0.0198	0.94
60896	0.0074	0.0086	1.16
87172	0.0053	0.0059	1.12
157264	0.0028	0.0034	1.20

image reconstruction in 3D ECT. All the results for these tests are presented in Table III and in Figure 5.

Using the distributed system for image reconstruction purposes the authors were able to speed-up the computations compared to local system by up to 20%. There was however one exception - for the image vector size of 20499. In this case the computations on a distributed system were slower than in local environment. This was caused by a combination of overheads resulting from synchronisation and network delays. Moreover, because of the specifics of GPU computations it is common to come across a combination of input data sizes and algorithm logic that will cause overall slow-down in specific cases. Nevertheless, the authors are sure that further work on the developed algorithms will result in even better results in the future.

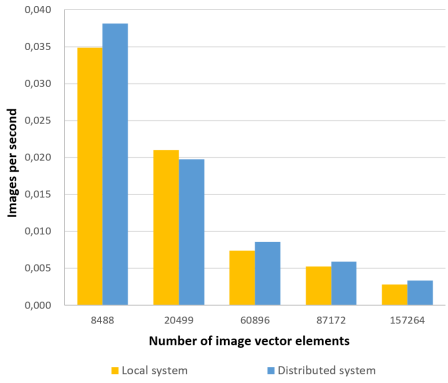


Fig. 5. Speed of non-linear reconstruction

IV. CONCLUSION

As a part of the research authors have developed a flexible, distributed computing system, intended for tomographic image reconstruction process. For both the linear and non-linear reconstruction algorithms parallel architecture developed by the authors is designed in such a way that it can be scaled to any number of computing nodes, assuming that the network medium is not a limiting factor. Performance tests have shown that the practical application of parallel algorithms executed on GPU allows for a 28-fold increase in the rate of performing calculations in the case of non-linear algorithms, compared to the optimized, sequential versions of algorithms.

Further research will also address challenges of calculations in heterogeneous and distributed environments. This work will be aimed at reducing delays, and therefore the response time of the system, due to the transmission of data over the network, as well as overall optimizations to increase the stability of the proposed solution. In addition, the authors are carrying out further work on the system, and the concept of complete reconstruction, utilizing both linear and non-linear algorithms, on the remote nodes. Ultimately, this will enable the transfer of all calculations to remote servers, connected to the data acquisition system over the Internet, thereby allowing monitoring and control of the industrial processes using smartphones or other mobile devices.

REFERENCES

[1] R. Banasiak, R. Wajman, T. Jaworski, P. Fiderek, H. Fidos, J. Nowakowski, D. Sankowski (2014), "Study on two-phase flow regime visualization and identification using 3D electrical capacitance tomography and fuzzy-logic classification," *International Journal of Multiphase Flow, Vol. 58*, 2014, pp. 1-14

[2] Chen, Ch., Woźniak P. W., Romanowski, A., Obaid, M., Jaworski, T., Kucharski, J., and Grudzień, K. and Zhao, S., Fjeld, M., "Using Crowdsourcing for Scientific Analysis of Industrial Tomographic Images," *ACM Trans. Intell. Syst. Technol., Vol. 7, No. 4*, 2016, ACM, pp. 52:1–52:25

[3] Garbaa, H., Jackowska-Strumiłło, L, Grudzień, K., Romanowski, A., "Neural network approach to ECT inverse problem solving for estimation of gravitational solids flow," *In Proc. of the 2014 Federated Conf. on Computer Science and Inf. Systems, AAIA'14, Vol. 2*, Warsaw, Poland, 2014, pp. 19-26

[4] Kapusta, P., Majchrowicz, M, "Accelerating Image reconstruction algorithms in Electrical Capacitance Tomography using Multi-GPU system," *Advanced Numerical Modelling, International Interdisciplinary PhD Workshop, Warsaw, Electrotechnical Institute*, 2011, pp. 47–49.

[5] Kapusta, P., Majchrowicz, M., Sankowski, D., Jackowska-Strumiłło, L., Banasiak, R., "Distributed multi-node, multi-GPU, heterogeneous system for 3D image reconstruction in Electrical Capacitance Tomography - network performance and application analysis," *Przegląd Elektrotechniczny, 89 (2 B)*, 2013, pp. 339-342.

[6] Majchrowicz, M., Kapusta, P., Banasiak, R. , "Applying parallel and distributed computing for image reconstruction in 3D Electrical Capacitance Tomography," *Zeszyty Naukowe AGH - Automatyka, Vol 14, Issue 3/2,* 2010, Kraków, Wydawnictwa AGH, pp. 711–722.

[7] Majchrowicz, M., Kapusta, P., Wąs, Ł., Wiak, S, "Application of General-Purpose Computing on Graphics Processing Units for Acceleration of Basic Linear Algebra Operations and Principal Components Analysis Method," *Man-Machine Interactions 3, Advances in Intelligent Systems and Computing Volume 242,* Springer International Publishing, 2014, pp. 519–527.

[8] Majchrowicz, M., Kapusta, P., Jackowska-Strumiłło, L., Sankowski, D., "Analysis of Application of Distributed Multi-Node, Multi-GPU Heterogeneous System for Acceleration of Image Reconstruction in Electrical Capacitance Tomography," *Image Processing & Communications, vol. 20, Issue 3*, 2015, pp. 5–14.

[9] Sankowski, D., Grudzień, K., Chaniecki, Z., Banasiak, R., Wajman, R., Romanowski, A., "Process tomograhy development at Technical University of Lodz," *Electrical Capacitance Tomography Theoretical Basis and Applications, edited by Dominik Sankowski and Jan Sikora,* Warszawa, 2010, pp. 70-95.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 685–694
DOI: 10.15439/2016F464
ACSIS, Vol. 8. ISSN 2300-5963

Influence of Locality on the Scalability of Method- and System-Parallel Explicit Peer Methods

Matthias Korch,
Thomas Rauber,
Matthias Stachowski
and Tim Werner
Department of Computer Science
University of Bayreuth
Email: {korch, rauber, matthias.stachowski, tim.werner}@uni-bayreuth.de

***Abstract*—Because the numerical solution of initial value problems (IVPs) of systems of ordinary differential equations (ODEs) can be computationally intensive, several parallel methods have been proposed in the past. One class of modern parallel IVP methods are the peer methods proposed by Schmitt and Weiner, some of which are publicly available in the software package EPPEER released in 2012. Since they possess eight independent stages, these methods offer natural parallelism across the method suitable for the typical numbers of CPU cores in modern multicore workstations. EPPEER is written in FORTRAN95 and uses OpenMP as parallel programming model.**

In this paper, we investigate the influence of the locality of memory references on the scalability of method- and system-parallel explicit peer methods. In particular, we investigate the interplay between the linear combination of the stages and the function evaluations by applying different program transformations to the loop structure and by evaluating their performance in detailed runtime experiments. These experiments point out that loop tiling is required to improve cache utilization while still allowing the compiler to vectorize along the system dimension.

To show that for certain classes of right-hand-side functions a stage-parallel execution is not optimal, and to enhance the scalability of the peer methods to core numbers larger than the number of stages of a method, system-parallel implementations have been derived. Runtime experiments show that there are IVPs for which these new implementations outperform stage-parallel implementations on numbers of cores less than or equal to the number of stages. Moreover, by exploiting the ability to utilize higher core numbers, higher speedups than the number of stages have been reached.

I. Introduction

THIS paper considers a class of parallel solution methods for initial value problems (IVPs) of systems of ordinary differential equations (ODEs), defined by

$$\mathbf{y}'(t) = \mathbf{f}(t, \mathbf{y}(t)), \quad \mathbf{y}(t_0) = \mathbf{y}_0, \quad t \in [t_0, t_e], \qquad (1)$$

where $\mathbf{f} : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n$ is the *right-hand-side function* defining the ODE system, $t \in \mathbb{R}$ is the *independent variable*, usually denoting "*time*", $\mathbf{y} : \mathbb{R} \to \mathbb{R}^n$ is the *solution function* to be computed within the *integration interval* $[t_0, t_e] \subset \mathbb{R}$, and $\mathbf{y}_0$ is the initial value, i.e., the value of $\mathbf{y}$ at time t_0.

Many ODE IVPs do not have an analytical solution and must be solved numerically. The classical numerical approach, which is also used by the class of methods considered in this paper, applies a time-stepping procedure that starts at t_0 and walks through the integration interval, computing a new approximation value $\mathbf{y}_m \approx \mathbf{y}(t_m)$ at each time step $m = 1, 2, \ldots$ until t_e is reached. A detailed treatment of the subject can be found in [1].

Because the numerical solution of ODE IVPs often is computationally intensive, several methods with potential for parallelism have been proposed. Usually, these methods are classified as exploiting *parallelism across the method*, *parallelism across the system*, and *parallelism across time* (also called *parallelism across the steps*), see [2] for an overview of classical parallel ODE methods.

This paper considers the explicit parallel two-step peer methods provided as part of the EPPEER software package [3]. Peer methods have been introduced by Schmitt and Weiner in 2004 [4]. The explicit methods included in EPPEER, which has been released in 2012, are described in [5], [6], [7]. These methods possess up to 8 independent stages, which can be computed in parallel on different cores. Hence, they are an example for methods providing parallelism across the method, which can be exploited without a (possibly difficult) parallelisation of the right-hand-side function. However, as the authors of EPPEER note, additional parallelisation across the system is possible for larger numbers of cores, but this functionality is currently not part of the EPPEER package and has not been investigated yet.

In their tests of the parallel performance of the explicit two-step peer methods, the authors of EPPEER notice that near-optimal speedups are possible for expensive right-hand-side functions, while the speedups observed for cheap right-hand-side functions were less satisfactory [6], [8]. They attribute this behavior to a part of the time step where a linear combination of vectors is computed [6], [8].

In this paper, we investigate the reasons for the lower performance of cheap right-hand-side functions in detail and identify the locality of memory references of the linear combination but also the interplay between the locality of the linear combination and the locality of the function evaluations as the main performance limiters. As part of this investigation, we apply several different program transformations to the loop structure of the peer methods and evaluate their effect on locality and scalability using runtime experiments on two

978-83-60810-90-3/$25.00

different hardware architectures. In particular, we make use of loop tiling to exploit both temporal and spatial locality while still enabling the compiler to vectorize the loops over the system dimension. In addition to the parallelization over the stages, we also investigate system-parallel implementations. The goal of this is not only to enable the use of larger numbers of processor cores, but primarily to compare the locality behavior and the scalability of a system-parallel and a stage-parallel execution on small numbers of cores. In particular, we are interested in the question whether a system-parallel or a stage-parallel execution is more efficient for IVPs with a cheap right-hand-side function.

The rest of the paper is structured as follows: Section II discusses related work. Section III describes the mathematical structure of peer methods. Section IV investigates the influence of the loop structure on locality and scalability. Section V considers the interplay of the function evaluations and the system-parallel execution. The last two sections conclude the article.

II. Background and Related Work

A. Parallel ODE Methods

The numerical solution of ODE IVPs can be computationally intensive. Therefore, many solution methods with potential for parallelism have been proposed. Most of the fundamental work on parallel ODE methods dates back to the 1980s and 1990s; an overview and further references can be found in [2].

Parallelism across time is generally difficult to obtain for IVPs because there needs to be an information flow from the initial value at t_0 to the end of the integration interval t_e. There are, however, promising approaches based on the Picard iteration, e.g., Parareal methods [9]. Most of the parallel ODE methods proposed concentrate on *parallelism across the method*, i.e., they provide a small number of independent coarse-grained computational tasks inherent in the computational structure of the method, for example, independent stages. Examples are Parallel Adams–Bashforth (PAB) and Parallel Adams–Moulton (PAM) methods [10], which belong to the class of *general linear methods* [1], parallel iterated RK (PIRK) methods [2], and extrapolation methods [1], [2]. Basically, all IVP methods possess a natural potential for *parallelism across the ODE system*, because usually the equations of the ODE system can be distributed to different processor cores. Of course, this approach is only feasible for reasonably large ODE systems.

One implementation strategy aiming at parallelism across the system is the use of parallel linear algebra operations which are parallelized along the system dimension. The project Odeint [11], for example, which is part of the Boost C++ library, contains several IVP methods and allows the use of different sequential or parallel state vector types. Another example of this type of parallelism is the PETSc library [12], which targets partial differential equations (PDEs), but also contains several ODE IVP solvers. One disadvantage of this approach is that parts of the code outside the linear algebra operations are not parallelized, and parallel performance tuning cannot be applied to the IVP solver as a whole.

System-parallel implementations covering all parts of a time step have been investigated in [13] for embedded RK methods and in [14] for PIRK methods.

Waveform relaxation methods [2] are specially tailored for a system-parallel execution of large ODE systems as they arise in the simulation of electrical circuits. They use the Picard iteration to decouple the equations of the ODE system and thus avoid synchronization and communication between cores. However, similar to Parareal methods, they have to deal with a possibly slow convergence of the Picard iteration.

The subject of this paper are explicit parallel two-step peer methods, which show similarities in their computational structure to PIRK methods and Parallel Adams methods. It investigates the locality and scalability of these methods when exploiting either method or system parallelism.

B. Performance Optimization

Computer systems are becoming more and more complex and make use of parallelism at various levels. To hide the increasing complexity, modern CPUs and compilers contain many mechanisms and techniques that aim at providing most of the available performance of the hardware in a transparent way to the application programmer. For example, CPUs contain multiple execution units, which can work in parallel, and they use dynamic instruction scheduling and *simultaneous multithreading* (SMT) to increase the utilization of the execution units. Modern compilers try to automatically unroll and vectorize loops to make use of SIMD (*single instruction multiple data*) extensions such as SSE and AVX [15].

Practically all modern CPUs used in high-performance computing possess a deep memory hierarchy with usually two or three levels of cache to temporarily store and reuse data read from or written to the external, slower main memory, thus hiding most part of the latency time that would otherwise be needed to access the main memory. In multi-core CPUs, the higher cache levels are often shared between several cores. To obtain maximum performance on a hardware platform with memory hierarchy, it is crucial to optimize the locality of memory references.

In the field of compiler design, efficient use of the memory hierarchy is tried to be achieved by reordering compute intensive loops in the source code. Loop tiling is considered as one of the most successful techniques. An important analytical model are *cache miss equations* [16], which can be used to estimate the effects of loop transformations. To analyze the dependencies of loops and to perform loop transformations, often the polyhedral model is used, e.g., [17]. A simple, but insightful visual model for the performance of computer programs on modern multi-core architectures is the *roofline model* [18]. A generalization of the roofline model to cover deep memory hierarchies is the *execution cache memory (ECM) model* [19].

To overcome the limits of static code analysis, some compiler-based approaches introduce source code annotations

[15] or propose domain specific languages (DSLs), which are then extended by *autotuning* techniques, e.g., [20]. Autotuning tries to determine the best configuration from a search space of possible code variants and parameters (e.g., block sizes for loop tiling, loop unroll factors, or number of threads). In [21], an online autotuning approach for system-parallel PIRK methods has been investigated.

III. Explicit Parallel Peer Methods in EPPEER

In this section, we will explain the computational structure of the explicit parallel two-step peer methods and describe their implementation in EPPEER. Then, we will investigate locality and scalability of the unmodified EPPEER code by several runtime experiments.

A. Computational Structure

Similar to classical RK methods, the explicit two-step peer methods in EPPEER use a time-stepping procedure to solve the IVP and compute s stages $\mathbf{Y}_{m,i}$, $i = 1, \ldots, s$, at each time step from t_{m-1} to t_m with $t_m = t_{m-1} + h_{m-1}$ for $m = 1, 2, \ldots$ [7]. However, in contrast to classical RK methods, all s stages have the same accuracy and stability properties and, to compute the stages, the s stages and the s function values of the previous time step are used:

$$\mathbf{Y}_{m,i} = \underbrace{\sum_{j=1}^{s} b_{ij} \mathbf{Y}_{m-1,j}}_{\mathbf{S_Y}} + \underbrace{h_m \sum_{j=1}^{s} a_{ij} \mathbf{f}(t_{m-1,j}, \mathbf{Y}_{m-1,j})}_{\mathbf{S_f}}, \quad (2)$$

where $t_{m-1,j} = t_{m-1} + h_{m-1} c_j$ for $j = 1, \ldots, s$, and a_{ij}, b_{ij}, and c_j, $i, j = 1, \ldots, s$ with $c_s = 1$ are the coefficients of the particular peer method. Therefore, these methods are a subclass of general linear methods (GLMs) [1].

Because only values from the previous time step are used, the s stages of the current time step do not depend on each other and can be computed in parallel on different cores, thus exhibiting parallelism across the method. In EPPEER, three methods with $s = 4, 6, 8$ are included which can use up to s cores. These methods have been introduced in [6]. In addition, four FSAL methods (first same as last: last stage of the previous time step is reused as the first stage of the current time step) with $s = 3, 5, 7, 9$, which have been introduced in [5], are included, which can use up to $s - 1$ cores.

Since all s stages have the same properties, in particular the same order $O(h_m^s)$, there is no determined value for $\mathbf{y}_m \approx \mathbf{y}(t_m)$. Instead, any stage value can be used as new approximation to the solution function.

Due to the two-step character, a starting procedure is required to generate $s-1$ stages in addition to the initial value $\mathbf{y}_0$ before the first time step can be performed with the peer method to compute $\mathbf{Y}_{1,i}$. Currently, EPPEER uses an explicit RK method (DOPRI 5(4)) to compute the start values. However, a parallel starting procedure has been proposed in [22].

```
!$OMP PARALLEL DEFAULT(SHARED)
!$OMP DO PRIVATE(stg,ic) SCHEDULE(STATIC)
     do stg = ist1,stages
       pyy(:,idx(new+stg)) = 0.D0
       do ic = 1,stages
         pyy(:,idx(new+stg)) = pyy(:,idx(new+stg))+
          & pa(stg,ic)*ff(:,idx(ic))
       end do
     end do
!$OMP END DO
!$OMP DO PRIVATE(stg,ic) SCHEDULE(STATIC)
     do stg = ist1,stages
       do ic = 1,stages
         pyy(:,idx(new+stg)) = pyy(:,idx(new+stg))+
          & pb(stg,ic)*pyy(:,idx(ic))
       end do
     end do
!$OMP END DO
!$OMP DO SCHEDULE(STATIC)
     do stg = ist1,stages
        call fcn(t+phs*pc(stg),pyy(:,idx(new+stg)),
          &ff(:,idx(new+stg)),cpar)
     end do
!$OMP END DO
!$OMP END PARALLEL
```

Listing 1. Ver-0: loop structure used in EPPEER.

B. Implementation

The following section introduces the implementation of the original EPPEER package. The source code of this implementation can be seen in Listing 1. Each time step of a peer method consists of two basic parts: The linear combination and the function evaluation.

EPPEER implements those two basic parts by three consecutive loop nests: The first and the second loop nest perform the linear combination. For this purpose the first loop nest initializes an $s \times n$ accumulation matrix for the computation of $\mathbf{Y}_{m,i}$ with zeroes, i.e.,

$$\text{for } i = 1, \ldots, s: \quad \mathbf{Y}_{m,i} \leftarrow \mathbf{0} \quad (3)$$

and uses this matrix to accumulate and add the second sum of Eq. (2), $\mathbf{S_f}$:

$$\text{for } i = 1, \ldots, s: \quad \mathbf{Y}_{m,i} \leftarrow \mathbf{Y}_{m,i} + \underbrace{h_m \sum_{j=1}^{s} a_{ij} \mathbf{F}_{m-1,j}}_{\mathbf{S_f}}, \quad (4)$$

where

$$\mathbf{F}_{m-1,j} = \mathbf{f}(t_{m-1,j}, \mathbf{Y}_{m-1,j}) \quad (5)$$

are stored function values computed in the previous time step. After that the second loop nest accumulates and adds the first sum of Eq. (2), $\mathbf{S_Y}$:

$$\text{for } i = 1, \ldots, s: \quad \mathbf{Y}_{m,i} \leftarrow \mathbf{Y}_{m,i} + \mathbf{S_Y}. \quad (6)$$

The third and final loop nest performs the function evaluation to compute $\mathbf{F}_{m,j}$, needed in the next time step, using Eq. (5).

The outermost loop of each loop nest iterates over the stages of the peer method, where the ith iteration of the outer loop computes the argument vector $\mathbf{Y}_{m,i}$ for stage i. Thus, we will refer to these loops as "stage loops". The iterations of the stage loops are independent of each other. That is why the stages

can be computed in parallel for each loop nest. This is a major advantage of the peer methods. The EPPEER package takes this advantage by parallelizing each stage loop with OpenMP.

The stage loops of the first and the second loop nest contain inner loops, which also iterate over the stages in order to compute the linear combinations $\mathbf{S_f}$ and $\mathbf{S_Y}$, respectively, for the stage i corresponding to the current iteration of the outer stage loop. Thus, we will refer to these loops as "combination loops".

The combination loops perform operations on vectors of dimension n and, thus, contain an innermost loop, which iterates over the system dimension and which we therefore call "system loop". Instead of explicitly implementing these system loops as FORTRAN `do` loops, EPPEER uses the FORTRAN vector notation, thus making it easy for the compiler to generate vectorized code for these loops.

C. Performance

In this section, we use the results of runtime experiments to analyze the scalability of the original EPPEER package with the number of cores. The first target system for these experiments is an 8 core Intel Xeon E5-2630V3 (Haswell-EP) CPU. However, the Haswell-EP CPU has two features which might alter the scalability in an unexpected way: Turbo Boost (on-demand increase of clock frequency) and Hyper Threading (2-way simultaneous multithreading (SMT) per core). To avoid potential negative influences of these features on our measurement results, we disabled both features for all measurements. Our second target system is an Intel Xeon Phi 31S1 coprocessor. The Xeon Phi does not have a turbo mode, but it provides 4-way SMT per core, which cannot be disabled. However, to ensure that each thread runs on a separate physical core, we make use of the OpenMP runtime environment to distribute the threads evenly among the cores until the number of threads exceeds the total number of cores. On both target systems, the Intel FORTRAN compiler in version 16.0.2 was used with optimization level 2 (`-O2`) to compile the EPPEER package. All computations were performed using double precision. For profiling purposes we use PAPI in Version 5.4.1. PAPI is a powerful profiling library, which can read the values of CPU-internal performance counters, which can count, for example, the number of cache misses or the number of load/store operations.

We chose two different ODE systems as test problems: The first one is the 2D Brusselator *brus*. It models an oscillating chemical reaction-diffusion system using a two dimensional grid as spatial discretization. The resulting access pattern to the grid points is a five-point 2D stencil. Therefore, *brus* is a sparse problem and its right-hand-side function $\mathbf{f}$ has a time complexity of $\Theta(n)$. For all measurements with *brus*, we chose a vector size of 500 000 elements, where each vector element is a double precision floating point number, requiring 8 bytes of storage space. Thus, a vector corresponds to a grid of 500×500 cells, where each cell contains two double values, so that one vector requires 3.81 MB of storage space. The second test problem is the N-body problem *mbod*, which is also known as the many body problem. It simulates the movement of particles, which interact with each other by the gravitational force. Since the gravitational field of a particle influences every other particle, the N-body problem is a dense problem and its right-hand-side function $\mathbf{f}$ has a time complexity of $\Theta(n^2)$. To obtain the experimental results shown in the following, a scene consisting of 2000 particles was used, where each particle added 6 equations to the ODE system, resulting in a vector size of 12 000 elements (93.75 KB). Table I shows the speedups observed in our experiments.

TABLE I
SPEEDUPS MEASURED WITH THE EPPEER PACKAGE.

		Speedups for different #threads				
Hardware	**Problem**	1	2	3	4	8
Haswell-EP	*brus*	1	1.839	2.006	2.232	2.311
Haswell-EP	*mbod*	1	1.961	2.582	3.779	7.047
Xeon Phi	*brus*	1	1.941	2.619	3.609	6.702
Xeon Phi	*mbod*	1	1.966	2.594	3.804	7.133

Similar to the authors of the EPPEER package, our measurements with the *mbod* problem show a very good scalability. The use of 8 threads yields an almost ideal speedup of 7.0 on Haswell-EP and a speedup of 7.1 on the Xeon Phi. In contrast, the best speedups measured for the *brus* problem are only 2.3 and 6.7 on the Haswell-EP and the Xeon Phi, respectively (on both systems obtained using 8 threads). Thus, obviously, the scalability of the peer methods is influenced by the IVP to be solved.

Interestingly, Table I shows a smaller efficiency when 3 threads are used, which is caused by the following: It is not possible to distribute s equally sized tasks among p threads evenly if $s \bmod p \neq 0$. In our case it is impossible to distribute the 8 stages equally among 3, 5, 6 and 7 threads. This load imbalance forces some of the threads to idle while all other threads complete their last remaining task.

One apparent difference between the two test problems is the access pattern of the right-hand-side function $\mathbf{f}(t_{m,j}, \mathbf{Y}_{m,j})$ to the argument vector $\mathbf{Y}_{m,j}$ and the resulting time complexity. While the dense *mbod* problem has a time complexity of $\Theta(n^2)$, the time complexity of the sparse *brus* problem is only $\Theta(n)$. We can therefore expect that for *brus* the time needed to perform the function evaluations constitutes a smaller fraction of the runtime of a time step than for the *mbod*. In fact, further measurements (Figure 1) have shown that on average evaluating the right hand side of *mbod* takes about 97 % of the total runtime of a time step, whereas the linear combination only takes the remaining 1 to 2 %. In contrast, for *brus*, less than 18 % of the total runtime of a time step is required to compute the function evaluations, but nearly 75 % are needed to perform the linear combination to compute the argument vectors $\mathbf{Y}_{m,j}$.

In total, the results of the runtime experiments show that the scalability depends on the IVP to be solved. For sparse and computationally inexpensive systems like *brus*, computing the linear combinations takes a major part of the runtime (see Figure 1 (b)) and can therefore be expected to have

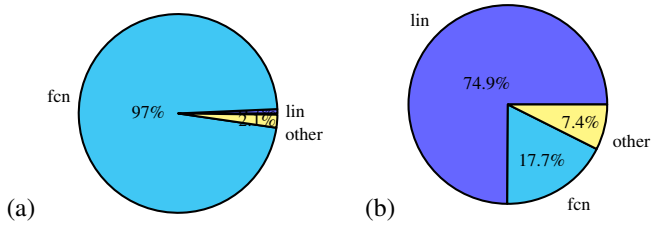


Fig. 1. Average runtime fractions of the linear combination (lin), the function evaluation (fcn) and other parts of the code for (a) *mbod* and (b) *brus*.

a significant influence on scalability. Since the arithmetic intensity, i.e., the ratio of the number of arithmetic instructions to the number of memory instructions, of this part of the time step is low, we can expect the performance of the linear combination to be bound by the performance of the cache and memory subsystem. Moreover, the working set of one time step, i.e., the amount of data accessed in a time step, amounts to four $s \times n$ matrices, which corresponds to 122.1 MB in our experiments and, thus, is significantly larger than the cache size of typical workstations today. We can therefore expect that many expensive accesses to the main memory are required in the experiments with *brus*.

However, for dense and computationally intensive systems like *mbod*, the evaluation of the right-hand-side function of the ODE system dominates the runtime. Since the function evaluations are independent of each other, this leads to the high scalability observed. For the problem size we used, the working set of one time step amounts to 2.92 MB, that means all data accessed during one time step fits completely in the 20 MB L3 cache of the Haswell-EP. The 512 KB L2 cache of a Xeon Phi core can not store the whole working set of a time step, but, to perform the stage-parallel function evaluation efficiently, it is only important that it can store two n-vectors (93.75 KB each) to hold the argument of the function evaluation and its result. The reason for this is that to compute the gravitational force acting on one body the interacting forces with all other $N-1$ bodies are accumulated. Thus, to compute all entries of the function result $\mathbf{f}(t_{m,j}, \mathbf{Y}_{m,j})$, half of the argument vector $\mathbf{Y}_{m,j}$ is read $n/2$ times.[1] Since all important data fits in the cache and the arithmetic intensity is higher than that of *brus*, we can expect the performance of the *mbod* problem to be bound by the compute performance of the processor used. Therefore, improving the locality of memory references further would probably not improve the already high scalability.

In the following, we therefore focus on the sparse case, where an improvement of the scalability is more important and where we can expect to be able to improve the scalability by an improvement of the locality behavior. As test problem we consider only *brus* from now on.

[1] *mbod* is a first order system derived from a second order system by substitution.

IV. Loop Structure of the Linear Combination

A. Variants of the Loop Structure

After the identification of the linear combination as the major fraction of the runtime for sparse ODE systems in the previous section, this section focuses solely on the improvement of the loop structure of the linear combination, possible loop transformations, and their influence on the resulting locality and scalability. The interplay with the evaluation of the right-hand-side function (the third loop nest in the original EPPEER package) will be considered afterwards in Section V.

The data access pattern of the original stage-parallel EPPEER loop structure (referred to as "Ver-0" in the following) is illustrated in Figure 2 for a method with 8 stages and one thread per stage. Only the first two threads are shown, because the data access pattern is similar for all threads: to compute the argument vector $\mathbf{Y}_{m,i}$ of their stage i, the whole two $s \times n$ matrices Y_{m-1} and F_{m-1} are read, and after each loop nest the result of the computations within the loop nest are written back to memory. Hence, in total there are $2s+1$ write accesses to each element of the matrix Y_m: initialization with zero, accumulation of sum $\mathbf{S_Y}$ and accumulation of sum $\mathbf{S_Y}$. Though the stages are computed in parallel by different threads, in most current shared-memory computers several threads share parts of the memory subsystem (e.g., shared higher level caches, main memory modules connected to a socket) so that they compete for these limited resources and may quickly reach their limits.

Since the system dimension is the stride-1 dimension, the innermost loops of Ver-0 iterate over the stride-1 dimension, which leads to high spatial locality and allows the compiler to vectorize the loads and stores using sequential SIMD load/store instructions.

The first improvement of the loop structure we consider can be seen in Listing 2 and will be referred to as "Ver-1" in the following. Ver-1 adopts the parallelization across the stages from Ver-0, but it contains two modifications. First, it eliminates the zero initialization of Y_m by peeling off the first iteration of the combination loop. This saves one pass over the matrix Y_m and, thus, many expensive memory accesses. Moreover, the first and the second loop nest of Ver-0 are fused, which both contribute to the computation of $\mathbf{Y}_{m,i}$, to a single loop nest. This loop fusion is legal, because the only dependencies between those two loop nests are the write accesses to the matrix Y_m accumulating the sums in Eq. 2. That is why changing the order of the loops, and thus the order of the accumulation, only influences the round-off error. An advantage of this new fused loop structure is that only s write accesses to each element of Y_m are necessary.

Trying to overcome some disadvantages of Ver-0 and Ver-1, Ver-2 (Listing 3) parallelizes the linear combinations across the system dimension. This version is based on an earlier version of the EPPEER package. It is obtained by interchanging the loops of each loop nest of Ver-0 so that the outermost loop iterates over the system dimension and the two inner loops iterate over the stages. To enable this interchange,

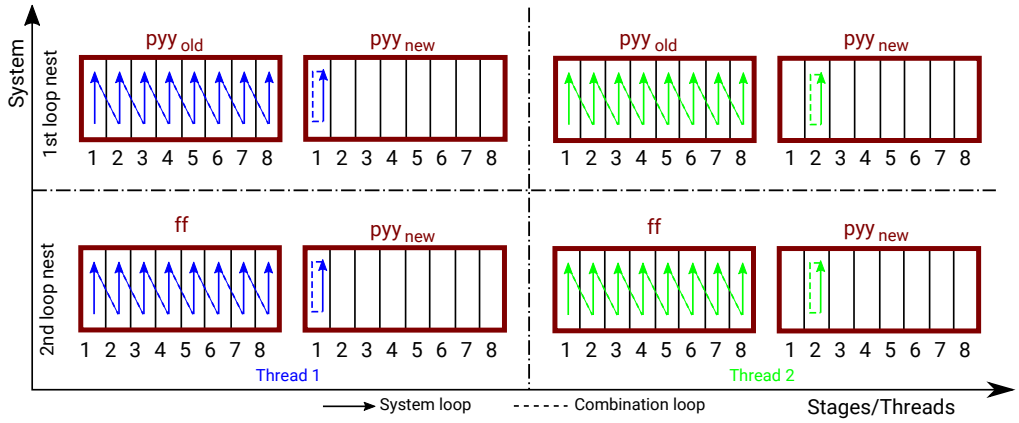


Fig. 2. Data access pattern of the original EPPEER package (Ver-0).

```
!$OMP PARALLEL DEFAULT(SHARED)
!$OMP DO PRIVATE(stg,ic) SCHEDULE(STATIC)
    do stg = ist1,stages
      pyy(:,idx(new+stg)) = pa(stg,1)*ff(:,idx(1)) +
         & pb(stg,1)*pyy(:,idx(1))
      do ic = 2,stages
        pyy(:,idx(new+stg)) = pyy(:,idx(new+stg)) +
           & pa(stg,ic)*ff(:,idx(ic)) + pb(stg,ic) *
           & pyy(:,idx(ic))
      end do
    end do
!$OMP END DO
!$OMP END PARALLEL
```

Listing 2. Ver-1: Ver-0 improved by loop peeling and fusion.

```
!$OMP PARALLEL DEFAULT(SHARED)
!$OMP DO PRIVATE(id,stg,ic) SCHEDULE(STATIC)
    do id = 1,nprob
      do stg = ist1,stages
        pyy(id,idx(new+stg)) = 0.D0
        do ic = 1,stages
         pyy(id,idx(new+stg)) = pyy(id,idx(new+stg))+
           &  pa(stg,ic)*ff(id,idx(ic))
        end do
      end do
    end do
!$OMP END DO
!$OMP DO PRIVATE(id,stg,ic) SCHEDULE(STATIC)
      do id = 1,nprob
        do stg = ist1,stages
          do ic = 1,stages
           pyy(id,idx(new+stg)) = pyy(id,idx(new+stg))+
             & pb(stg,ic)*pyy(id,idx(ic))
          end do
        end do
      end do
!$OMP END DO
!$OMP END PARALLEL
```

Listing 3. Ver-2: System parallel execution.

the previously innermost system loops, which were defined implicitly using vector notation, had to be transformed into an explicit FORTRAN do loop. As a result, a higher temporal locality for the write accesses to the elements of the matrix Y_m is generated. The compiler can now keep the temporary partial sums in a register during the s^2 iterations over the stages by the inner loops of each loop nest, so that now each element of Y_m needs only be written to main memory twice.

Even more important, the amount of data accessed by each thread is reduced significantly: While in Ver-0 and Ver-1 each thread needs to read all elements of the matrices Y_{m-1} and F_{m-1} to compute the stage(s) assigned to it, in Ver-2 it reads only the range of elements of Y_{m-1} and F_{m-1} assigned to it and it writes only the corresponding range of elements of Y_m. Hence, if there are p threads, each thread only reads at most $2s\lceil\frac{n}{p}\rceil$ elements and writes at most $s\lceil\frac{n}{p}\rceil$ elements, whereas in Ver-0 and Ver-1 each thread reads $2sn$ elements and writes at most $\lceil\frac{s}{p}\rceil$ elements.

While Ver-1 can use only as many threads as the peer method used has stages, an additional benefit of the system parallel execution is that more threads can participate in the computation of the linear combination.

Unfortunately, the new loop structure of Ver-2 also has disadvantages. Since the iteration over the stride-1 dimension is performed by the outermost loop, the reads of the matrix elements in the innermost loops have a large stride of n. Hence, the higher temporal locality comes at the expense of lower spatial locality. Further, the compiler cannot vectorize the loads and stores using sequential SIMD load/store instructions. However, there are SIMD gather or scatter instructions, which make a vectorization of strided memory accesses possible. Although all modern Intel CPUs since Haswell support AVX2 gather instructions, our Haswell-EP still emulates those instructions by microcode. Thus, gather instructions still have a low throughput on Haswell-EP.

In order to derive a loop structure with high temporal as well as spatial locality, we can make use of loop tiling (Ver-3, Listing 4). Ver-3 also makes use of both optimizations from Ver-1 (elimination of the zero initialization phase and loop fusion) yielding a similar loop structure as in Ver-1. However, in contrast to Ver-1, it has an outer parallel tile loop, which iterates over the system dimension in larger steps dividing the system into tiles of a user-defined size. Inside the tile loop run the stage loop and the combination loop. The innermost loop is the intra-tile loop, which iterates over the elements of the

```
!$OMP PARALLEL DEFAULT(SHARED)
!$OMP DO PRIVATE(stg,ic, idi) SCHEDULE(STATIC)
do id = 1,nprob, tile_size
  do stg = ist1,stages
    do idi = id, min(id+tile_size-1,nprob)
      pyy(idi,idx(new+stg)) = pb(stg,1)*pyy(idi,idx(1))+
        & pa(stg,1)*ff(idi,idx(1))
    end do
    do ic = 2,stages
      do idi = id, min(id+tile_size-1,nprob)
        pyy(idi,idx(new+stg)) = pyy(idi,idx(new+stg))+
          & pb(stg,ic)* pyy(idi,idx(ic)) +
          & pa(stg,ic) * ff(idi,idx(ic))
      end do
    end do
  end do
end do
!$OMP END DO
!$OMP END PARALLEL
```

Listing 4. Ver-3: Ver-2 improved by loop peeling, fusion, and tiling.

TABLE II
RUNTIME COMPARISON FOR 100 TIME STEPS ON HASWELL-EP AND XEON PHI WITH GRID SIZE 500×500

		Runtime for different #threads				
Hardware	**Version**	1	4	8	60	120
Haswell-EP	Ver-0	10.73	4.98	5.02	-	-
Haswell-EP	Ver-1	8.64	4.00	3.45	-	-
Haswell-EP	Ver-3	2.53	0.95	0.76	-	-
Xeon Phi	Ver-0	50.01	13.06	6,69	-	-
Xeon Phi	Ver-3	19.54	4.90	2.47	0.39	0.29

corresponding tile, i.e., the stride-1 dimension. The resulting memory access pattern is illustrated in Figure 3. Only the same amount of data needs to be accessed as in Ver-2, but, since the tile loop now contains the stage loop and the combination loop, an iteration of the tile loop computes the linear combination for one tile for all the stages. This provides a blocking effect across the stages: Assuming the tile size is small enough so that the data accessed in one iteration of the tile loop fits in the cache, the final values of the elements of Y_m have to be sent to main memory only once. Since the innermost loop iterates over the stride-1 dimension, there also is a high spatial locality, so that cache lines can be reused once they have been loaded into the cache, and efficient SIMD vectorization is possible. Because of that, choosing a good tile size for Ver-3 is important.

B. Performance of the Loop Structure Variants

In this section we will analyze how the different versions so far scale with the amount of active cores. For this purpose we have measured the runtimes of the different versions on the Haswell-EP and on the Xeon Phi. The runtimes measured are given in Table II. The scaling behavior is shown in Figures 4 and 6 in terms of the inverted runtime (reciprocal of the runtime) as a function of the number of threads. As in a speedup diagram, a linear growth of the inverted runtime corresponds to a good scalability, but, in contrast to a speedup diagram, there is no need to chose a sequential reference runtime.

We can divide the versions so far into two groups: The first group consists of the original EPPEER version (Ver-0) and Ver-1, which both only utilize stage parallelism, while the second group consists of Ver-2 and Ver-3, which both only utilize system parallelism for the linear combination.

Because the peer methods part of EPPEER possess at most 8 stages, the stage-parallel variants use up to 8 cores only. Furthermore, there is a load imbalance when 3, 5, 6, or 7 threads are used. Ver-1 of the first group, which improves the original EPPEER version by removing the zero initialization and fusing the two loop nests, is about 1.6 times faster than the original EPPEER version.

In contrast to the stage-parallel variants, the system-parallel variants can potentially use more than 8 cores and are not affected by the load imbalance. The cause of this is the system dimension offering a higher degree of parallelism. Ver-2 uses system parallelism without loop tiling. That is why it eventually became up to 3 times faster than the best stage-parallel variant on the Haswell-EP. However, when Ver-2 was used on less than 4 cores, it was slower than the stage-parallel variants.

Unlike Ver-2, Ver-3 achieves system parallelism with an efficient memory access pattern by utilizing loop tiling. Preparatory runtime measurements had shown that a tile size of about 128 was best for almost all problem sizes. For the following experiments, therefore 128 was used as tile size for Ver-3. Figure 5 shows the normalized runtime of Ver-3 with different tile sizes for a grid size of 500×500 and 700×700 and for one core and eight cores on the Haswell-EP. For both problem sizes and both core numbers there is a runtime minimum between tile size 64 and 256.

Because of its more efficient memory access pattern, Ver-3 was about three times faster than Ver-2 running on the same amount of cores on the Haswell-EP. Ver-3 was also about 8 times faster than the best stage-parallel variant, if both used the same amount of cores. This suggests Ver-3 also having a more cache friendly memory access pattern than the stage-parallel variants.

On the Xeon Phi also Ver-3 obtains the best runtime. On this processor we could observe a reduction of the runtime for up to 120 threads, where for large numbers of threads Ver-3 clearly outperforms Ver-2.

To measure the locality behavior, we measure the normalized runtime (Figures 7 and 8), i.e., the runtime per time step divided by the number of equations, n, on the Haswell-EP. Since the right-hand-side of *brus* has costs $\Theta(n)$, an increase in the normalized runtime usually indicates working sets falling out of a cache level. As we can see, the normalized runtime of Ver-3 is significantly lower than that of the other versions, and it does not increase as strongly when the problem size exceeds a certain threshold, which depends on the number of threads.

In addition, we measured the L3 cache misses and the total amount of store and load operations. The measurements confirm that Ver-3 has a smaller L3 cache miss rate in relation to the total load/store operations than both stage-parallel versions (see Figure 9). Furthermore, this plot shows also that the L3 cache miss rate is much lower on small problem sizes.

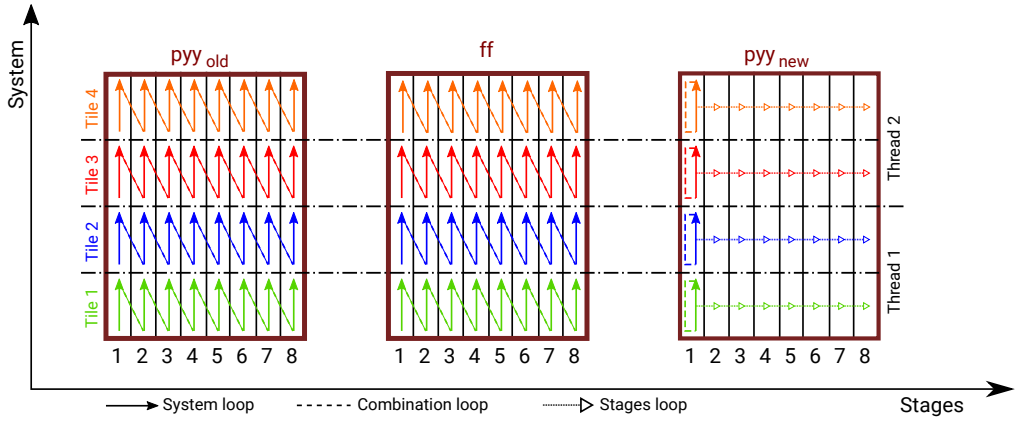


Fig. 3. Data access pattern of the optimized system-parallel loop structure (Ver-3).

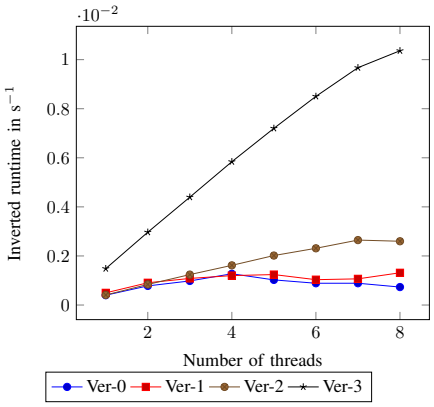


Fig. 4. Inverted runtime of different versions of the linear combination of *brus* on Haswell-EP in seconds.

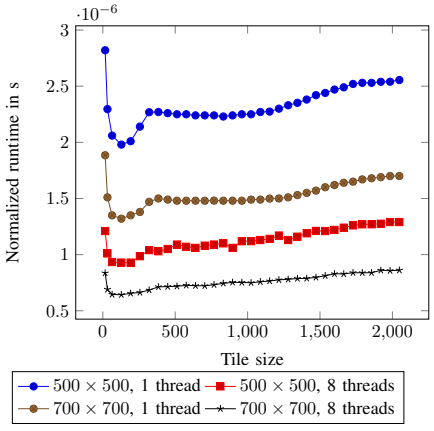


Fig. 5. Normalized runtime of Ver-3 in seconds for different tile sizes (x-axis), different grid sizes and numbers of threads on Haswell-EP.

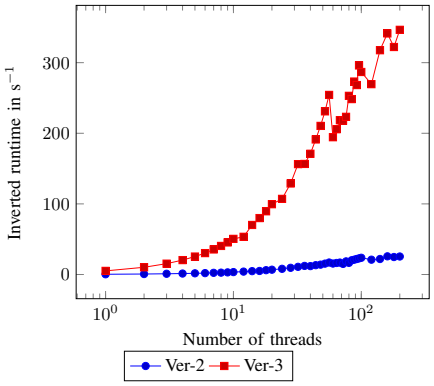


Fig. 6. Inverted runtime of system-parallel versions of the linear combination of *brus* on Xeon Phi for one time step in seconds.

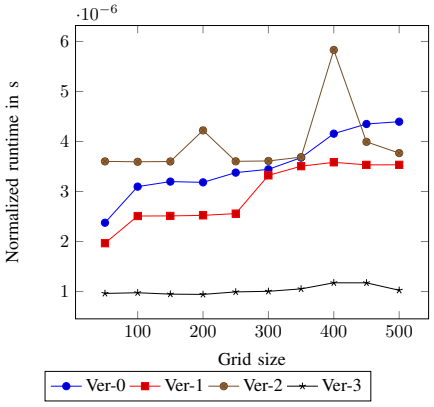


Fig. 7. Normalized runtime of the linear combination of *brus* on Haswell-EP in seconds for 1 thread.

V. Interplay with the Function Evaluation and Fully System-Parallel Execution

In the last section we have only focused on optimizing the linear combination without modifying the loop nest, which evaluates the problem function. In this section, however, we will improve the problem function evaluation and the

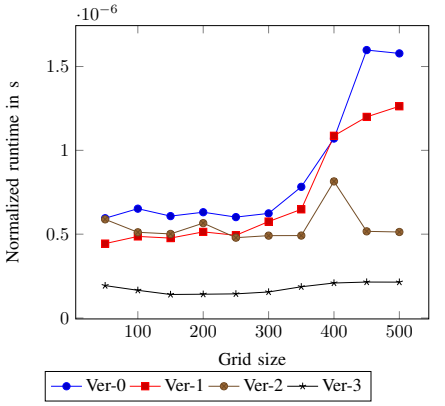


Fig. 8. Normalized runtime of the linear combination of *brus* on Haswell-EP in seconds for 8 thread.

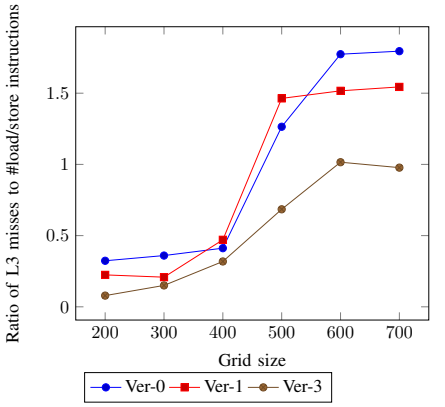


Fig. 9. Ratio of L3 misses to #load/store instructions for 8 threads and variable grid size.

interaction of the problem function with the linear combination.

The runtime experiments of the previous section have confirmed that an optimized system-parallel tiled loop variant of the linear combination such as Ver-3 outperforms stage parallel variants, because it requires less memory references and can efficiently exploit temporal and spatial locality. However, up until now all implementation variants evaluated the problem function using parallelism across the stages.

Unfortunately, combining the system-parallel linear combination with a stage-parallel function evaluation has several disadvantages: The stage-parallel function evaluation can use as many cores as there are stages only. Furthermore both a system parallel linear combination and method parallel function evaluation have different memory access patterns, which causes a data redistribution and reduces locality. That is why we modify the loop nest containing the problem function to adopt system parallelism. However, up to now the problem function has been evaluated by a single function call that computed the temporal derivatives for every element of a given vector, which is not suited for system parallelism.

Therefore, we implemented two new versions of the *brus*

```
!$OMP PARALLEL DEFAULT(SHARED)
!$OMP DO PRIVATE(stg) SCHEDULE(STATIC)
 do id = 1,nentries,tile_size
   do stg = ist1,stages
     do idi = id, min(id+tile_size,nentries)
       call fcn(idi,t+phs*pc(stg),pyy(:,idx(new+stg)),
            & ff(:,idx(new+stg)),cpar)
     end do
   end do
 end do
!$OMP END DO
!$OMP END PARALLEL
```

Listing 5. Elementwise function evaluation.

```
do stg = ist1,stages
!$OMP PARALLEL DEFAULT(SHARED)
!$OMP DO SCHEDULE(STATIC)
  do i = 1,nentries,block_size
     call fcn(i,block_size,t+phs*pc(stg),
       & pyy(:,idx(new+stg)), ff(:,idx(new+stg)),cpar)
  end do
!$OMP END DO
!$OMP END PARALLEL
end do
```

Listing 6. Blockwise function evaluation.

problem with a modified signature and a modified internal structure: The first new version (*element* version) calculates the temporal derivatives for one single element of the problem vector per call. Yet again, this design has two major disadvantages: The *element* version has to perform the boundary checks of the stencil for each call, and the compiler is not able to vectorize this function efficiently. Moreover the temporal derivatives of the two substances contained in a grid cell have some computations in common, which the *element* version has to compute redundantly twice. We can avoid all those inefficiencies if we do not call the right-hand-side function once per element, but once for a range of elements, and optimize the internal structure of the function accordingly. This results in the implementation of the *block* version. Finally, we have adapted the loop nest evaluating the problem functions to the *element* version in Listing 5 and to the *block* version in Listing 6. Since the *element* version cannot provide any blocking itself, we have added a simple 1D-blocking scheme to the loop nest.

Inverted runtimes measured for the two new variants of *brus* are shown in Figures 10 and 11. Here, the grid size used is 500×500, the number of stages is set to 8, and 200 is used as block size for the *block* version. As expected, both new versions introduced in this section are faster. The *block* version has the best runtime. The *element* version first starts slower than the original implementation, because this version invokes the function evaluation n times, each time performing a test whether the current index corresponds to a boundary point or not. This leads to a high overhead. But for large numbers of cores, the higher locality of the *element* version can compensate this. In particular, it can obtain speedups higher than the number of stages available.

Hence, all in all, the best runtime for *brus* is obtained by the system-parallel linear combination with loop tiling, loop

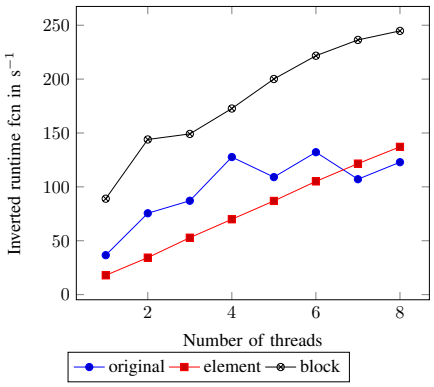


Fig. 10. Inverted runtime for different function evaluations (row, element, block) on Haswell-EP for one function invocation using Ver-3.

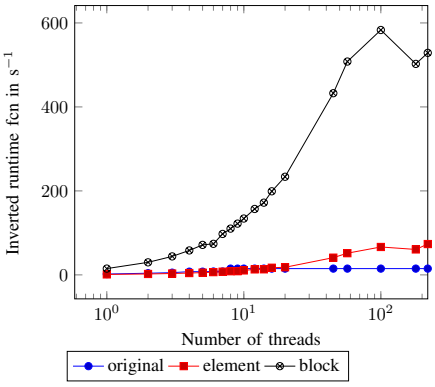


Fig. 11. Inverted runtime for different function evaluations (row, element, block) on Xeon-Phi for one function invocation using Ver-3.

fusion and loop peeling combined with the system-parallel blockwise function evaluation.

VI. Conclusions

In this paper, we have considered the influence of locality on the scalability of method- and system-parallel explicit peer methods. In particular, we have focused on improving the scalability for large sparse ODE systems with cheap function evaluation costs. After confirming that the linear combination to compute the argument vectors requires the major part of the runtime for such ODE systems, we have considered several transformations of the loop structure, analyzed their memory access pattern and measured the resulting runtimes and scalability on a Haswell-EP and on a Xeon Phi processor. As expected, the results of the runtime experiments confirmed that a system-parallel computation of the linear combination leads to a better performance than a stage-parallel computation because less memory references are required and a more cache-efficient data access pattern can be employed. A further performance improvement is possible when also the function evaluation can be performed in a system-parallel way, so that no data-redistribution is necessary.

Acknowledgment

We would like to thank Bernhard Schmitt for providing the EPPEER package and Simon Melzner for hardware support.

References

[1] E. Hairer, S. P. Nørsett, and G. Wanner, *Solving Ordinary Differential Equations I: Nonstiff Problems*, 2nd ed. Berlin: Springer, 2000.
[2] K. Burrage, *Parallel and Sequential Methods for Ordinary Differential Equations*. New York: Oxford University Press, 1995.
[3] B. A. Schmitt, "Peer methods for ordinary differential equations," http://www.mathematik.uni-marburg.de/~schmitt/peer/, last checked 2015-08-17.
[4] B. A. Schmitt and R. Weiner, "Parallel two-step W-methods with peer variables," *SIAM J. Numer. Anal.*, vol. 42, no. 1, pp. 265–282, 2004.
[5] B. A. Schmitt, R. Weiner, and S. Beck, "Two-step peer methods with continuous output," *BIT Numer. Math.*, vol. 53, pp. 717–739, 2013.
[6] B. A. Schmitt, R. Weiner, and S. Jebens, "Parameter optimization for explicit parallel peer two-step methods," *Appl. Numer. Math.*, vol. 59, pp. 769–782, 2009.
[7] R. Weiner, K. Biermann, B. A. Schmitt, and H. Podhaisky, "Explicit two-step peer methods," *Comput. Math. Appl.*, vol. 55, no. 609–619, 2008.
[8] B. A. Schmitt and R. Weiner, *Manual for explicit parallel peer code EPPEER*, Aug. 2012.
[9] Y. Maday and G. Turinici, "A parareal in time procedure for the control of partial differential equations," *C.R.A.S. Sér. I Math*, vol. 335, pp. 387–391, 2002.
[10] P. J. van der Houwen and E. Messina, "Parallel Adams methods," *J. Comput. Appl. Math.*, vol. 101, pp. 153–165, Jan. 1999.
[11] K. Ahnert, D. Demidov, and M. Mulansky, "Solving ordinary differential equations on gpus," in *Numerical Computations with GPUs*, V. Kindratenko, Ed. Springer, 2014, ch. 7, pp. 125–157.
[12] S. Balay, W. D. Gropp, L. C. McInnes, and B. F. Smith, "Efficient management of parallelism in object oriented numerical software libraries," in *Modern Software Tools in Scientific Computing*, E. Arge, A. M. Bruaset, and H. P. Langtangen, Eds. Birkhäuser Press, 1997, pp. 163–202.
[13] M. Korch and T. Rauber, "Parallel low-storage Runge-Kutta solvers for ODE systems with limited access distance," *Int. J. High Perf. Comput. Appl.*, vol. 25, no. 2, pp. 236–255, 2011. doi: 10.1177/1094342010384418
[14] ——, "Locality optimized shared-memory implementations of iterated Runge-Kutta methods," in *Euro-Par 2007*, ser. LNCS, vol. 4641. Springer, 2007, pp. 737–747.
[15] R. Karrenberg, "Automatic SIMD vectorization of SSA-based control flow graphs," Dissertation, Universität des Saarlandes, Saarbrücken, Jul. 2014.
[16] S. Ghosh, M. Martonosi, and S. Malik, "Cache miss equations: A compiler framework for analyzing and tuning memory behavior," *ACM Trans. Prog. Lang. Syst. (TOPLAS)*, vol. 21, no. 4, pp. 703–746, 1999.
[17] D. Feld, T. Soddemann, M. Jünger, and S. Mallach, "Facilitate SIMD-Code-Generation in the Polyhedral Model by Hardware-aware Automatic Code-Transformation," in *Proc. of the 3rd International Workshop on Polyhedral Compilation Techniques*, A. Größlinger and L.-N. Pouchet, Eds., Berlin, Germany, Jan. 2013, pp. 45–54.
[18] S. Williams, A. Waterman, and D. Patterson, "Roofline: An Insightful Visual Performance Model for Multicore Architectures," *Commun. ACM*, vol. 52, no. 4, pp. 65–76, Apr. 2009.
[19] G. Hager, J. Treibig, J. Habich, and G. Wellein, "Exploring performance and power properties of modern multi-core chips via simple machine models," *Concurrency and Computation: Practice and Experience*, vol. 28, pp. 189–210, 2016.
[20] J. Ansel, "Autotuning programs with algorithmic choice," Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA, Feb. 2014.
[21] N. Kalinnik, M. Korch, and T. Rauber, "Online auto-tuning for the time-step-based parallel solution of ODEs on shared-memory systems," *Journal of Parallel and Distributed Computing*, vol. 74, no. 8, pp. 2722–2744, 2014. doi: 10.1016/j.jpdc.2014.03.006
[22] B. A. Schmitt and R. Weiner, "Parallel start for explicit parallel two-step peer methods," *Numerical Algorithms*, vol. 53, no. 2-3, pp. 363–381, 2010.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 695–704
DOI: 10.15439/2016F35
ACSIS, Vol. 8. ISSN 2300-5963

Block Iterators for Sparse Matrices

Daniel Langr*†, Ivan Šimeček* and Tomáš Dytrych‡
* Czech Technical University in Prague
Faculty of Information Technology
Department of Computer Systems
Thákurova 9, 160 00, Praha, Czech Republic
Email: {langrd,xsimecek}@fit.cvut.cz
† Výzkumný a zkušební letecký ústav, a.s.
Beranových 130, 199 05, Praha, Czech Republic
‡ Czech Academy of Sciences
Nuclear Physics Institute
Řež 130, 250 68, Řež, Czech Republic
Email: tdytrych@ujf.cas.cz

***Abstract*—Finding an optimal block size for a given sparse matrix forms an important problem for storage formats that partition matrices into uniformly-sized blocks. Finding a solution to this problem can take a significant amount of time, which, effectively, may negate the benefits that such a format brings into sparse-matrix computations. A key for an efficient solution is the ability to quickly iterate, for a particular block size, over matrix nonzero blocks. This work proposes an efficient parallel algorithm for this task and evaluate it experimentally on modern multi-core and many-core high performance computing (HPC) architectures.**

I. Introduction

STORAGE formats prescribe a way how sparse matrices are stored in a computer memory. Many designed formats are based on partitioning of matrices into blocks where

1) blocks have a uniform size,
2) this size is not fixed for a given format and may be chosen for each matrix individually.

We call such formats *uniformly-blocking formats* or, shortly, *UB formats*.

Considering a particular sparse matrix A and a particular UB format, we thus face a problem of finding an optimal block size (whatever this means). Typically, we want to find a block size that will provide highest performance of sparse matrix-vector multiplication (SpMV) performed with A. Generally, this task cannot be accomplished until the matrix is fully assembled or at least until its structure of nonzero elements is fully known, which implies that matrices cannot be assembled in UB formats directly (there are usually other reasons as well). Instead, one needs to

1) assemble A in some suitable (not-parametrized) simple storage format,
2) find an optimal block size for A,
3) transform A in memory from its original storage format to a given UB format.

The second and third steps form an important problem related to a given UB format. If the solution of this problem takes too long, it might effectively negate the benefits that a UB format brings into subsequent computations with A.

To find an optimal block size, there is usually no option other than

1) to form some set of possibly-optimal block sizes,
2) to evaluate an optimization criterion for all of them.

Note that this approach generally gives a pseudooptimal block size instead of an optimal one. For sake of simplicity, we do not distinguish between these two cases and call them both optimal throughout this text.

Evaluation of an optimization criterion for a given UB format, given matrix A, and a particular tested block size typically involves gathering some information about all nonzero blocks of A. We therefore need to examine all these nonzero blocks. Such a procedure can be described briefly as follows: *for all nonzero blocks of A, perform some calculations that contribute to the evaluation of an optimization criterion*. Thus, in fact, we need to *iterate over nonzero blocks of A*.

When an optimal block size is found, this iterative process has to be run once again within the third step mentioned above, i.e., during the final transformation of A to a given UB format.

This paper addresses the problem of fast iteration over nonzero blocks of a sparse matrix. We propose an efficient scalable parallel algorithm for a solution to this problem and evaluate it experimentally on modern multi-core and many-core HPC architectures, where matrices frequently emerge in multi-threaded programs. (In distributed-memory environments, objects of our concern are "local" matrices formed by nonzero elements mapped to particular application processes.)

II. Case Study

As an illustrative case study, we work throughout this text with the *adaptive-blocking hierarchical storage format*

This work was supported by the Czech Science Foundation under Grant No. 16-16772S. This work was supported by the IT4Innovations Centre of Excellence project (CZ.1.05/1.1.00/02.0070), funded by the European Regional Development Fund and the national budget of the Czech Republic via the Research and Development for Innovations Operational Programme, as well as Czech Ministry of Education, Youth and Sports via the project Large Research, Development and Innovations Infrastructures (LM2011033).

Algorithm 1: Transformation of A to the ABHSF

Input: A: sparse matrix
Input: $\mathcal{S} = \{s_1, s_2, \ldots\}$: set of possibly-optimal block sizes
Data: M_{opt}, M, s_{opt}, i: auxiliary variables

```
M_opt ← 0
for i ← 1 to |S| do
    M ← 0
    for each nonzero block B of size s_i of A do
        find a space-optimal way W to store B in memory
        considering ⌈log2 s_i⌉ bits for in-block indexes;
        calculate the contribution c(B, W) of B stored in W
        to the memory footprint of A;
        M ← M + c(B, W)
    end
    if i = 1 or M < M_opt then
        s_opt ← s_i
        M_opt ← M
    end
end
for each nonzero block B of size s_opt of A do
    store B in memory in the ABHSF format
end
```

(ABHSF) [1], [2]. This format partitions A into uniform square blocks of size s and stores each block in memory in a space-optimal way. The optimization criterion of ABHSF is represented by the total memory footprint of A which is being minimized. This is a very common optimization criterion for storage formats in general (not only UB formats), since the performance of SpMV is limited by bandwidths of memory subsystems on modern HPC architectures [3].

Many UB formats work with in-block row and column indexes. Optimal (space-optimal) block sizes are then typically those that employ most or all of the available indexing bits. In case of the ABHSF and byte-padded in-block indexes, setting $s = 256$ is almost generally optimal or at least close to being optimal [1]. (Such a choice eliminates the discussed optimization problem, however, it does not eliminate the need to iterate over nonzero blocks of A; this process is still required for transformation of A into the ABHSF.)

On the other hand, if we really want to minimize the memory footprint of A stored in the ABHSF, we need to use the minimum possible number of bits for in-block indexes, i.e., $\lceil \log_2 s \rceil$. In such cases, any block size s can be generally optimal. Let $\mathcal{S} = \{s_1, s_2, \ldots\}$ denotes some set of possibly-optimal block sizes. The transformation of A into the ABHSF can then be written as Algorithm 1.

Algorithm 1 iterates over nonzero blocks of A exactly $|\mathcal{S}|+1$ times. Therefore, we want $|\mathcal{S}|$ to be

1) large enough to find the best possible block size,
2) small enough to prevent long algorithm running times.

One way to get close to both these outcomes is to consider only block sizes

$$\mathcal{S} = \{2^k : 1 \leq k \leq k_{\max}\}, \tag{1}$$

which implies maximum utilization of all k bits for in-block indexes. The $k_{\max}$ parameter, which corresponds to $|\mathcal{S}|$, determines the upper bound for tested block sizes. In practice, setting $k_{\max} = 10$ is typically sufficient, which implies 11 iterations over nonzero blocks of A while testing block sizes $s = 2, 4, 8, \ldots, 1024$ within Algorithm 1.

III. Notation

Let A be an $m \times n$ sparse matrix, where $a_{i,j}$ denotes the value of an element of A located in its ith row and jth column. As a mathematical object, A can be written as

$$A = \begin{pmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{m,1} & \cdots & a_{m,n} \end{pmatrix}. \tag{2}$$

However, within computer programs, we typically work only with nonzero elements of sparse matrices (or, even only with nonzero elements from a single triangular part if a matrix exhibits some kind of symmetry). An element of A is determined by its value, row index, and column index; let us write it as a triplet $(i, j, a_{i,j})$. As a data structure, we can consider A as a *set of matrix nonzero elements*:

$$A = \big\{(i, j, a_{i,j}) : 1 \leq i \leq m,\ 1 \leq j \leq n,\ a_{i,j} \neq 0\big\}. \tag{3}$$

Moreover, nonzero elements stored in memory are accessible in some order, which is typically prescribed by a given storage format. If this order matters, we can consider A as a *sequence of matrix nonzero elements*:

$$A = \big((i_l, j_l, a_{i_l,j_l})\big)_{l=1}^{nnz}, \quad a_{i_l,j_l} \neq 0, \tag{4}$$

where nnz denotes the number of nonzero elements of A.

In the text below, we use forms (2), (3), and (4) interchangeably, while preferring the particular one in dependence on the actual context. For the sake of simplicity, we also consider partitioning into square blocks only; the generalization for rectangular blocks is straightforward.

Partitioning A into square blocks of size s yields an $M \times N$ block matrix, where $M = \lceil m/s \rceil$ and $N = \lceil n/s \rceil$. For indexing block rows and columns, we use capital letters I and J, respectively. A block is called nonzero if it contains at least one nonzero matrix element.

A matrix element $(i, j, a_{i,j})$ belongs to a block with indexes

$$I = \lfloor (i-1)/s \rfloor + 1 \quad \text{and} \quad J = \lfloor (j-1)/s \rfloor + 1. \tag{5}$$

By using $\backslash$ for integer division, we can rewrite (5) as

$$I = (i-1) \backslash s + 1 \quad \text{and} \quad J = (j-1) \backslash s + 1. \tag{6}$$

Element's in-block indexes can be found correspondingly as $\big[(i-1) \mod s\big] + 1$ and $\big[(j-1) \mod s\big] + 1$.

Note that the calculations of block indexes and local in-block indexes for nonzero matrix elements involve integer division and modulo, which are relatively expensive arithmetic operations [4]. When possibly-optimal block sizes are chosen according to (1), both integer division and modulo can be substituted by much faster *logical shift* and *bitwise AND* operations.

Algorithm 2: Iteration over nonzero blocks of A: variant 1

```
Input: A: sparse matrix
Input: s: block size
Data: B: nonzero elements of a single block
Data: I, J: indexes
1  for I ← 1 to ⌈m/s⌉ do
2      for J ← 1 to ⌈n/s⌉ do
3          B ← {}
4          for all (i, j, a_{i,j}) ∈ A do
5              if (i − 1)\s + 1 = I and (j − 1)\s + 1 = J then
6                  B ← B ∪ {(i, j, a_{i,j})}
7              end
8          end
9          if B ≠ {} then
10             process block B with indexes I and J
11         end
12     end
13 end
```

Algorithm 3: Iteration over nonzero blocks of A: variant 2

```
Input: A: sparse matrix
Input: s: block size
Data: B_{I,J}: nonzero elements of a block in Ith block row and
      Jth block column
Data: I, J: indexes
1  for I ← 1 to ⌈m/s⌉ do
2      for J ← 1 to ⌈n/s⌉ do B_{I,J} ← {}
3  end
4  for all (i, j, a_{i,j}) ∈ A do
5      I ← (i − 1)\s + 1
6      J ← (j − 1)\s + 1
7      B_{I,J} ← B_{I,J} ∪ {(i, j, a_{i,j})}
8  end
9  for I ← 1 to ⌈m/s⌉ do
10     for J ← 1 to ⌈n/s⌉ do
11         if B_{I,J} ≠ {} then
12             process block B_{I,J} with indexes I and J
13         end
14     end
15 end
```

IV. Algorithms

Let us now analyse the problem of iteration over the nonzero blocks of A. In the most generic case, we have, at the outset, no knowledge which blocks of A are nonzero and which nonzero elements of A belong to these blocks. There are basically two ways to find this out:

1) to iterate over all blocks of A and for each block find its nonzero elements;
2) to iterate over all nonzero elements of A, find for each of them the corresponding block (6), and save the information that the element belongs to this block.

Pseudocodes for these two options are provided as Algorithms 2 and 3, respectively. Processing of blocks is application-dependent; it might, e.g., represent the calculation of blocks contributions to the optimization criterion (line 5–7 of Algorithm 1) or the storage of blocks in memory (line 15).

Algorithm 2 have low memory requirements; its auxiliary space is

$$S_2(A, s) = O(s^2),$$

since, at a given time, only nonzero elements for a single block need to be kept in memory. The drawback of this algorithm is its high time complexity

$$T_2(A, s) = \Theta(m \cdot n \cdot nnz/s^2).$$

As for Algorithm 3, its time complexity is considerably lower, namely

$$T_3(A, s) = \Theta(m \cdot n/s^2 + nnz).$$

However, the auxiliary space of Algorithm 3 is

$$S_3(A, s) = O(m \cdot n/s^2 + nnz),$$

since one needs to save the information about all nonzero elements for each nonzero block. Moreover, an implementation of this algorithm would likely require some complex dynamic data structure, which might introduce problems with memory fragmentation and expensive insertion/look-up operations.

Whenever working with sparse matrices, we generally want to avoid algorithms with $\Omega(nnz)$ auxiliary space as much as possible. Within many running instances of HPC programs, matrices are the largest objects in a computer memory and their sizes determine an extent of underlying computational problems. Any $\Omega(nnz)$ auxiliary space algorithm (such as Algorithm 3) thus, in effect, considerably limits the size of a problem being solved.

To avoid the high time complexity of Algorithm 2 as well as the high auxiliary space of Algorithm 3, we propose another solution for iteration over nonzero block of A that works as follows:

1) The nonzero elements of A are reordered such that the nonzero elements of each block are laid out consecutively (grouped together) in memory. In other words, *the nonzero elements are sorted with respect to blocks*.
2) A single iteration over nonzero elements is performed while elements of each nonzero block are identified and processed.

The pseudocode of such a solution is provided as Algorithm 4. Its time complexity and auxiliary space is dominated by the sorting step (line 1). Let us assume that we use an in-place randomized quicksort with time complexity $O\big(nnz \cdot log_2(nnz)\big)$ and auxiliary space $O\big(\log_2(nnz)\big)$. The overall time complexity of Algorithm 4 then will be

$$T_4(A, s) = O\big(nnz \cdot \log_2(nnz)\big)$$

and its auxiliary space

$$S_4(A, s) = O\big(\log_2(nnz)\big)$$

as well.

Algorithm 4 reduces both the time complexity of Algorithm 2 and the auxiliary space of Algorithm 3, however, at the following price: it requires A to be provided in such a format that facilitates reordering/sorting its nonzero elements. There is

Algorithm 4: Iteration over nonzero blocks of A: variant 3

Input: A: sparse matrix
Input: s: block size
Data: I, I', J, J', l, l_1: indexes

1 $\left((i_l, j_l, a_{i_l,j_l})\right)_{l=1}^{nnz} \leftarrow$ sort A with respect to blocks
2 $l_1 \leftarrow 1$
3 $I \leftarrow (i_1 - 1) \backslash s + 1$
4 $J \leftarrow (j_1 - 1) \backslash s + 1$
5 **for** $l \leftarrow 2$ to nnz **do**
6 $\quad I' \leftarrow (i_l - 1) \backslash s + 1$
7 $\quad J' \leftarrow (j_l - 1) \backslash s + 1$
8 $\quad$ **if** $I' \neq I$ **or** $J' \neq J$ **then**
9 $\quad\quad$ process block with indexes I and J that contains nonzero elements $\left((i_q, j_q, a_{i_q,j_q})\right)_{q=l_1}^{l-1}$
10 $\quad\quad l_1 \leftarrow l$
11 $\quad\quad I \leftarrow I'$
12 $\quad\quad J \leftarrow J'$
13 $\quad$ **end**
14 **end**
15 process block with indexes I and J that contains nonzero elements $\left((i_q, j_q, a_{i_q,j_q})\right)_{q=l_1}^{nnz}$

practically only one candidate—the *coordinate* storage format (COO) [5], [6]; it consists of three arrays containing row indexes, column indexes, and values of nonzero elements. At the same time, it does not prescribe any particular ordering for these arrays.

To require A to be initially in COO is not as restrictive in practice as it might seem, since:

1) any sparse matrix can be easily and quickly transformed into COO regardless of its original storage format,
2) COO is the most convenient format for assembling sparse matrices (newly generated nonzero elements are simply appended to the corresponding COO arrays).

A scenario where matrices are first assembled in COO and then transformed to another, computationally more suitable, storage format (such as some UB format) is thus perfectly viable for HPC programs.

To sort the nonzero elements with respect to blocks, we can define sorting keys by using the pairs of I and J block indexes calculated by (5). For example, if we want blocks to be sorted lexicographically, we can calculate sorting keys as $I \cdot N + J$. Again, note that choosing (1) for possibly-optimal block sizes implies faster calculation of sorting keys and therefore, in effect, likely faster sorting step within Algorithm 4.

A. Parallelization

Parallelization of (expensive) Algorithms 2 and 3 is straightforward. In Algorithm 2, we can parallelize the inner-most loop (line 4) while synchronizing concurrent updates of B at line 6. In Algorithm 3, we can parallelize the loops over blocks (lines 1–2 and 9–10) as well as the loop over nonzero matrix elements (line 4) while using thread-local I and J indexes and synchronizing concurrent updates to $B_{I,J}$ at line 7.

Parallelization of Algorithm 4 is a bit more complex; we propose its multi-threaded variant as Algorithm 5. Note that

Algorithm 5: Parallel iteration over nonzero blocks of A

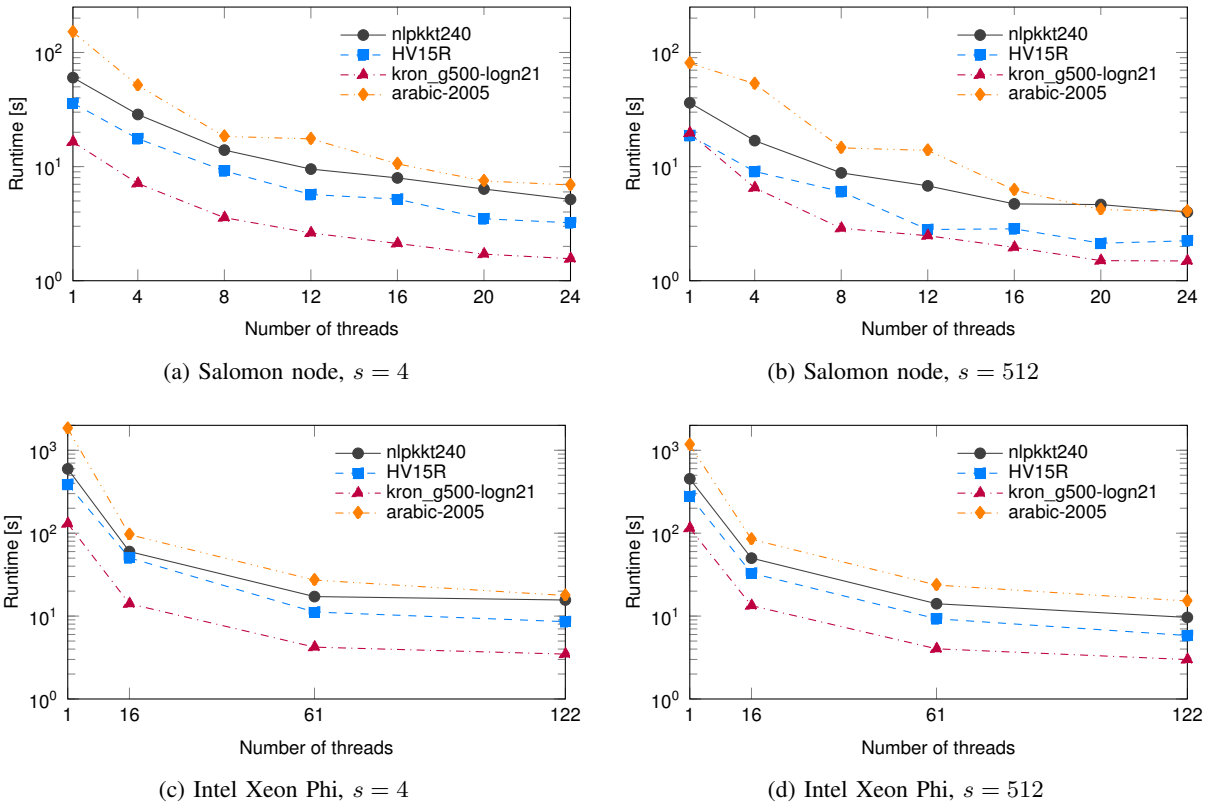
Input: A: sparse matrix
Input: s: block size
Input: T: number of threads
Data: I, I', J, J', l, l_1, t: thread-private indexes
Data: $tb[]$: thread-shared integer array of size $T + 1$

1 $\left((i_l, j_l, a_{i_l,j_l})\right)_{l=1}^{nnz} \leftarrow$ sort A in parallel with respect to blocks
2 $tb[1] \leftarrow 1$
3 $tb[T + 1] \leftarrow nnz + 1$
4 **for all** threads **do in parallel**
5 $\quad t \leftarrow$ current thread number (between 1 and T)
6 $\quad$ **if** $t > 1$ **then**
7 $\quad\quad l \leftarrow \lceil nnz \cdot (t - 1) \rceil \backslash T + 1$
8 $\quad\quad I \leftarrow (i_1 - 1) \backslash s + 1$
9 $\quad\quad J \leftarrow (j_1 - 1) \backslash s + 1$
10 $\quad\quad l \leftarrow l + 1$
11 $\quad\quad$ **while** $l \leq nnz$ **do**
12 $\quad\quad\quad I' \leftarrow (i_1 - 1) \backslash s + 1$
13 $\quad\quad\quad J' \leftarrow (j_1 - 1) \backslash s + 1$
14 $\quad\quad\quad$ **if** $I' \neq I$ **or** $J' \neq J$ **then break**
15 $\quad\quad\quad l \leftarrow l + 1$
16 $\quad\quad$ **end**
17 $\quad\quad tb[t] \leftarrow l$
18 $\quad$ **end**
19 $\quad$ perform barrier to synchronize threads
20 $\quad l_1 \leftarrow tb[t]$
21 $\quad I \leftarrow (i_{l_1} - 1) \backslash s + 1$
22 $\quad J \leftarrow (j_{l_1} - 1) \backslash s + 1$
23 $\quad$ **for** $l \leftarrow tb[t] + 1$ to $tb[t + 1] - 1$ **do**
24 $\quad\quad I' \leftarrow (i_l - 1) \backslash s + 1$
25 $\quad\quad J' \leftarrow (j_l - 1) \backslash s + 1$
26 $\quad\quad$ **if** $I' \neq I$ **or** $J' \neq J$ **then**
27 $\quad\quad\quad$ process block with indexes I and J that contains nonzero elements $\left((i_q, j_q, a_{i_q,j_q})\right)_{q=l_1}^{l-1}$
28 $\quad\quad\quad l_1 \leftarrow l$
29 $\quad\quad\quad I \leftarrow I'$
30 $\quad\quad\quad J \leftarrow J'$
31 $\quad\quad$ **end**
32 $\quad$ **end**
33 $\quad$ process block with indexes I and J that contains nonzero elements $\left((i_q, j_q, a_{i_q,j_q})\right)_{q=l_1}^{tb[t+1]-1}$
34 **end**

we cannot simply parallelize the main loop of Algorithm 4 (line 5), since its uniform splitting would generally cause threads to start with nonzero elements that are not, in sequence (4), first within corresponding blocks. Algorithm 5 therefore splits the load among threads such that:

1) an amortized number of nonzero elements processed by each thread is nnz/T, where T denotes the number of threads;
2) all nonzero elements of each particular block are processed by a single thread only.

Such splitting is calculated at lines 2–18 of Algorithm 5 and stored into an auxiliary array $tb[]$. Each thread can then process its exclusive portion of nonzero elements independently of other threads (lines 20–33). Threads are required to be indexed from 1 to T; if not, some mapping from thread IDs to such indexing must be provided.



(a) Salomon node, $s = 4$

(b) Salomon node, $s = 512$

(c) Intel Xeon Phi, $s = 4$

(d) Intel Xeon Phi, $s = 512$

Fig. 1: Strong scalability of Algorithm 5 with the *do-nothing processor* measured for different architectures, different matrices, and different block sizes.

V. Experiments

We have conducted an extensive experimental study to evaluate Algorithm 5. Within this study, we worked with matrices from the University of Florida Sparse Matrix Collection (UFSMC) [7]. Matrices that we used are listed in Appendix; their characteristics can be found at the UFSMC web pages[1]. We tried to choose matrices emerging in a wide range of scientific and engineering disciplines and thus having different properties, such as:

- different types of elements—real, complex, integer, binary;
- different sizes and shapes—square, rectangular;
- different kinds of symmetries—unsymmetric, symmetric, Hermitian;
- different numbers of nonzero elements—from $1.1 \cdot 10^7$ of the kim2 matrix to $6.4 \cdot 10^8$ of the arabic-2005 matrix;
- different densities, i.e., relative counts $nnz/(m \cdot n)$, of nonzero elements—from $5.12 \cdot 10^{-7}$ of the nlpkkt240 matrix to $1.11 \cdot 10^{-2}$ of the TSOPF_RS_b2328 matrix;
- different patterns of nonzero elements.

The matrices were read on the input from files downloaded from the UFSMC. All these files stored nonzero elements of matrices in the *reverse lexicograhical order* (RLO) and in the same order, we stored the elements in memory in the COO format as the first step of our benchmark program.

The measurements were performed on the following two shared-memory HPC architectures:

1) nodes of the Salomon supercomputer operated by IT4Innovations National Supercomputing Center in Ostrava, Czech Republic, having two 12-core Intel Xeon E5-2680v3 CPUs and 128 GB RAM per node;
2) Intel Xeon Phi coprocessor type 7120P with 16 GB RAM.

Benchmark codes were written in C++ and we used the GNU g++ compiler version 5.1.0 on Salomon and Intel icpc compiler version 16.0.1 for Intel Xeon Phi builds.

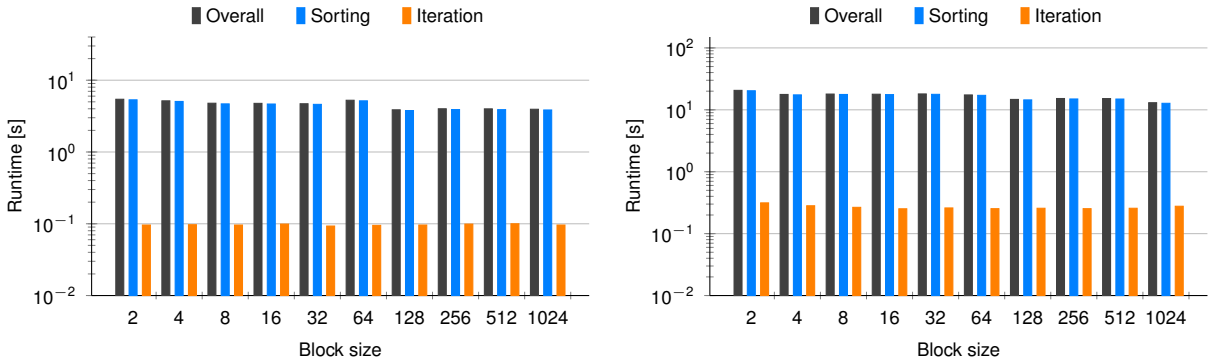
Parallelization was implemented with OpenMP. As for sorting (line 1 of Algorithm 5), we used AQsort[2]—an OpenMP-based multi-threaded variant of in-place quicksort that can work with multiple arrays, such as the arrays of the COO storage format in our case.

A. Processors

Lines 27 and 33 of Algorithm 5 contain processing of found nonzero blocks. Within this study, we invoked two different block *processors* at these points. The first one did nothing useful at all, which allowed us to evaluate the algorithm itself

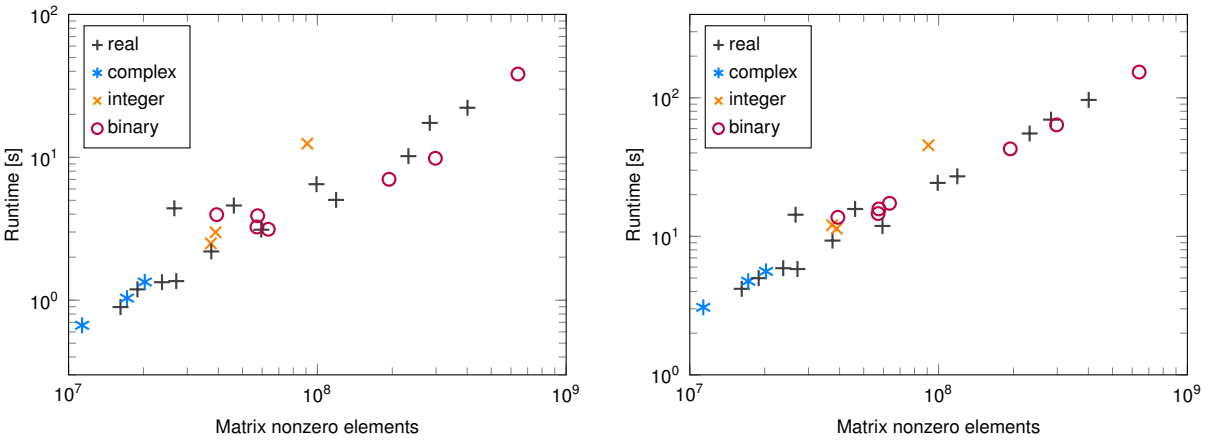
[1] http://www.cise.ufl.edu/research/sparse/matrices/

[2] https://github.com/DanielLangr/AQsort



(a) Salomon node, nlpkkt240, $T = 24$ (b) Intel Xeon Phi, arabic-2005, $T = 122$

Fig. 2: Runtime of Algorithm 5 (*Overall*) and its two phases (*Sorting* and *Iteration*) with the *do-nothing processor*, measured for different architectures, different matrices, different block sizes, and the optimal number of threads.



(a) Salomon node, $T = 24$ (b) Intel Xeon Phi, $T = 122$

Fig. 3: Aggregated runtime of Algorithm 5 with the *do-nothing processor* run 10 times in a row for block sizes $s = 2, 4, 8, \ldots, 1024$, measured for 26 matrices from the UFSMC, different architectures and the optimal number of threads.

without any application-dependent computations; we call this processor a *do-nothing processor*.[3]

The second processor was designed for the problem of finding an optimal block size when storing A in the ABHSF; we call it the *ABHSF-opt processor*. This processor calculated and summed the contributions of blocks to the overall memory footprint of A.

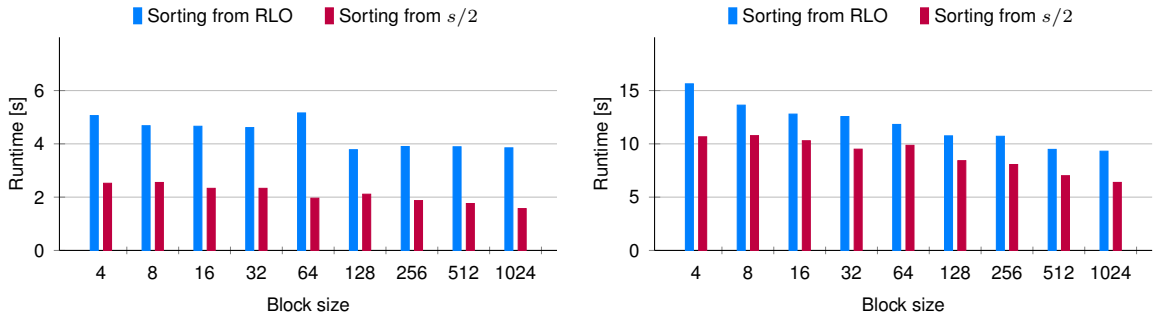
B. Scalability

First, we measured the strong scalability of Algorithm 5; the results for 4 different matrices and 2 different block sizes are shown in Fig. 1. In all cases, parallelization led to a considerable reduction of runtime required for the iteration over nonzero blocks of A. This runtime was dominated by the sorting phase of the algorithm (see Section V-C for details), thus, consequently, the overall scalability of Algorithm 5 was determined by the scalability of AQsort within our study. The maximum number of threads, i.e., 24 for Salomon nodes and 122 for Intel Xeon Phi, was chosen experimentally; beyond these points, runtime of AQsort started to grow significantly.

[3]A processor that would do anything at all might be optimized away by the compiler. We therefore designed the do-nothing processor such that it summed the number of nonzero elements of blocks, which also allowed us to verify that the algorithm correctly iterated over all nonzero blocks of A.

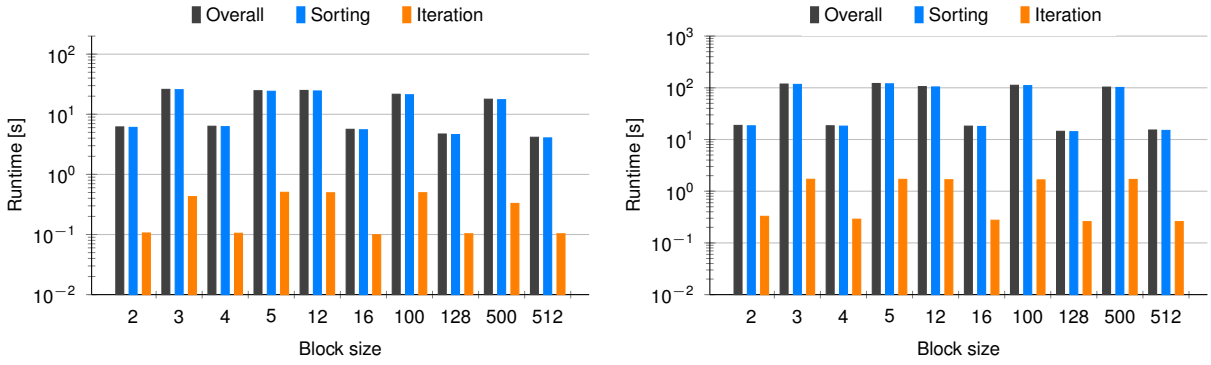
C. Algorithm Phases

The second experiment evaluated the contributions of the *sorting* and *iterations* phases of Algorithm 5 to its *overall* runtime; the results are presented by Fig. 2. The set of tested block sizes was selected according to (1) while setting $k_{\max} = 10$. The sorting phase of Algorithm 5 clearly dominates the overall algorithm runtime. The runtime of the iteration phase (with the *do-nothing* processor in this case) is practically negligible. We can also notice that larger block sizes yielded slightly faster sorting due to lower number of distinct sorting keys (less nonzero elements need to be swapped in memory).



(a) Salomon node, nlpkkt240, $T = 24$ (b) Intel Xeon Phi, nlpkkt240, $T = 122$

Fig. 4: Runtime of the sorting phase of Algorithm 5 for different block sizes when nonzero elements were sorted from the RLO (*Sorting from RLO*) and from the ordering given by half a block size (*Sorting from* $s/2$).



(a) Salomon node, nlpkkt240, $T = 24$ (b) Intel Xeon Phi, arabic-2005, $T = 122$

Fig. 5: Runtime of Algorithm 5 (*Overall*) and its two phases (*Sorting* and *Iteration*) with the *do-nothing processor*, measured for different architectures, different matrices, different block sizes, and the optimal number of threads.

D. Multiple Block Sizes

Within the problem of finding an optimal block size, we need to iterate over nonzero blocks of matrices multiple times while testing different block sizes. In the next experiment, we therefore run Algorithm 5 ten times in a row, while testing block sizes $s = 2, 4, 8, \ldots, 1024$ as proposed in Section II. We used 26 matrices from the UFSMC (listed in Appendix) and, for each one, measured the aggregated runtime of all 10 algorithm runs. The results are presented by Fig. 3, which shows runtimes as a function of the number of matrix nonzero elements.

We can observe that the relation of runtime and nnz was roughly linear. Though, for a constant T, the time complexity of Algorithm 5 is $O\big(n \cdot \log_2(n)\big)$, modern implementations of parallel quicksorts yield in practice such a linear growth of runtime on modern multi-core and many-core architectures (details are beyond the scope of this text).

E. Initial Ordering Effects

Fig. 2 shows the runtimes for sorting of nonzero elements of matrices from the RLO to the block-aware ordering. However, when we are looking for an optimal block size from $\mathcal{S}$ for a given matrix, we initially need to sort its nonzero elements from the input ordering (the RLO in our case) only once for the tested block size s_1. Then, for all other tested block sizes $s_k : k > 1$, the sorting algorithm takes as an input nonzero elements sorted with respect to the block size s_{k-1}.

Within our study, we considered block sizes $s_k = 2^k$, which implies $s_k = 2 \cdot s_{k-1}$ for $k > 1$. Moreover, we defined sorting keys according to the lexicographical ordering of blocks. Consequently, for $k > 1$, the nonzero elements were on the input of Algorithm 5 partially sorted, which should result in shorter sorting times. We performed an experiment to verify this assumption; the results are shown in Fig. 4. They clearly indicate that AQsort was able to take the advantage of such partially sorted data; the amount of spared time was significant, especially on Salomon CPU-based nodes.

F. Block Sizes Effects

Recall that in the previous text, we made an assumption that setting block sizes $s_k = 2^k$ should provide faster runs of Algorithm 5 due to the possibility of calculation of block indexes I and J by using cheap logical and bitwise operations. However, if we need to test block sizes other than the powers of 2, we cannot avoid integer division (6). To evaluate the

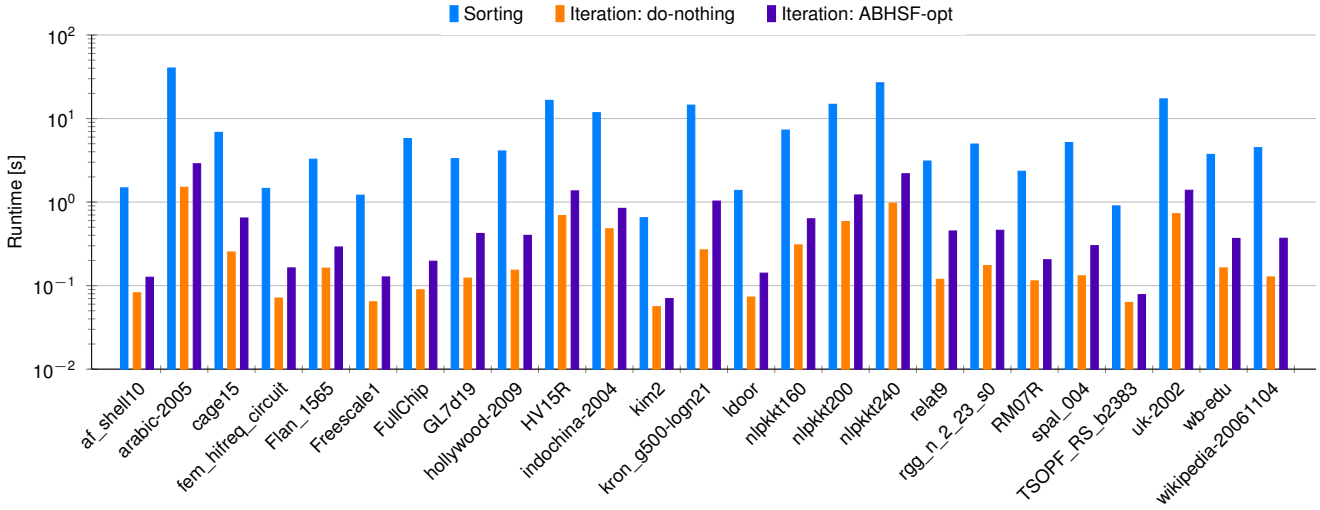


Fig. 6: Aggregated runtime of sorting and iteration phases of Algorithm 5 run 10 times in a row for block sizes $s = 2, 4, 8, \ldots, 1024$, measured for different matrices on a Salomon node using 24 threads. The iteration phases were measured for both the *do-nothing* and *ABHSF-opt* processors.

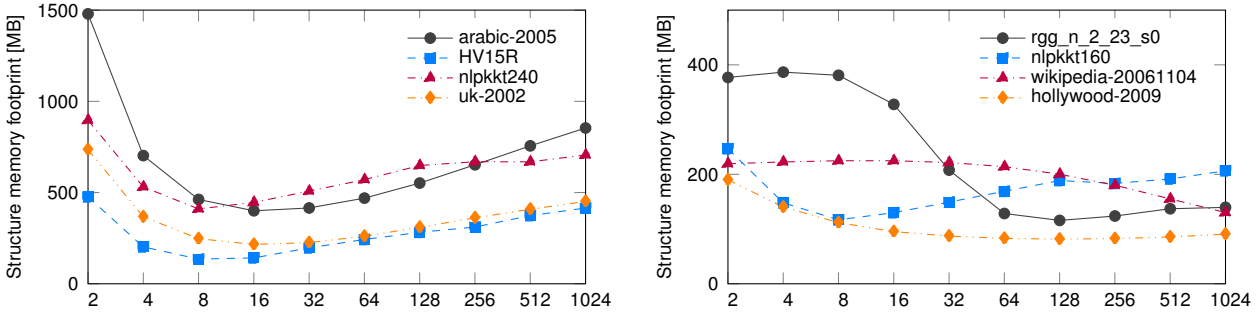


Fig. 7: Matrix structure memory footprints for different matrices stored in the ABHSF and different block sizes.

difference between both cases, we measured runtimes of Algorithm 5 and both its phases (sorting and iteration) for several block sizes of both classes $s = 2^k$ and $s \neq 2^k$; the results are presented in Fig. 5.

The conclusion is obvious—the way of deriving block indexes for the nonzero elements had a tremendous impact on the algorithm. Its runtime grew by almost a factor of 4 and 7 on Salomon nodes and Intel Xeon Phi, respectively, when using integer division instead of bitwise/logical operations.

G. ABHSF

Up to now, we presented measurements that used the *do-nothing processor* designed to evaluate Algorithm 5 itself. However, in practice, we iterate over nonzero blocks of sparse matrices to do something useful and the question is how the algorithm runtime will change in such cases. To answer this question, we substituted the *do-nothing processor* with the *ABHSF-opt* processor introduced in Section V-A and used Algorithm 5 for finding optimal block sizes for all tested matrices. The results of this experiment are presented in Fig. 6. They show aggregated runtimes of all 10 sorting phases as well as 10 iteration phases, and, for comparison, we show results for both types of block processors. We can observe that with the *ABHSF-opt processor*, the iteration phase took considerably longer times in comparison with the *do-nothing* processor. However, the overall runtime of the whole algorithm was still dominated by its sorting phase.

In regard to memory footprints of sparse matrices, we can usually focus only on *matrix structure memory footprints*, i.e., memory footprints of the information describing the structure of nonzero elements (compression of the values of nonzero elements pays off only for special kinds of matrices where same values emerge many times). For illustration, we show in Fig. 7 the relation between block sizes and the matrix structure memory footprints of selected matrices stored in the ABHSF. For most of the tested matrices, we have found only a single minimum, which typically corresponded to block sizes of 8 or 16 (left side of Fig. 7 and the nlpkkt160 matrix). However, we have also observed few "pathological" cases with different behavior (right side of Fig. 7). For example, the minimum

TABLE I: Matrix structure memory footprints in MB for selected matrices stored in the COO and CSR storage formats with 32-bit indexes and the ABHSF with optimal block sizes.

Matrix	COO	CSR	ABHSF
arabic-2005	4882.8	2528.2	400.1
hollywood-2009	438.8	223.8	81.5
HV15R	2159.7	1087.5	135.4
nlpkkt160	907.4	485.5	116.9
nlpkkt240	3061.2	1637.4	410.3
rgg_n_2_23_s0	484.5	274.2	115.8
uk-2002	2274.4	1207.9	215.7
wikipedia-20061104	300.5	162.2	130.1

for the wikipedia-20061104 matrix was not even found within the whole tested range $s = 2, 4, 8, \ldots, 1024$; such a result might indicate that the ABHSF is not a suitable format for this matrix.

We also present in Table I the comparison of the matrix structure memory footprints for selected matrices and 3 storage formats—COO, the *compressed sparse row* (CSR) format, and the ABHSF. CSR is likely the most commonly used format for sparse matrices, together with its *compressed sparse column* (CSC) counterpart (they are also often abbreviated as CRS and CCS). The measurements revealed that storing sparse matrices in the ABHSF can result in substantial memory savings.

VI. Related Work

We have proposed an algorithm for the purpose of storing matrices in a file system in the ABHSF [2, Algorithm 1]. This algorithm served as a starting point for the development of Algorithm 4 that was generalized for any UB format.

For examples of designed UB formats, see, e.g., [1], [5], [8]–[21]

VII. Conclusions

The contribution of this paper is an efficient scalable parallel algorithm for fast iteration over nonzero blocks of sparse matrices. This algorithm is a building block of a process of transformation of sparse matrices into UB storage formats. We have presented an extensive experimental study with the proposed algorithm using matrices from the UFSMC that came from different scientific and engineering disciplines and thus featured different characteristics. Measurements conducted on modern multi-core and many-core HPC architectures revealed that if the set of tested block sizes is chosen properly, the process of finding an optimal block size takes up to tens of seconds even for very large matrices.

The remaining question is whether or not it pays off to transform matrices into UB formats. The answer to this question is highly application-dependent. For instance, if a matrix is used within an iterative linear solver or an eigensolver, we would first need to know how many SpMV operations are applied to a given matrix and how much time this operation takes. In our future work, we want to focus on the ABHSF and undertake a research that should tell how many SpMV-based iterations need to be done with a given matrix to reduce the overall application runtime when considering matrix storage in this format.

Appendix

The list of sparse matrices from the UFSMC used in the experiments: 3Dspectralwave, af_shell10, arabic-2005, cage15, fem_hifreq_circuit, Flan_1565, Freescale1, FullChip, GL7d19, hollywood-2009, HV15R, indochina-2004, kim2, kron_g500-logn21, ldoor, nlpkkt160, nlpkkt200, nlpkkt260, relat9, rgg_n_2_23_s0, RM07R, spal_004, TSOPF_RS_b2383, uk-2002, wb-edu, wikipedia-20061104.

Acknowledgments

The authors acknowledge support from P. Tvrdík from the Czech Technical University in Prague, P. Vrchota and J. Fiala Výzkumný a zkušební letecký ústav, a.s., and M. Pajr from CQK Holding and IHPCI. The authors would like to thank M. Václavík of the Czech Technical University in Prague for providing an access to an Intel Xeon Phi accelerator installed at the Star university cluster.

References

[1] D. Langr, I. Šimeček, P. Tvrdík, T. Dytrych, and J. P. Draayer, "Adaptive-blocking hierarchical storage format for sparse matrices," in *Proceedings of the Federated Conference on Computer Science and Information Systems (FedCSIS 2012)*. IEEE Xplore Digital Library, 2012, pp. 545–551.

[2] D. Langr, I. Šimeček, and P. Tvrdík, "Storing sparse matrices in the adaptive-blocking hierarchical storage format," in *Proceedings of the Federated Conference on Computer Science and Information Systems (FedCSIS 2013)*. IEEE Xplore Digital Library, 2013, pp. 479–486.

[3] D. Langr and P. Tvrdík, "Evaluation criteria for sparse matrix storage formats," *IEEE Transactions on Parallel and Distributed Systems*, vol. 27, no. 2, pp. 428–440, 2016. doi: 10.1109/TPDS.2015.2401575

[4] A. Fog, "Instruction tables: Lists of instruction latencies, throughputs and micro-operation breakdowns for Intel, AMD and VIA CPUs," 2016, accessed April 8, 2016 at http://www.agner.org/optimize/instruction_tables.pdf.

[5] R. Barrett, M. Berry, T. F. Chan, J. Demmel, J. Donato, J. Dongarra, V. Eijkhout, R. Pozo, C. Romine, and H. V. der Vorst, *Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods*, 2nd ed. Philadelphia, PA: SIAM, 1994.

[6] Y. Saad, *Iterative Methods for Sparse Linear Systems*, 2nd ed. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2003. ISBN 0898715342

[7] T. A. Davis and Y. F. Hu, "The University of Florida Sparse Matrix Collection," *ACM Transactions on Mathematical Software*, vol. 38, no. 1, pp. 1:1–1:25, 2011. doi: 10.1145/2049662.2049663

[8] M. Belgin, G. Back, and C. J. Ribbens, "Pattern-based sparse matrix representation for memory-efficient SMVM kernels," in *Proceedings of the 23rd International Conference on Supercomputing*, ser. ICS '09. New York, NY, USA: ACM, 2009. doi: 10.1145/1542275.1542294. ISBN 978-1-60558-498-0 pp. 100–109.

[9] ——, "A library for pattern-based sparse matrix vector multiply," *International Journal of Parallel Programming*, vol. 39, no. 1, pp. 62–87, 2011. doi: 10.1007/s10766-010-0145-2

[10] A. Buluç, J. T. Fineman, M. Frigo, J. R. Gilbert, and C. E. Leiserson, "Parallel sparse matrix-vector and matrix-transpose-vector multiplication using compressed sparse blocks," in *Proceedings of the 21st Annual Symposium on Parallelism in Algorithms and Architectures*, ser. SPAA '09. New York, NY, USA: ACM, 2009. doi: 10.1145/1583991.1584053. ISBN 978-1-60558-606-9 pp. 233–244.

[11] A. Buluc, S. Williams, L. Oliker, and J. Demmel, "Reduced-bandwidth multithreaded algorithms for sparse matrix-vector multiplication," in *Proceedings of the 2011 IEEE International Parallel & Distributed Processing Symposium*, ser. IPDPS '11. IEEE Computer Society, 2011. doi: 10.1109/IPDPS.2011.73 pp. 721–733.

[12] J. W. Choi, A. Singh, and R. W. Vuduc, "Model-driven autotuning of sparse matrix-vector multiply on GPUs," in *Proceedings of the 15th ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming*, ser. PPoPP '10. New York, NY, USA: ACM, 2010. doi: 10.1145/1693453.1693471 pp. 115–126.

[13] E.-J. Im and K. Yelick, "Optimizing sparse matrix computations for register reuse in SPARSITY," in *Proceedings of the International Conference on Computational Science (ICCS 2001), Part I*, ser. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2001, vol. 2073, pp. 127–136.

[14] E.-J. Im, K. Yelick, and R. Vuduc, "Sparsity: Optimization framework for sparse matrix kernels," *International Journal of High Performance Computing Applications*, vol. 18, no. 1, pp. 135–158, 2004. doi: 10.1177/1094342004041296

[15] V. Karakasis, G. Goumas, and N. Koziris, "A comparative study of blocking storage methods for sparse matrices on multicore architectures," in *Computational Science and Engineering, 2009. CSE '09. International Conference on*, vol. 1, Aug 2009. doi: 10.1109/CSE.2009.223 pp. 247–256.

[16] R. Nishtala, R. W. Vuduc, J. W. Demmel, and K. A. Yelick, "Performance modeling and analysis of cache blocking in sparse matrix vector multiply," University of California, Tech. Rep. UCB/CSD-04-1335, 2004.

[17] ——, "When cache blocking of sparse matrix vector multiply works and why," *Applicable Algebra in Engineering, Communication and Computing*, vol. 18, no. 3, pp. 297–311, 2007. doi: 10.1007/s00200-007-0038-9

[18] I. Šimeček, D. Langr, and P. Tvrdík, "Space-efficient sparse matrix storage formats for massively parallel systems," in *Proceedings of the 14th IEEE International Conference of High Performance Computing and Communications (HPCC 2012)*. IEEE Computer Society, 2012. doi: 10.1109/HPCC.2012.18 pp. 54–60.

[19] I. Šimeček and D. Langr, "Space and execution efficient formats for modern processor architectures," in *Proceedings of the 17th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC 2015)*. IEEE Computer Society, 2015. doi: 10.1109/SYNASC.2015.24 pp. 98–105.

[20] F. S. Smailbegovic, G. N. Gaydadjiev, and S. Vassiliadis, "Sparse Matrix Storage Format," in *Proceedings of the 16th Annual Workshop on Circuits, Systems and Signal Processing, ProRisc 2005*, 2005, pp. 445–448.

[21] P. Stathis, S. Vassiliadis, and S. Cotofana, "A hierarchical sparse matrix storage format for vector processors," in *Proceedings of the 17th International Symposium on Parallel and Distributed Processing*, ser. IPDPS '03. Washington, DC, USA: IEEE Computer Society, 2003, p. 61.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 705–708
DOI: 10.15439/2016F48
ACSIS, Vol. 8. ISSN 2300-5963

An Iteration Space Visualizer for Polyhedral Loop Transformations in Numerical Programming

Marek Palkowski, Wlodzimierz Bielecki
West Pomeranian University of Technology in Szczecin
ul. Zolnierska 49, 71-210 Szczecin, Poland
Email: mpalkowski@wi.zut.edu.pl, wbielecki@wi.zut.edu.pl

***Abstract*—An iteration space visualizer is presented to analyze parallelism in loop nests including parallelism in tiled code of numerical programs. The tool visualizes exact data dependences available in arbitrarily nested loops as well as tiles generated with TRACO by means of the transitive closure of a loop nest dependence graph. Various graphical operations such as rotation, zooming, coloring and filtering allow for a detailed examination of dependences, iteration space slices, and shapes of generated tiles. The visualizer is a built-in TRACO module which collects results generated with TRACO and it is launched automatically when TRACO finishes code generation. The visualizer helps high-performance application developers discover parallelism available in loop nests and analyze tiled code produced by means of the polyhedral model.**

I. Introduction

AUTOMATIC parallelization in numerical programs has been the topic of research for many decades. In the majority of cases, the techniques focus on two basic steps: dependence analysis and program transformations. Despite the great steps forward in this area, sophisticated dependences, the construction of loop transformations, and statement instance mappings are beyond what the programmer is able to see at first glance [1].

This paper focuses on a graphical support for the automatic parallelizer and optimizer, TRACO, – the source-to-source compiler based on the transitive closure of a dependence graph to transform affine loop nests. A proposed visualizer assists both experts and non-expert programmers to understand the results generated with TRACO [2] and code generated by it.

The TRACO visualizer is based on Python scripts which use wrappers to the Integer Set Library (ISL) [3] and matplotlib [4]. ISL allows users to manipulate on sets and maps while matplotlib is a plotting library to produce 2-D figures and 3-D interactive projections. The tool collects and visualizes data generated by TRACO, for example, dependences, code fragments to be executed in parallel, and shapes of tiles.

The remainder of the paper is organized as follows. The next section discusses the basics of the polyhedral model, iteration space dependence graph and operations to manipulate maps and sets. Section 3 briefs algorithms implemented in TRACO. Section 4 explores visualizing functions and theirs capabilities. Section 5 introduces related work. The last section concludes the paper.

II. Background

In this paper, we deal with affine loop nests where, for given loop indices, lower and upper bounds as well as array subscripts and conditionals are affine functions of surrounding loop indices and possibly of structure parameters (defining loop index bounds), and the loop steps are known constants.

To implement algorithms, we have chosen the dependence analysis proposed by Pugh and Wonnacott [5], where dependences are represented by dependence relations. A dependence relation is a tuple relation of the form [*input list*]→[*output list*]: *formula*, where *input list* and *output list* are the lists of variables and/or expressions used to describe input and output tuples, and *formula* describes the constraints imposed upon input and output lists and it is a Presburger formula built of constraints represented by algebraic expressions and using logical and existential operators [5].

Standard operations on relations and sets are used, such as intersection (∩), union (∪), difference (-), domain (dom *R*), range (ran *R*), relation application (S'= *R*(*S*): $e' \in S'$iff exists *e* s.t. $e \rightarrow e' \in R$, $e \in S$). In detail, the description of these operations is presented in [5], [3].

The positive transitive closure for a given relation *R*, R^+, is defined as follows [6] $R^+ = \{e \rightarrow e' : \ e \rightarrow e' \in R \vee \exists e'' s.t. \ e \rightarrow e'' \in R \wedge e'' \rightarrow e' \in R^+\}$. It describes which vertices e' in a dependence graph (represented by relation *R*) are connected directly or transitively with vertex *e*. Transitive closure, R^*, describes the same connections in a dependence graph (represented by R) that R^+ does plus connections of each vertex with itself.

In sequential loop nests, the iteration i executes before j if i is *lexicographically less* than j, denoted as $i \prec j$, i.e., $i_1 < j_1 \vee \exists k \geq 1 : i_k < j_k \wedge i_t = j_t, for \ t < k$.

An ultimate dependence source is a source that is not the destination of another dependence. Set including all dependence sources,S_{UDS} is calculated as follows: $S_{UDS} = domain(R) - range(R)$, where a dependence relation *R* describes all the dependences in a loop nest.

An (iteration-space) slice is defined as follows. Given a dependence graph defined by a set of dependence relations, a slice *S* is a weakly connected component of this graph, i.e., a maximal sub-graph such that for each pair of vertices in the sub-graph there exists a forward or backward path.

Let IS denotes the loop nest iteration space. A function is called a schedule if it maps each iteration of IS onto another

978-83-60810-90-3/$25.00

space so that all data dependences available in the loop nest are preserved. The schedule that maps every $x \in IS$ to the first possible time step allowed by the dependences is called the free schedule.

III. The TRACO Compiler

TRACO allows us to generate parallel code. Parallelization is based on extracting synchronization-free slices or producing and applying the free-schedule. An approach to extract synchronization-free slices takes two steps [2]. First, for each slice, a representative statement instance is defined (it is the lexicographically minimal statement instance from all the ultimate sources of a slice). Next, slices are reconstructed from their representatives and code scanning these slices is generated.

In order to find representatives of slices, we build a relation, R_{USC} that describes all pairs of the ultimate dependence sources being transitively connected in a slice. The relation is constrained with the intersection of the sets $R*(e)$ and $R*(e') : (R*(e) \cap R*(e'))$ which guarantees that vertices e and e' are transitively connected, i.e., they are the sources of the same slice.

Next, set, S_{repr}, containing representatives of each slice is found as S_{repr} = S_{UDS} - range(R_{USC}). Then the remaining sources of this slice can be found by applying the relation (R_{USC})* to set S_{repr}. A set, representing slice elements, is formed by applying R* to the sources of a slice. To generate code, we apply the CLooG library [7] or ISL [3] to the set comprising statement instances of independent slices.

To parallelize loop nests which expose a single synchronization-free slice, time partitioning is applied. The algorithm, presented in paper [8], allows us to generate the free schedule and next fine-grained parallel code; all statement instances of a time partition can be executed in parallel, while partitions are enumerated sequentially.

Given relations R, representing all dependences in a loop nest, we calculate R^k, where $R^k = \underbrace{R \circ R \circ ... R}_{k}$, "$\circ$" is the composition operation. Given set *UDS* comprising all loop nest statement instances that are ready to execution at time *k*=0, each vertex, belonging the set S_k = $R^k(UDS)$ - $R^+ \circ R^k(UDS)$, is connected in the dependence graph, defined by relation *R*, with some vertex(ices) represented by set *UDS* with a path of length *k*. Hence at time *k*, all the statement instances belonging to the set S_k can be scheduled for execution and it is guaranteed that *k* is as few as possible.

Tiling is a very important iteration reordering transformation for both improving data locality and extracting loop nest parallelism. TRACO allows users generate parallel tiled code by means of algorithms based on the transitive closure of a dependence graph [2]. First, we form rectangular set *TILE*(***II***,***B***) including iterations belonging to a parametric tile as follows *TILE*(***II***, ***B***) = $\{[\boldsymbol{I}] \mid \boldsymbol{B}*\boldsymbol{II} + \boldsymbol{LB} \leq \boldsymbol{I} \leq \min(\boldsymbol{B}*(\boldsymbol{II}+\mathbf{1}) + \boldsymbol{LB} - \mathbf{1}, \boldsymbol{UB})$ AND $\boldsymbol{II} \geq 0\}$, where vectors ***LB*** and ***UB*** include the lower and upper loop index bounds of an original loop nest, respectively; diagonal matrix ***B*** defines the size of a rectangular original tile; elements of vectors ***I*** and ***II*** represent the original loop nest indices and the identifiers of tiles, respectively; **1** is the vector whose all elements are equal to 1.

TRACO, instead of program transformations represented by a set of affine functions, one for each statement, uses the transitive closure of a loop nest dependence graph to carry out corrections of original rectangular tiles so that all dependences of the original loop nest are preserved under the lexicographic order of target tiles. This may lead to changing shapes of original rectangular tiles; target tiles can be of an arbitrary shape which is affected with dependences available in a loop nest. Recognizing shapes from a mathematical representation of target tiles can be difficult. The visualizer helps recognize what are tile shapes. After code generation, the visualizer forms a graphical representation of the iteration space, dependences, and target tiles.

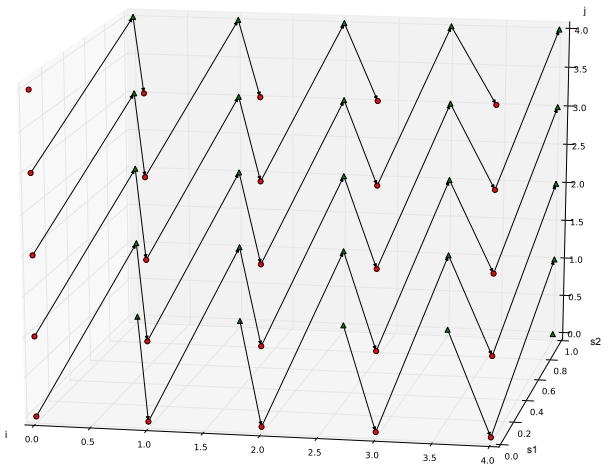


Fig. 1. Iteration space and dependences of Example 1 (n=4); statement instances are marked with the red circles and green triangles.

IV. Applying the TRACO Visualizer

Let us consider the example from paper [9] presented below:

```
// Example 1.
for(i=0; i<=n; i++)
 for(j=0; j<=n; j++){
  a[i][j] = a[i][j] + b[i-1][j];//s1
  b[i][j] = a[i][j-1 * b[i][j]; //s2
}
```

Figure 1 shows the iteration space and dependences for Example 1 generated with the visualizer. Analyzing this figure, we can discover that coarse-grained parallelism represented with synchronization-free slices is available in the considered loop nest (9 threads when n=4), but slices are load imbalanced. We also can see that there exist fine-grained parallelism, i.e., exist time partitions: for each value of index j, we can execute all instances of statement s2 in parallel for all values of index i, then all instances of statement s1. Visualization helps to

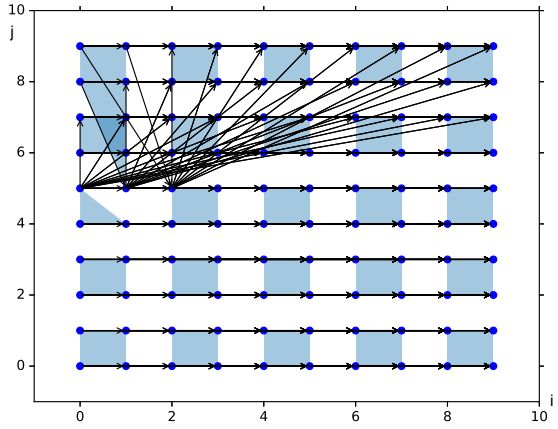


Fig. 2. Dependences, tiles of the size 2x2, and slices for Example 2.

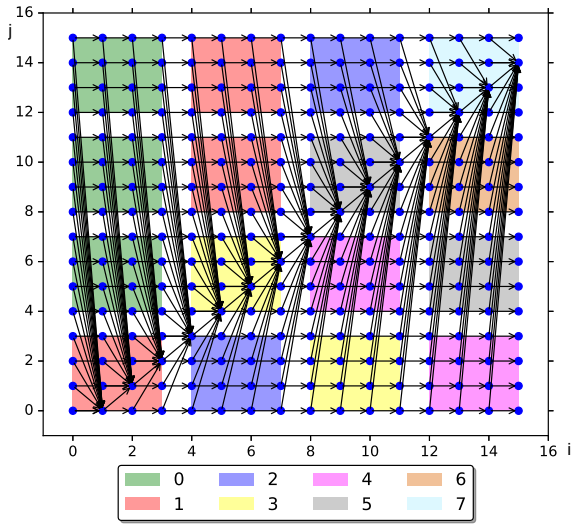


Fig. 3. Dependences, tiles of the size 4x4 and the schedule for Example 3.

discover parallelism available in Example 1 and what is the type of parallelism.

For Example 2 below

```
// Example 2.
for(i=0; i<=9; i++)
 for(j=0; j<=9; j++)
  a[j+4][j+1]=a[i+2*j+1][i+j+3];
```

Figure 2 depicts the iteration space with dependences. The loop tiling algorithm, implemented in TRACO, moves iteration [1,5] from tile [0,2] to tile [0,4] because it is the destination of the dependence whose source belongs to iteration [0,8]. For this example, TRACO is able to find three independent slices whose elements are tiles.

The iteration space with dependences for Example 3 is illustrated in Figure 3. Original rectangular tiling is preserved for this example. The figure shows tiles which are executed according to the free schedule. Time partitions are marked by different colors starting with the three independent green tiles.

```
// Example 3.
for(i=0; i<=15; i++)
 for(j=0; j<=15; j++)
    a[i][j] = a[i+1][i]+a[i+1][j];
```

Figure 4 illustrates dependences and tiles for the loop nest below in the 3D space.

```
// Example 4
for(k=0; k<=15; k++)
 for(i=0; i<=15; i++)
  for(j=0; j<=15; j++)
   a[i][j][k] =  a[i+1][j-1][k];
```

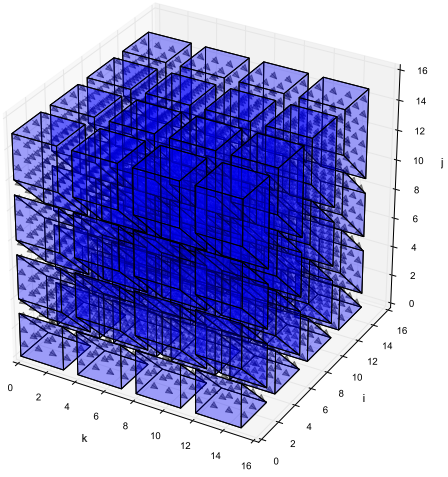


Fig. 4. Dependences and tiles of the size 2x2 for Example 4.

We use also a 3D projection to visualize 2D tiles for the arbitrary nested loops of Example 5, see Figure 5. The third axis indicates numbers of statements in the loop nest.

```
// Example 5
for(i=0; i<=15; i++){
 for(j=0; j<=15; j++)
  a[i][j] = a[i+1][j-1];
 for(j=0; j<=i; j++)
  b[i][j] = b[i][j+1]+a[i][0]; }
```

Figure 6 presents four independent slices whose elements are tiles for Example 6, the Planckian distribution. A 3D-projection reveals synchronization-free parallelism after appropriate rotating.

```
// Example 6 - Planckian distribution, loop=n=7
for ( l=1 ; l<=loop ; l++ )
  for ( k=0 ; k<n ; k++ ){
    y[k] = u[k] / v[k];
    w[k] = x[k] / ( exp( y[k] ) -1.0 );
  }
```

Summing up, we can conclude that visualization allows the programmer to find suitable loop transformations and discover available parallelism or code optimization.

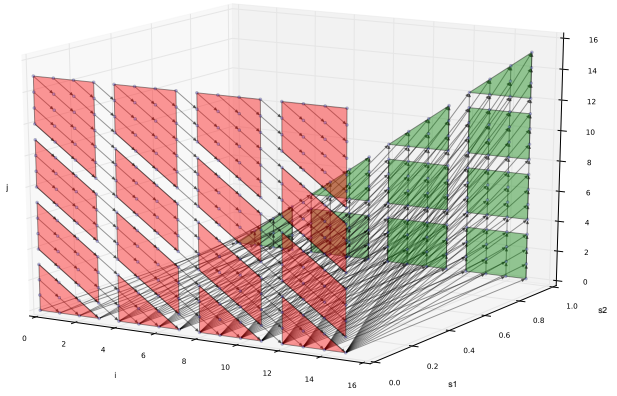


Fig. 5. Dependences and tiles of the size 4x4 for Example 5.

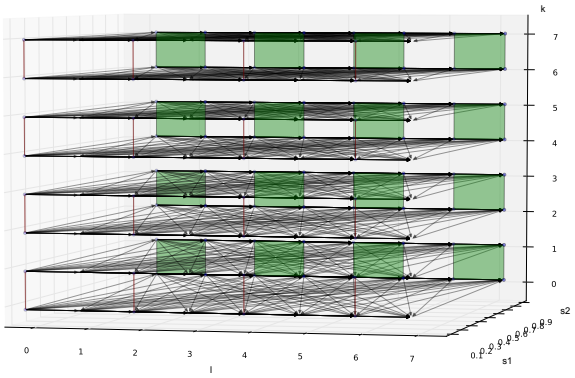


Fig. 6. Dependences and tiles of the size 2x2 for Example 6.

V. Related Work

Popular polyhedral libraries and compilers provide interface to visualization. The PolyLib tool [10] projects loop iteration domains by means of the VisualPolylib. LooPo [11] visualizes loop dependences before and after automatic parallelization. The 3D iteration space visualizer [12], [1] allows programmers to visualize and manipulate 3D dependence graphs, and to select a desired iteration space to start automatic parallelization search based mainly on unimodular transformations.

Clint [13] translates manipulations back to the polyhedral representation and ultimately transforms code to match visualization. This tool seems to be the most advanced visualizer and corresponds to extended versions of classical loop transformations (reordering, shifting, interchange, fusion, splitting, index set splitting, grain, reversal, skewing, tiling etc.).

The islplot[1] and Linpy tools [14] are based on islpy[2], a Python wrapper around Sven Verdoolaeges isl [3], the library for manipulating sets and relations of integer points bounded by linear constraints. Both the libraries are based on the matplotlib framework. Given the fact that TRACO is implemented also by means of the islpy interface, we considered these libraries for the presented approach. Islplot allows us to draw 2D maps and sets directly from islpy classes. However, domains of unions in isplot are drawn as separated figures. Therefore, we draw 2D polygons of tile points building convex hulls by means of Jarvis' March. Moreover, islplot is not integrated with matplotlib in 3D and provides only one function to draw sets in WebGL. Hence, for solids we used the interface provided by LinPy, which allows us to build the Polyhedron object from its constraints.

[1]http://tobig.github.io/islplot/

[2]https://documen.tician.de/islpy/

VI. Conclusion

Visualization of results of a dependence analysis and TRACO transformations carried out assists the development of parallel numerical programs. The approach complements analytical methods used in traditional automatic parallelizing compilers. Furthermore, the tool is helpful to investigate the use of interactive visualization for learning parallelization and the polyhedral model. In future, we plan to provide more interactive functions to the visualizer using event handling and object picking provided by matplotlib. We are going also to use the approach in order to design algorithms of arbitrary shaped tiles.

References

[1] Y. Yu and E. H. D'Hollander, "Loop parallelization using the 3d iteration space visualizer," *J. Vis. Lang. Comput.*, vol. 12, no. 2, pp. 163–181, Apr. 2001.

[2] M. Palkowski, T. Klimek, and W. Bielecki, "Traco: An automatic loop nest parallelizer for numerical applications," in *Computer Science and Information Systems (FedCSIS)*, Sept 2015, pp. 681–686.

[3] S. Verdoolaege, "Integer set library - manual," Tech. Rep., 2011. [Online]. Available: www.kotnet.org/~skimo//isl/manual.pdf,

[4] J. D. Hunter, "Matplotlib: A 2d graphics environment," *Computing In Science & Engineering*, vol. 9, no. 3, pp. 90–95, 2007.

[5] W. Pugh and D. Wonnacott, "An exact method for analysis of value-based array data dependences," in *Sixth Annual Workshop on Programming Languages and Compilers for Parallel Computing*. Springer-Verlag, 1993.

[6] Wayne Kelly et al., "The omega library interface guide," College Park, MD, USA, Tech. Rep., 1995.

[7] C. Bastoul, "Code generation in the polyhedral model is easier than you think," in *PACT'13 IEEE Intern. Conf. on Parallel Architecture and Compilation Techniques*, Juan-les-Pins, 2004, pp. 7–16.

[8] W. Bielecki, M. Palkowski, and T. Klimek, "Free scheduling for statement instances of parameterized arbitrarily nested affine loops," *Parallel Computing*, vol. 38, no. 9, pp. 518–532, Sep. 2012.

[9] A. Lim, G. I. Cheong, and M. S. Lam, "An affine partitioning algorithm to maximize parallelism and minimize communication," in *In Proceedings of the 13th ACM SIGARCH International Conference on Supercomputing*. ACM Press, 1999, pp. 228–237.

[10] V. Loechner, "PolyLib: A library for manipulating parameterized polyhedra," 1999. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.9.8197

[11] M. Griebl, "Automatic parallelization of loop programs for distributed memory architectures," 2004.

[12] Y. Yu, "A 3d-java tool to visualize loop-carried dependences," in *Applications, Proceedings of the International Conference ParCo'99*. College Press, 1999, pp. 17–20.

[13] O. Zinenko, C. Bastoul, and S. Huot, "Manipulating visualization, not codes," in *International Workshop on Polyhedral Compilation Techniques 2015 (IMPACT)*, 2015, p. 8.

[14] Mines Paristech, "Linpy documentation release 1.0," 2014.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 709–716
DOI: 10.15439/2016F366
ACSIS, Vol. 8. ISSN 2300-5963

Efficient parallel evaluation of block properties of sparse matrices

Ivan Šimeček, Daniel Langr
Czech Technical University in Prague
Faculty of Information Technology
Department of Computer Systems
Thákurova 9, 160 00, Praha, Czech Republic
Email: {xsimecek,langrd}@fit.cvut.cz

***Abstract*—Many storage formats for sparse matrices have been developed. Majority of these formats can be parametrized, so the algorithm for finding optimal parameters is crucial. For overall efficiency, it is important to reduce the execution time of this preprocessing. In this paper, we propose a new algorithm for the determination of the number of nonzero blocks of the given size in a sparse matrix. The proposed algorithm requires relatively a small amount of auxiliary memory. Our approach is based on the Morton reordering and bitwise manipulations. We also present a parallel (multithreaded) version and evaluate its performance and space complexity.**

I. Introduction

COMPUTATIONS with sparse matrices are widespread in scientific projects. Many storage formats for sparse matrices have been developed. The most straightforward approach is to accompany values of nonzero elements with their row and column indexes, which forms the *coordinate* format (COO, for details see Sec. I-C). If the local matrix nonzero elements are ordered lexicographically, then row indexes in COO can be substituted by the number of nonzero elements of each row. Such idea is represented by the *compressed sparse row* (CSR, for details see Sec. I-D) format. Due to matrix sparsity, the memory access patterns in common formats (like COO or CSR) are irregular and the utilization of cache suffers from low spatial and temporal locality, hence other (so called advanced) formats are used in practice. These advanced formats are usually parametrized,

A. General notation

We consider a matrix A of order $n \times n$, $A = (a_{i,j})$. The number of its nonzero elements is denoted by N. Matrix A is considered *sparse* if it is worth (for performance or any other reason) not to store this matrix in memory in a dense array. In the following text:

- We assume that indexes of all vectors and matrices start from zero.
- We assume that $1 \ll n \leq N \ll n^2$.
- The number of nonzero elements in submatrix B of matrix A is denoted by $\eta(B)$, so $\eta(A) = N$
- For any submatrix C, if $\eta(C) = 0$ then the submatrix C is called zero submatrix, otherwise it is called nonzero submatrix.
- If all elements in A have the same probability N/n^2 to be nonzero (independently on other elements) then A has a *uniform distribution* of nonzero elements.
- The parameter th denotes the number of threads used for the execution of an algorithm.

This work was supported by internal CTU grant No. SGS16/122/OHK3/1T/18 of the Czech Technical University in Prague. This work was supported by the IT4Innovations Centre of Excellence project (CZ.1.05/1.1.00/02.0070), funded by the European Regional Development Fund and the national budget of the Czech Republic via the Research and Development for Innovations Operational Programme, as well as Czech Ministry of Education, Youth and Sports via the project Large Research, Development and Innovations Infrastructures (LM2011033).

B. Banded matrices

Citing from Golub and Van Loan [1]:

Definition 1:

If all matrix elements are zero outside a diagonally bordered band whose range is determined by constants k_1 and k_2:

$$a_{i,j} = 0 \quad \text{if} \quad j < i - k_1 \quad \text{or} \quad j > i + k_2, \quad k_1, k_2 \geq 0.$$

Then the quantities k_1 and k_2 are called the left and right half-bandwidth, respectively. The bandwidth of the matrix (denoted by $\omega(A)$) is $k_1 + k_2 + 1$.

Definition 2: If $\omega(A) \ll n$, then A is *banded*.

C. The Coordinate (COO) format

The coordinate (COO) format is the simplest format for storing sparse matrices (see [2], [3]). The matrix A is represented by three linear arrays $values$, $xpos$, and $ypos$. The array $values[0, \ldots, N-1]$ stores the nonzero values of A, arrays $xpos[0, \ldots, N-1]$ and $ypos[0, \ldots, N-1]$ contain column and row indexes, respectively, of these nonzero values. The ordering of elements in this format is not prescribed.

D. The Compressed Sparse Row (CSR) format

The most common format for storing sparse matrices is the *compressed sparse row* (CSR) format (see [2]–[7]). The matrix A stored in the CSR format is represented by three linear arrays: $values$, $addr$, and ci. The array $values[0, \ldots, N-1]$ stores the nonzero elements of A, the array $addr[0, \ldots, n]$ contains indexes of initial nonzero elements of rows of A. The array $ci[0, \ldots, N-1]$ contains column indexes of nonzero elements of A.

E. Hierarchical formats

Many hierarchical formats are based on the idea of partition the matrix into square disjoint *blocks* of size $2^c \times 2^c$ rows/columns, where $c \in N^+$ is a formal parameter. In the further text, we will denote them as basic hierarchical format (BHF), for details see [6], [8], [9]. Coordinates of the upper left corners of these blocks are aligned to multiples of 2^c. Thus, indexes of nonzero elements are separated in two parts, indexes of blocks and indexes inside the blocks. Every such a region has *block row* and *block column* indexes. Let $B(c)$ denote the number of nonzero blocks for matrix A. The minimal number of nonzero blocks is equal to $B(c)_{min} = \lceil \frac{N}{2^{2c}} \rceil$, if all nonzero blocks contain only nonzero elements (i.e., are 100% dense). The maximal number of nonzero blocks is equal to $B(c)_{max} = \min\left(N, \lceil \frac{n}{2^c} \rceil^2\right)$, if each nonzero block contains exactly one nonzero element or if the whole matrix A is covered by nonzero blocks. This idea is for example behind formats: COOCOO format [10], ABHSF [8], [9], multilevel format [11], and so on. For all these formats, the optimal value of bits for each level should be computed. For this decision, the information about number of blocks $B(c)$ for given c is required.

F. Our requirements for an algorithm

Our assumptions and the requirements for an algorithm for computation of the number of nonzero blocks are as follows:

- The algorithm should be execution efficient (even in the multithreaded environment).
- Since, we are processing large sparse matrices, we assume that the space complexity (memory footprint) of the sparse matrix A is significant. We define an *in-place* algorithm as an algorithm that needs auxiliary data structures with space complexity strictly lower than the number of matrix nonzero elements. By in-place computation we mean an algorithm with $o(N)$ space complexity, in contrast to $O(N)$. It is the typical situation in HPC (high performance computing) that the matrix typically represents the largest object stored in the main memory. Hence, the algorithm should be also space-efficient, i.e., space complexity of all its temporary data should be much lower than space complexity of the matrix. When the computation of the number of nonzero blocks of a matrix cannot be performed in-place, then there might not be enough free memory to make this computation at all.
- In real situation, we don't need to compute the number of nonzero blocks for all c from $[1, \ldots, \lceil \log n \rceil]$ because some block sizes from this interval are simply too small or too large for the given purpose. Thus, we need to compute the number of nonzero blocks only for c from the given interval $[c\_min, \ldots, c\_max]$.

G. Overview of state-of-the-art

As far as we know, there are only two algorithms for the computation of the number of nonzero blocks (e.g., in [10]), both of them are based on sets. There are two main reasons for such low number:

- Authors rarely present algorithms for preprocessing of matrices, which are necessary for assessment of the suitability of their formats for a given application [6]. We have not found a case where format authors provided efficient parallel implementations of algorithms for the computation of the number of nonzero blocks in non-experimental forms.
- Register blocking formats (like SPARSITY [12] or [13] or CARB [5], [14]) store a matrix as a set of small dense blocks. Blocks can be linear (horizontal, vertical, or diagonal) or rectangular, a usual range of size is from 2 to 20. Since the block-size is not limited to the power of 2 and optimization criteria are more complex, a special transformation algorithm is used.
- Some formats (e.g., [11], [15]–[17]) skip this computation and use "typically good" values of the block-size.

The idea behind the two found solutions is to evaluate the number of blocks $B(c)$ for all values of parameter c from $[c\_min, \ldots, c\_max]$ using a set. We will consider four implementations of such an algorithm using different data-structure for implementation of the set.

1) First algorithms based on sets: All nonzero elements are mapped to block coordinates. These block coordinates are mapped to index $i \in \langle 0, \ldots, \lceil n/2^c \rceil^2 \rangle$ and put into set U. Finally, the cardinality of U is determined. It is equal to the number of blocks $B(c)$.

Algorithm 1 Determination of the number of blocks $B(c)$

```
 1: procedure NumberOfBlocks1(I,c)
Input: In = a matrix in the CSR format
Input: c = the parameter (logarithm of block size)
Output: B = the number of blocks
 2:     B ← 0;  d ← 2^c;  p ← ⌈n/d⌉;
 3:     construct the set U;
 4:     U = ∅;
 5:     for y ← 1, In.n do
 6:         for j ← In.addr[y], In.addr[y+1] − 1 do
 7:             x ← In.ci[j];
 8:             i ← ⌊y/d⌋ · p + ⌊x/d⌋;
 9:             put the element i into U;
10:     B ← |U|;
11:     return B;
```

The time complexity of Algorithm 1, $T_1(n, N)$, consists of

- t_1 = time complexity of creating an empty set U at codeline (4),
- $t_2 = N \cdot t_{\text{ins}}$ = time complexity of inserting N elements at codeline (9) into the set U,
- t_3 = time complexity of computing the cardinality of U at codeline (10).

The time complexity depends on the data structure used for implementing the set U. Let $d = 2^c$ and $p = \lceil n/d \rceil$. Let us consider four basic implementations:

1) a bit array (of size of $p^2 = \lceil n/d \rceil^2 = \lceil n/2^c \rceil^2$ bits): Then

- t_1 is proportional to the range of indices i: $t_1 = O(p^2)$,
- $t_2 = N \cdot O(1) = O(N)$,
- t_3 is also proportional to the range of indices i: $t_3 = O(p^2)$.

The total time complexity is $T_1(n, N) = O(N + p^2) = O(N + \lceil n/2^c \rceil^2)$ and the space complexity is $O(p^2) = O(\lceil n/2^c \rceil^2)$. These complexities are high (especially for small values of c) which makes this approach inefficient.

2) A linked list:

- $t_1 = N \cdot O(1) = O(N)$,
- $t_2 = N \cdot O(B(c)_{max})$,
- $t_3 = O(B(c)_{max})$.

The total time complexity is $T_1(n, N) = O(N \cdot B(c)_{max})$ and the space complexity is $O(B(c)_{max})$. This complexity is high which makes this approach inefficient.

3) A balanced binary search tree:

- $t_1 = O(1)$,
- t_2 is proportional to N and to the logarithm of the size of binary representation of index i: $t_{\text{ins}} = O(2 \log p) = O(\log(n/2^c)) = O(\log n - c)$,
- $t_3 = O(B(c)_{max})$.

The total time complexity is $T_1(n, N) = O(B(c)_{max} + N(\log n - c))$ and the space complexity is $O(B(c)_{max})$, hence the approach is quite execution efficient, but space inefficient (especially for small values of c).

4) A hash table of size l (we assume a closed hashing scheme):

- $t_1 = O(l)$,
- $t_2 = N \cdot O(1)$ in ideal case,
- $t_3 = O(l)$.

We must carefully choose the value of parameter l to satisfy the no-collision assumption during the insertion of elements, this parameter should be greater than $B(c)_{max}$. Then, the total time complexity is $T_4(n, N) = O(N + l) = O(N + B(c)_{max})$ and the space complexity is $O(B(c)_{max})$, hence the approach is execution efficient, but space inefficient (especially for small values of c).

2) An improved algorithm: We can improve Algorithm 1 as follows: The matrix A is divided into disjoint horizontal strips whose height is equal to 2^c. Then the cardinality of set U (i.e., the number of blocks) is determined in each strip separately. The value of $B(c)_{max}$ is decreased to $\lceil \frac{n}{2^c} \rceil$ that occurs if the whole strip is covered by nonzero regions.

Algorithm 2 Determination of the number of blocks $B(c)$ (improved)

```
1:  procedure NumberOfBlocks2(In,c)
Input: In = a matrix in the CSR format
Input: c = the parameter (logarithm of block size)
Output: B = the number of blocks
2:      B ← 0;  old ← −1;
3:      d ← 2^c;
4:      create the set U;
5:      U = ∅;
6:      for y ← 1, In.n do
7:          if ⌊y/d⌋ > old then
8:              B ← B + |U|;
9:              U = ∅;
10:             old ← ⌊y/d⌋;
11:         for j ← In.addr[y], In.addr[y + 1] − 1 do
12:             x ← In.ci[j];
13:             i ← ⌊x/d⌋;
14:             put the element i into U;
15:     B ← B + |U|;
16:     return B;
```

The time complexity of Algorithm 2, $T_2(n, N)$, consists of:

- $t_1 = p \cdot t_{\text{init}}$ = time complexity of creating empty sets at codelines (5) and (9),
- $t_2 = N \cdot t_{\text{ins}}$ = time complexity of inserting N elements at codeline (14),
- $t_3 = p \cdot t_{\text{enum}}$ = time complexity of determining the cardinality of the set at codelines (8) and (15).

Four basic implementations of the set data type provide the following time complexities:

1) a bit array (of size of $\lceil n/d \rceil = \lceil n/2^c \rceil$ bits):
   - $t_{\text{init}} = O(p)$,
   - $t_{\text{ins}} = O(1)$,
   - $t_{\text{enum}} = O(p)$,

   $T_2(n, N) = O(N + p^2)$ and the space complexity is $O(p)$. So, the space complexity decreased compared to same implementation of the set data type in Algorithm 1, but the time complexity remains high, especially for small values of c.
2) a linked list:
   - t_{init} remain the same as in the previous case,
   - $t_{\text{enum}} = t_{\text{ins}} = O(B(c)_{max})$ drops due to $B(c)_{max}$ decrease.

   The total time and the space complexities remain the same compared to the same implementation of the set data type in Algorithm 1.
3) a balanced binary search tree:
   - t_{init} remains the same as in the previous case,
   - $t_{\text{ins}} = O(\log p) = O(\log(n/2^c)) = O(\log n - c)$.
   - $t_{\text{enum}} = O(B(c)_{max})$ drops due to $B(c)_{max}$ decrease.

The total time complexity remains the same compared to the same implementation of the set data type in Algorithm 1.

4) a hash table with the open hashing scheme (the size of the hash table is equal to l). It is much easier to satisfy the no collision assumption by insertion of elements (with lesser value of l).
   - t_{ins} is the same,
   - t_{init} and t_{enum} drops due to $B(c)_{max}$ and l decrease.

   The total time complexity remains the same.

3) The summary of state-of-the-art algorithms: None of the algorithm and implementation satisfies the requirements describe in Sec. I-F. The improved algorithm reduces memory requirements which allows implementations to be space efficient. On the other hand, existing algorithms are execution inefficient (especially for small values of c). The situation is getting worse for multithreaded execution. If each strip is computed in parallel then every thread has independent instance of the set and memory requirements will be th-times higher. If single strip is computed by multiple threads then the only one (shared) set is used, but every access is a critical section, hence the speedup would be minimal.

II. Our new approach

A. Main idea

In our approach, we utilize reordering of nonzero elements according to so-called Morton order (for details see [18]). Morton ordering is a mapping from an i-dimensional space onto a linear list of numbers. If we want to convert a certain set of integer coordinates to a Morton code, we have to interleave the binary representations of each coordinate. Here is an example of transformation from 3D coordinates into Morton code.

$$(x, y, z) = (5, 9, 1)_{10} = (0101, 1001, 0001)_2$$

Interleaving the bits results in: $(010\ 001\ 000\ 111)_2 = (1095)_{10}$-th cell along the co called Z-curve. For the determination of the number of blocks, we use the following lemma.

Lemma 1: Morton codes for all elements inside the block of size 2^c that is aligned to multiple of 2^c are the same except $2 \cdot c$ least significant bits.

The proof of Lemma 1 is obvious (based on the construction of Morton code). Hence, we can design an algorithm based on this lemma: in a sorted sequence of nonzero elements, we count differences in Morton codes of two adjacent items (more exactly: the positions of highest bit set in the result of logical XOR of two adjacent items). We call this algorithm Morton-based (see Algorithm 3).

Algorithm 3 Determination of the number of blocks $B(c)$ (new)

```
 1: procedure NumberOfBlocks3(In,c)
Input: A = a matrix in the COO format
Input: c = the parameter (logarithm of block size)
Output: B = the number of blocks
 2:     for j ← 1, c_max do
 3:         number[j] ← 1;
 4:     add to every nonzero element its Morton code;
 5:     sort nonzero elements on its Morton code;
 6:     old ← A[0].Morton;
 7:     for i ← 1, N do
 8:         new ← A[i].Morton;
 9:         diff ← XOR(new, old);
10:         old ← new;
11:         k ← round_up(Highest1(diff)/2);
12:             ▷ Highest1 = the index of the position of the
    highest bit set
13:         for j ← 1, k do
14:             number[j] ← number[j] + 1;
15:     return number[];
```

The time complexity of Algorithm 3, $T_3(n, N)$, consists of:

- $t_1 = N$ = time complexity of generating Morton codes at codeline (4),
- $t_2 = N \log N$ = time complexity of sorting. We assume the sorting algorithm with complexity $O(i \log i)$ for a array of length i.
- $t_3 = N \cdot c\_max$ = time complexity of for-cycle at codeline (7) mainly inner loop at codeline (13).

Overall time complexity is $T_3(n, N) = N(c\_max + \log N)$, so this algorithm is efficient, but requires additional space for Morton codes proportional to N. To avoid this, we propose two solutions:

1) Morton codes are not stored explicitly. They can overwrite COO storage format (arrays $xpos$ and $ypos$). After determination of the number of blocks, the coordinates (original values in these arrays) will be restored from Morton codes.
2) Computation of Morton code can be included in a compare function of sorting algorithm at codeline (5) in Algorithm 3.

B. Example of algorithm usage

Let us assume a very small example of a sparse matrix with $n = 8$ and $N = 12$. Instead of the values of the matrix elements, we deal only with binary flags indicating the existence of nonzero elements.

$$\mathbf{M^{(0)}} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

The steps of the new algorithm are as follows:

Step 1: (codeline 4) For each nonzero element, the Morton code is computed. Morton codes (for elements in lexicographic order) = { 000000, 010101, 000011, 010110, 001100, 001111, 011010, 110011, 101000, 111100, 101011, 111111 }.

Step 2: (codeline 5) The whole sequence is sorted according to Morton codes.

Step 3: (codeline 7-14) We count difference in Morton codes of adjacent items. Morton codes (for elements the highest different bit set is marked) ={ 000000, 0000$\bar{1}$1, 00$\bar{1}$100, 0011$\bar{1}$1, 0$\bar{1}$0101, 0101$\bar{1}$0, 01$\bar{1}$010, $\bar{1}$01000, 1010$\bar{1}$1, 1$\bar{1}$0011, 11$\bar{1}$100, 1111$\bar{1}$1 }. The counters (in variable $numbers$) are increased according to the position of the highest bit set (e.g., 0000$\bar{1}$1 $\Rightarrow$ increase $numbers[1]$ by one, 00$\bar{1}$100 $\Rightarrow$ increase $numbers[1]$ and $numbers[2]$ by 1, etc.).

Step 4: (codeline 15) After traversing the example sequence, the counters are set to $B(c = 1) = numbers[1] = 7$, $B(2) = 4$, $B(3) = 1$.

C. Parallelization

1) SW technologies: The OpenMP API specification [19] is defined by a collection of compiler directives, library routines and environment variables extending the C, C++ and Fortran languages. These can be used to create portable parallel programs utilizing shared memory. The core of OpenMP is the so called *fork-join model* execution model. An application employing OpenMP usually begins as a single thread program and during execution uses multiple threads or even other devices to perform parallel tasks.

The OpenMP API provides a relaxed-consistency, shared memory model. All threads have access to the memory and each may have its own temporary view of the memory (which represents cache or other local storage used for caching). Each thread also have access to thread private memory, which cannot be accessed by any other thread. A single access to a variable is not guaranteed to be atomic with respect to other accesses of that variable, since it may be implemented with multiple load or store instructions. If multiple threads write without synchronization to the same memory unit, the data race occurs.

2) Multithreaded execution: The parallel (multithreaded) version of the algorithm is represented by Alg. 4. We simply make all steps (described in Sec. II-B) parallel:

Step 1 (Computation of Morton codes): The parallelization of this step is straightforward.

Step 2 (Sorting of Morton codes): The parallelization of this step is straightforward by using any parallel in-place sort (e.g., `sort` method from `std::algorithm` [20] or AQsort [21])

Step 3 (Counting the differences): The whole sequence of sorted Morton codes is partitioned into th disjoint chunks of consecutive elements. Each chunk contains approximately the same number of elements. Each thread counts the differences in the assigned chunk independently. After this computation, local (thread private) instances of $l\_number[]$ are summed up to the global values ($number[]$).

Algorithm 4 Determination of the number of blocks $B(c)$ (new, parallel version)

```
1:  procedure NumberOfBlocks4(In,c)
Input: A = a matrix in the COO format
Input: c = the parameter (logarithm of block size)
Output: B = the number of blocks
2:      parallel add to every nonzero element its Morton code;
3:      parallel sort nonzero elements on its Morton code;
4:      len = N/th;
5:      start of parallel block
6:      tid = get_tid_of_current_thread();
7:      if tid = 0 then
8:          old ← A[0].Morton;
9:          start ← 1;
10:         for j ← 1, c_max do
11:             l_number[j] ← 1;
12:     else
13:         old ← A[tid · len − 1].Morton;
14:         start ← tid · len;
15:         for j ← 1, c_max do
16:             l_number[j] ← 0;
17:     for i ← start, (tid + 1) · len − 1 do
18:         new ← A[i].Morton;
19:         diff ← XOR(new, old);
20:         old ← new;
21:         k ← round_up(Highest1(diff)/2);
22:         for j ← 1, k do
23:             l_number[j] ← l_number[j] + 1;
24:     end of parallel block
25:     parallel reduction(+) of l_number into number;
26:     return number[];
```

D. Modification of algorithm suitable for the CSR format

Our algorithms (see Alg. 3 and 4) require the input storage format that allows to reorder/sort its nonzero elements. Hence, we use the COO format. This format is used in HPC, but the CSR format is more frequent (see for example [22], [23]). For the CSR format, we propose the following modification.

Step 1 The whole matrix is partitioned into 2D disjoint regions (=chunks of consecutive rows). Starting row of each

region are aligned to multiple of 2^{c_max}.

Step 2 For each region all nonzero elements in this region are extracted and their Morton codes are stored into a temporary array. These regions are proceeded by Alg. 3.

Step 3 The results for all regions are summed up to the global values.

The parallelization of this algorithm is easy: each region can be computed independently by a different thread. For good load balancing in OpenMP API even for matrices with non-uniform distribution (e.g., for banded matrices), so-called dynamic scheduling strategy is used.

Matrix	abbr	n	N	avg_per_row
circuitM5	m1	$5.56 \cdot 10^6$	$5.95 \cdot 10^7$	10.7
nlpkkt120	m2	$3.54 \cdot 10^6$	$5.02 \cdot 10^7$	14.1
ldoor	m3	$9.52 \cdot 10^5$	$2.37 \cdot 10^7$	24.9
TSOPF_RS_b2383	m4	$3.81 \cdot 10^4$	$1.62 \cdot 10^7$	42.5
mouse_gene	m5	$4.51 \cdot 10^4$	$1.45 \cdot 10^7$	32.1
t2em	m6	$9.25 \cdot 10^5$	$4.59 \cdot 10^6$	5.0
bmw7st_1	m7	$1.41 \cdot 10^5$	$3.74 \cdot 10^6$	26.5
amazon0312	m8	$4.01 \cdot 10^5$	$3.20 \cdot 10^6$	8.0
thread	m9	$2.97 \cdot 10^4$	$2.25 \cdot 10^6$	75.8
gupta2	m10	$6.21 \cdot 10^4$	$2.16 \cdot 10^6$	34.8

TABLE I: Characteristics of the testing matrices and their abbrevation in the further text.

III. Evaluation of the results

A. Testing matrices

We have used 10 testing matrices from various application domains from the University of Florida Sparse Matrix Collection [24]. Table I shows the characteristics of the testing matrices.

B. Used HW and SW

The execution times were measured on a server with following HW and SW parameters:

- 2 × CPU Intel Xeon Processor E5-2620 v2 (15MB L3 cache per CPU),
- CPU cores: 6 per CPU, 12 in total,
- Memory size: 32 GB RAM, total max. memory bandwidth: 51.2 GB/s,
- Peak single precision floating point performance 0.48 Tflops (using base clocks),
- OS Linux, C++ compiler (g++) version 4.8.3 with switches `-O3 -march=native -mavx -fopenmp`.

We measure elapsed wall clock times using OpenMP function `omp_get_wtime()`.

C. Evaluation of results

1) Comparison of sequential algorithms: Tables II and III show the comparison of measured times for different algorithms for the determination of the number of blocks. From this table, we can conclude that our Morton-based algorithm is always faster for larger matrices. The reason is the following: the time complexity of classical algorithm is $O(n^2+N)$, hence for smaller matrices (with small order) both components are approximately the same and algorithm is execution-efficient. On the other hand, for larger matrices in the complexity of the classical algorithm the component n^2 become dominant and algorithm is execution-inefficient. For small matrices the initial overhead of the Morton-based algorithm is significant. For larger matrices, this overhead become negligible and this algorithm is execution-efficient.

Matrix	CL($th=1$)	CL($th=12$)	NEW($th=1$)	NEW($th=12$)
m1	2834	265.9	3.56	0.348
m2	1149	107.8	3.19	0.267
m3	82.8	7.88	1.28	0.107
m4	0.313	0.053	0.959	0.243
m5	0.523	0.044	1.19	0.103
m6	77.3	7.22	0.209	0.018
m7	1.49	0.124	0.204	0.017
m8	14.2	1.26	0.284	0.024
m9	0.091	0.008	0.132	0.012
m10	0.284	0.024	0.142	0.020

TABLE II: Measured times in seconds for the determination of the number of blocks ($c\_max = 8$): CL denotes the classical improved algorithm, NEW denotes the Morton-based algorithm.

Matrix	CL($th=1$)	CL($th=12$)	NEW($th=1$)	NEW($th=12$)
m1	2839	414	4.38	1.52
m2	1151	168	4.19	0.476
m3	83.4	11.7	1.67	0.212
m4	0.562	0.566	0.563	0.565
m5	0.742	0.746	0.743	0.749
m6	77.4	11.5	0.273	0.041
m7	1.41	0.670	0.259	0.135
m8	14.4	2.39	0.388	0.130
m9	0.128	0.128	0.157	0.157
m10	0.309	0.325	0.154	0.154

TABLE III: Measured times in seconds for the determination of the number of blocks ($c\_max = 16$): CL denotes the classical (improved) algorithm, NEW denotes the Morton-based algorithm.

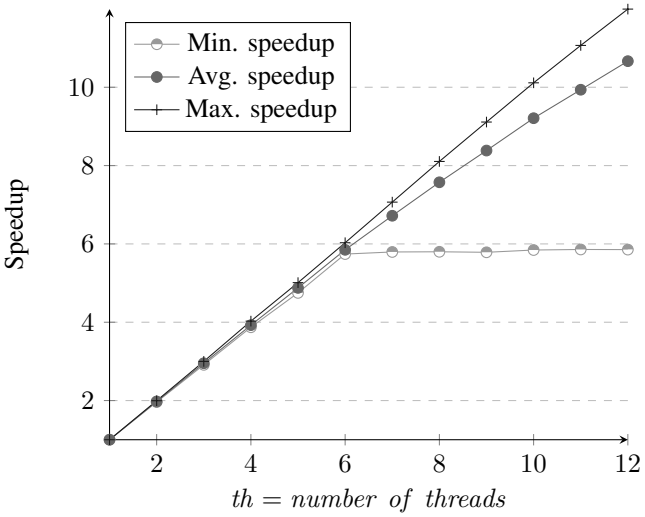


Fig. 1: Speedups for classical (improved) algorithm with $c\_max = 8$.

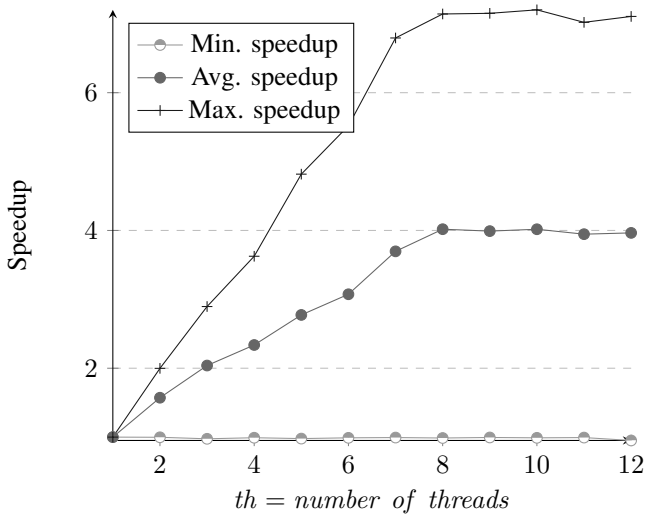


Fig. 2: Speedups for classical (improved) algorithm with $c\_max = 16$.

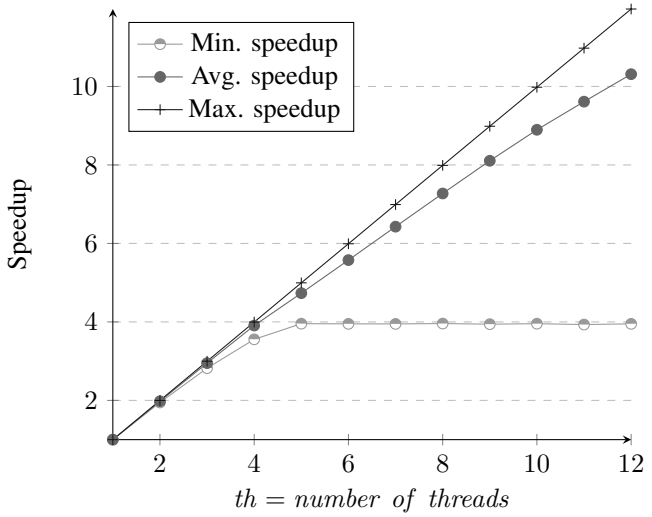


Fig. 3: Speedups for Morton-based algorithm with $c\_max = 8$.

2) Comparison of parallel algorithms: Tables II and III show the comparison of measured times for different algorithms for the determination of the number of blocks. From Fig. 1, 2, 3, and 4 we can conclude that both algorithms scale very well (speedup is equal to the number of threads) for $c\_max = 8$. For $c\_max = 16$, the scalability is getting worse since parallelization is based on regions (see Section I-G3 and II-D). For small matrices (with small order), the parameter 2^{c_max} is comparable with the order of the matrix. In this case, the load-balance is not good and majority of threads is idle and the speedup is almost independent on the number of threads.

IV. Conclusions

This paper presents the design of a new algorithm for the for the calculation of the number of blocks in sparse matrices. This

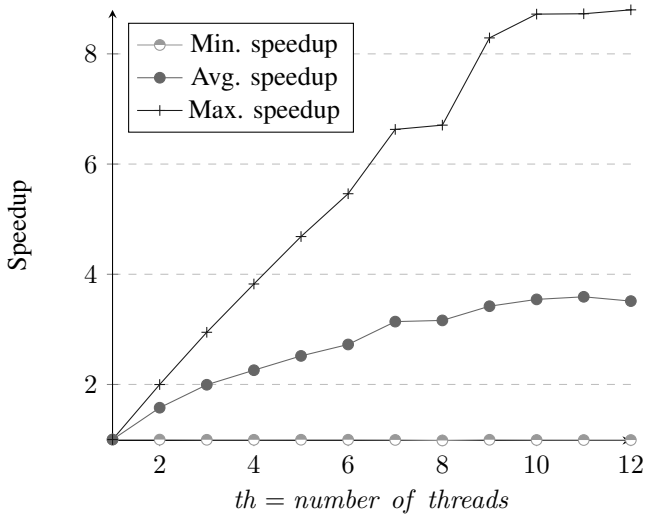


Fig. 4: Speedups for Morton-based algorithm with $c\_max = 16$.

algorithm is crucial for preprocessing of matrices into some advanced storage formats. We have also developed a parallel version of this algorithm. We performed experiments on the parallel system and their results showed that the proposed algorithm is both execution- and space-efficient.

Acknowledgments

The authors would like to thank M. Václavík of the Czech Technical University in Prague for providing an access to the Star university cluster.

References

[1] G. H. Golub and C. F. Van Loan, *Matrix Computations (3rd ed.)*. Baltimore: Johns Hopkins, 1996.

[2] Y. Saad, *Iterative Methods for Sparse Linear Systems*, 2nd ed. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2003.

[3] R. Barrett, M. Berry, T. F. Chan, J. Demmel, J. Donato, J. Dongarra, V. Eijkhout, R. Pozo, C. Romine, and H. V. der Vorst, *Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods*, 2nd ed. Philadelphia, PA: SIAM, 1994.

[4] I. Šimeček and P. Tvrdík, "A new approach for accelerating the sparse matrix-vector multiplication," in *Proceedings of 8th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC '06)*. Los Alamitos: IEEE Computer Society, 2006, pp. 156–163. [Online]. Available: http://dl.acm.org/citation.cfm?id=1264261

[5] ——, "Sparse matrix-vector multiplication — final solution?" in *Parallel Processing and Applied Mathematics*, ser. PPAM'07, vol. 4967. Berlin, Heidelberg: Springer-Verlag, 2008, pp. 156–165. [Online]. Available: http://www.springerlink.com/content/48x1345471067304/

[6] D. Langr and P. Tvrdík, "Evaluation criteria for sparse matrix storage formats," *IEEE Transactions on Parallel and Distributed Systems*, vol. 27, no. 2, pp. 428–440, 2016.

[7] B. Bylina, J. Bylina, P. Stpiczyński, and D. Szałkowski, "Performance analysis of multicore and multinodal implementation of spmv operation," in *Proceedings of the 2014 Federated Conference on Computer Science and Information Systems*, ser. Annals of Computer Science and Information Systems, M. P. M. Ganzha, L. Maciaszek, Ed., vol. 2. IEEE, 2014, pp. pages 569–576. [Online]. Available: http://dx.doi.org/10.15439/2014F313

[8] D. Langr, I. Šimeček, P. Tvrdík, T. Dytrych, and J. P. Draayer, "Adaptive-blocking hierarchical storage format for sparse matrices," in *Federated Conference on Computer Science and Information Systems (FedCSIS)*. 345 E 47TH ST, NEW YORK, NY 10017 USA: IEEE Xplore Digital Library, September 2012, pp. 545–551.

[9] D. Langr, I. Šimeček, and P. Tvrdík, "Storing sparse matrices in the adaptive-blocking hierarchical storage format," in *Proceedings of the Federated Conference on Computer Science and Information Systems (FedCSIS 2013)*. IEEE Xplore Digital Library, September 2013, pp. 479–486.

[10] I. Šimeček, D. Langr, and P. Tvrdík, "Space-efficient sparse matrix storage formats for massively parallel systems," in *High Performance Computing and Communication and 2012 IEEE 9th International Conference on Embedded Software and Systems (HPCC-ICESS)*, ser. HPCC'12, Liverpool, Great Britain, june 2012, pp. 54–60.

[11] I. Šimeček and D. Langr, "Space and execution efficient formats for modern processor architectures," in *2015 17th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC)*, Sept 2015, pp. 98–105.

[12] E. Im, *Optimizing the Performance of Sparse Matrix-Vector Multiplication - dissertation thesis*, University of California at Berkeley, 2001.

[13] M. Martone, M. Paprzycki, and S. Filippone, "An improved sparse matrix-vector multiply based on recursive sparse blocks layout," in *Large-Scale Scientific Computing*, ser. Lecture Notes in Computer Science, I. Lirkov, S. Margenov, and J. Waśniewski, Eds. Springer Berlin Heidelberg, 2012, vol. 7116, pp. 606–613. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-29843-1_69

[14] P. Tvrdík and I. Šimeček, "A new diagonal blocking format and model of cache behavior for sparse matrices," in *Proceedings of the 6th International Conference on Parallel Processing and Applied Mathematics*, ser. PPAM'05, vol. 12, no. 4. Poznan, Poland: Springer-Verlag, 2005, pp. 164–171. [Online]. Available: http://dl.acm.org/citation.cfm?id=2096870.2096894

[15] I. Šimeček and D. Langr, "Space-efficient sparse matrix storage formats with 8-bit indices," in *Seminar on Numerical Analysis*. Liberec: Technical University of Liberec, 2012, pp. 161–164. [Online]. Available: http://shimi.webzdarma.cz/vyzkum/sna12/article.pdf

[16] M. Martone *et al.*, "On the usage of 16 bit indices in recursively stored sparse matrices," *Symbolic and Numeric Algorithms for Scientific Computing*, vol. 0, pp. 57–64, 2010.

[17] ——, "Use of hybrid recursive CSR/COO data structures in sparse matrices-vector multiplication," in *Proceedings of the International Multiconference on Computer Science and Information Technology*, Wisla, Poland, October 2010.

[18] G. M. Morton, *A computer Oriented Geodetic Data Base; and a New Technique in File Sequencing*. IBM Ltd., 1966.

[19] OpenMP Architecture Review Board, "Openmp application program interface," online, 2013. [Online]. Available: http://www.openmp.org/mp-documents/OpenMP4.0.0.pdf

[20] J. Singler and B. Konsik, "The gnu libstdc++ parallel mode: Software engineering considerations," in *Proceedings of the 1st International Workshop on Multicore Software Engineering*, ser. IWMSE '08. New York, NY, USA: ACM, 2008, pp. 15–22. [Online]. Available: http://doi.acm.org/10.1145/1370082.1370089

[21] D. Langr, "Parallel multi-array in-place sort with openmp." [Online]. Available: https://github.com/DanielLangr/AQsort

[22] J. Cáceres, B. Barán, and C. Schaerer, "Implementation of a distributed parallel in time scheme using petsc for a parabolic optimal control problem," in *Proceedings of the 2014 Federated Conference on Computer Science and Information Systems*, ser. Annals of Computer Science and Information Systems, M. P. M. Ganzha, L. Maciaszek, Ed., vol. 2. IEEE, 2014, pp. pages 577–586. [Online]. Available: http://dx.doi.org/10.15439/2014F340

[23] S. Fialko, "Parallel finite element solver for multi-core computers," in *Computer Science and Information Systems (FedCSIS), 2012 Federated Conference on*, Sept 2012, pp. 525–532.

[24] T. A. Davis, "The university of florida sparse matrix collection," *NA DIGEST*, vol. 92, 1994.

1st Workshop on Constraint Programming and Operation Research Applications

THE aim of the CPORA-Workshop on Constraint Programming and Operation Research Applications is to bring together interested researchers from constraint programming/constraint logic programming (CP/CLP), operations research (OR) and artificial intelligence (AI) to present new techniques or new applications in decision support, combinatorial optimization, modeling and control processes arising in manufacturing, transportation, telecommunication, computer networks, logistic systems etc. and to provide an opportunity for researchers in one area to learn about techniques in the others. The aim of this workshop is share ideas, projects, researches results, models, experiences etc. associated with CP/CLP/OR/AI and to give researchers the opportunity to show how the integration of techniques from different fields can lead to interesting results on large and complex problems. Additionally, we would like to stimulate the communication between researchers working on different fields and practitioners who need reliable and efficient modelling and computational methods for industrial and business processes.

Contributions containing of both: the theoretical and practical results obtained in this area are welcome.

TOPICS

- Constraint programming/Constraint logic programming,
- Mathematical programming,
- Constraint Satisfaction Problem,
- Logic programming,
- Hybrid methods,
- Network programming,
- Petri-Nets,
- Knowledge methods,
- Soft computing (FL, GA, NN etc.),
- Answer Set Programming (ASP),
- The boolean satisfiability problem (SAT).
- Manufacturing,
- Multimodal processes management,
- Project management,
- Supply chain management,
- Modeling and planning production flow,
- Production scheduling,
- Multimodal social networks,
- Intelligent transport and passenger routing,
- Network knowledge modeling,
- Transportation networks.

EVENT CHAIRS

- **Bocewicz, Grzegorz,** Koszalin University of Technology, Poland
- **Sitek, Pawel,** Kielce University of Technology, Poland

PROGRAM COMMITTEE

- **Banaszak, Zbigniew,** Warsaw University of Technology, Poland
- **Bzdyra, Krzysztof,** Koszalin University of Technology
- **Gola, Arkadiusz,** Lublin University of Technology, Poland
- **Hajduk, Mikuláš,** Technical University of Kosice, Slovakia
- **Nielsen, Izabela Ewa,** Aalborg University, Denmark
- **Nielsen, Peter,** Aalborg University, Denmark
- **Relich, Marcin,** University of Zielona Gora, Poland
- **Terkaj, Walter**
- **Türkyılmaz, Ali,** Fatih University
- **Wikarek, Jarosław,** Kielce University of Technology, Poland

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 719–724
DOI: 10.15439/2016F228
ACSIS, Vol. 8. ISSN 2300-5963

Combinatorial Portfolio Selection with the ELECTRE III method: Case study of the Stock Exchange of Thailand (SET)

Veera Boonjing
International College,
King Mongkut's Institute of Technology Ladkrabang,
Ladkrabang, Bangkok, Thailand
(email: veera.bo@kmitl.ac.th)

Laor Boongasame
Department of Business Computer,
Bangkok University, Bangkok,
Thailand (email: laor.b@bu.ac.th)

***Abstract*—**Various techniques of portfolio selection are applied to interpret the status of the market and predict the market's future trend, but they are not beneficial to small investors because these techniques should be administered by an expert. In addition, these techniques desire accumulation of data about the market and complicated calculations, which is too much effort for individual small investors. Therefore, portfolio selection with two significant financial ratios using the ELECTRE III method is proposed for these investors to make trading decisions. In order to demonstrate the effectiveness of this new method, it is compared to the situation where a fix percentage allocation existed and data was collected from the Stock Exchange of Thailand (SET).

I. Introduction

The Stock Exchange of Thailand (SET) is the national stock exchange of Thailand. Retail investors in Thailand have long been a majority of players. They account for approximately 41%, whereas foreigner and institution flows are 36% and 21% of daily trading value respectively [13]. Since many factors (such as political and economic factors) are likely to influence the trend of the market [3], forecasting the trend requires various market analysis techniques [2, 5, 10, 11, 19, 22]. In addition, there are a number of machine learning techniques proposed as a solution to the problem such as reinforcement learning [17, 18, 20], neural networks [9, 12], genetic algorithms [1, 15, 16], decision trees [24], support vector machines [4, 14, 23], and boosting and expert weighting [6, 7, 8]. Arthur Samuel in 1959 defined machine learning as the "field of study that gives computers the ability to learn without being explicitly programmed". Such techniques build a model from an example training set of input observations in order to make data-driven predictions or decisions expressed as outputs. Although machine learning techniques are applied to interpret the status of the market and predict the market's future trend, but they are not beneficial to small investors because these techniques desire both accumulation of training data set about the market and complicated calculations to make data-driven predictions, which is too much effort for an individual small investor. Therefore, portfolio selection with only two significant financial ratios using the ELECTRE III method is proposed for those investors to make trading decisions. The two significant financial ratios are net profit margin and dividend yield. The Net profit margin is a ratio of profitability calculated as after-tax net income (net profits) divided by sales (revenue) and displayed as a percentage; whereby consideration focuses on stocks with high net profit margin. Dividend yield is the amount that a company pays to its shareholders annually for their investments; whereby consideration focuses on stocks with high dividend yield. It is expressed as a percentage and indicates attractiveness of investing in a company's stocks. The ELECTRE III method [21] is the most popular one of the outranking methods in Multi Criteria Decision Making (MCDM). Its performance of alternatives on each criterion in outranking is compared in pairs. An alternative a is said to outrank an alternative b if it performs better on some criteria and at least as well as b on all others. The outranking relation in the ELECTRE III is a fuzzy binary relation. It uses three distinct thresholds (indifference, preference, and veto) to incorporate the uncertainties that are inherent in most influence valuations. In addition, the ELECTRE III method is less of a complex than the machine learning techniques because it follows strictly static program instructions.

This report proceeds as follow. In the next section, previous related literatures are reviewed. Then the research methodology and data used are discussed. Empirical results found in the study are then presented and analyzed. Lastly, conclusion, implications, and limitations together with suggestion for further study are presented.

II. Backgrounds

An ELECTRE is a family of multi-criteria decision analysis methods (Roy, B. (1978)). The ELECTRE is working on the concrete, multiple criteria, and real-world problem of how firms can decide on new activities and had encountered problems using a weighted sum technique. It uses several mathematical functions to indicate the dominant degree of one alternative over the remaining ones. Additionally, it also facilitates comparisons between alternative schemes by using a weighted sum technique. The outranking relationships between alternatives are constructed and exploited eventually.

In order to be consistent with the basic concept of similarity in case based reasoning, different terminologies from the classical ELECTRE is used.

Let the distance between case a_k and case a_l on the j feature be denoted by $|a_{kj} - a_{lj}|$ or d_{lkj}. Let w_j express the weight of the feature.

Definition 1: The indifferent threshold of criterion j q_j: a_k and a_l are indifferent if $|a_{kj} - a_{lj}| < q_j$.

Definition 2: The strict preference threshold of criterion j p_j: a_k is strictly preferred to a_l if $|a_{kj} - a_{lj}| < q_j$.

Definition 3: The weak preference threshold of criterion j p_j: a_k is weakly preferred to a_l if $|a_{kj} - a_{lj}| \leq p_j$.

Definition 4: The veto threshold of criterion j v_j: reject the hypothesis of outranking of a_k over a_l if $|a_{kj} - a_{lj}| > v_j$.

The Index of Concordance and Discordance

Definition 5: The degree of concordance with the judgmental statement that a_k outranks a_l under the j the criterion $cr_j(k,l)$ is defined as

$$cr_j(k,l) = \begin{cases} 1 & if\ q_j \geq a_{lj} - a_{kj}, \\ 0 & if\ p_j \leq a_{lj} - a_{kj}, \\ \frac{p_j - (a_{lj} - a_{kj})}{p_j - q_j} & otherwise \end{cases}$$

Definition 6: A concordance index of each ordered pair (a_k, a_l) of alternatives $cr(k,l)$ is defined as

$$cr(k,l) = \frac{\sum_{j=1}^{r} w_j cr_j(k,l)}{\sum_{j=1}^{r} w_j}$$

Where w_j is the weight determining the relative importance of j th criterion.

Definition 7: The degree of discordance with the judgmental statement that a_k outranks a_l under the j the criterion $d_j(k,l)$ is defined as

$$d_j(k,l) = \begin{cases} 0 & if\ a_{lj} - a_{kj} \leq p_j, \\ 1 & if\ a_{lj} - a_{kj} \geq v_j, \\ \frac{(a_{lj} - a_{kj}) - p_j}{v_j - p_j} & otherwise \end{cases}$$

The Degree of Outranking

Definition 8: The degree of credibility of outranking with the judgmental statement that a_k outranks a_l is defined as

$$s(k,l) = \begin{cases} cr(k,l) & if\ J(k,l) = \varnothing, \\ cr(k,l) \prod_{j \in J(k,l)} \frac{1 - d_j(k,l)}{1 - cr(k,l)} & otherwise \end{cases}$$

where $J(k,l)$ is defined as the set of criterion for which $d_j(k,l) > cr(k,l)$. If $J(k,l) = \varnothing$, we have $d_j(k,l) > cr(k,l)$ for any criterion, then $s(k,l)$ is the same as $cr(k,l)$.

Definition 9: The ranking of the alternatives is defined as

$$\delta_k = \sum_{l=1}^{n} s(k,l) - \sum_{l=1}^{n} s(l,k), k = 1,2,\ldots,n$$

III. THE ELECTRE III MODEL FOR SELECTING STOCKS

A. *The ELECTRE III Method*

Let A = {a_1, a_2, …a_n} be a set of stock alternatives, $P_{purchase}$ be prices of the purchased stocks in any year, P_{sell} be prices of the sold stocks in any year, and C = {c_1, c_2} be a set of criteria in this research which are net profit margin and dividend yield. W = {w_1, w_2} is a set of weights of influence on criteria net profit margin and dividend yield, a_{kj} is the performance values of criterion c_j of stock alternative a_k, and (a_k, a_j) is any ordered pair of stock alternatives. Net Profits of any stocks are the difference between the price of the purchased stocks and their sold stocks. Total profit is summation of Net Profits and their Dividend yields.

In this section, Combinatorial Portfolio selection with the ELECTRE III method is described. There are three steps as follows:

Table 1: The ELECTRE III method for Selecting Stocks

Input: a list A of stock alternatives, $P_{purchase}$ is prices of the purchased stocks in any year, P_{sell} is prices of the sold stocks in any year, C is a set of criteria: net profit margin and dividend yield, W is a set of weights of influence on criteria net profit margin and dividend yield, Percent of ranking allocation
Output: Total profit of each allocation
1. Ranking the stocks. The results are ranking based on the ELECTRE III method.
2. Allocating percentage of top ranking stocks.
3. Calculating total profit from each allocation.

B. Description of the Scenario

In this section, we present an application of the ELECTRE III method to select any stocks. Suppose that Somsri want to select stocks of any company. Table 2 shows all stocks that the she wants to purchase. Criteria that are considered in selecting each stock are their weights, their preference threshold, their indifference index, and their veto threshold are defined for this application, as in Table 3. The criteria net profit margin and dividend yield are to be maximized. The last price and dividend yield of stocks from 2011 to 2014 are shown in Table 4 and 5 respectively. Finally, the final ranking of the ELECTRE-III methods is shown in Table 6.

Table 2: Stock alternatives

Stock alternatives	Description
A1	BTS: BTS GROUP HOLDINGS PUBLIC COMPANY LIMITED
A2	PTT: PTT PUBLIC COMPANY LIMITED
A3	SPALI: SUPALAI PUBLIC COMPANY LIMITED
A4	SCB: THE SIAM COMMERCIAL BANK PUBLIC COMPANY LIMITED
A5	AHC: AIKCHOL HOSPITAL PUBLIC COMPANY LIMITED

Table 3: Indifference, preference, and veto thresholds values

Criteria	Description	Units	Indifference Threshold (q)	Preference Threshold (p)	Veto Threshold (q)	Weight
C1	Net Profit Margin (%)	%	5	10	20	0.6
C2	Dividend Yield (%)	%	0.5	3	5	0.3

Table 4: Price of stocks of any company in any years

Alternatives	Last Price(Baht)			
	30/12/2011	28/12/2012	27/12/2013	30/12/2014
A1: BTS	0.7	7.15	8.70	9.65
A2: PTT	318	332	286	324
A3: SPALI	14.30	17.70	14.6	24.10
A4: SCB	116.5	181.5	143.5	182.0
A5: AHC	77.50	21.40	19.40	28.50

Table 5: Dividend yields of stocks of the company in any years

Alternatives	Dividend Yield (%)			
	30/12/2011	28/12/2012	27/12/2013	30/12/2014
A1: BTS	5.03	3.55	4.21	6.2
A2: PTT	3.21	3.91	4.55	4.01
A3: SPALI	4.2	3.67	4.45	2.9
A4: SCB	2.57	1.93	3.14	2.88
A5: AHC	3.23	1.64	2.58	2.25

Table 6: Final rankings of the stock alternatives

Stock Alternative	ELECTRE III
A1: BTS	3
A2: PTT	5
A3: SPALI	2
A4: SCB	1
A5: AHC	4

Finally, we can rank stock alternatives. In ELECTRE III method, the best case is A4 and it is followed by A3, A1, A5, and A2 respectively.

IV. EMPIRICAL RESULTS AND ANALYSIS

The empirical experiment to measure the total profit of ELECTRE III algorithm is conducted in this research. Several parameters are needed to be defined when applying the ELECTRE III algorithm:

A. Experiment setting

To demonstrate the effectiveness of the proposed schema the total profits under the two scenarios (that is, the ELECTRE-III method allocation and fixed-percentage allocation) are compared.

Then, assume that a budget is equal to 100,000 Thai Bahts. The number of years that the stock is hold is 4 years. The result of combination of two, three, four highest ranking stocks in the the ELECTRE-III method are shown in Tables 7, 8, and 9 below. Here, a fixed-percentage allocation of 20 % has been assumed for all alternatives and are shown in Table 10.

Table 7: Results of the ELECTRE III method allocation with two highest ranking stocks (ELECTRE III (2))

Stocks	Ranking	Percentage	Budgets	$P_{purchased}$	Units	P_{sold}	Profits	Total
A3: SPALI	2	50	50000	14.3	3496.50	24.1	9.88	87,762.24
A4: SCB	1	50	50000	116.5	429.18	182	65.5	32,626.61
Total								120,388.8

Table 8: Results of the ELECTRE III method allocation with three highest ranking stocks (ELECTRE III (3))

Stocks	Ranking	Percentage	Budgets	$P_{purchased}$	Units	P_{sold}	Profits	Total
A1: BTS	3	33.33	33333	0.7	47619.05	9.65	8.95	1,330,476
A3: SPALI	2	33.33	33333	14.3	2331	24.18	9.88	58,508.16
A4: SCB	1	33.33	333333	116.5	286.12	182	65.5	21,751.07
Total								1,410,735

Table 9: Results of the ELECTRE III method allocation with four highest ranking stocks (ELECTRE III (4))

Stocks	Ranking	Percentage	Budgets	$P_{purchased}$	Units	P_{sold}	Profits	Total
A1: BTS	3	25	25000	0.7	35,714.29	9.65	8.95	997,857.1
A3: SPALI	2	25	25000	14.3	1,784.25	24.18	9.88	43,881.12
A4: SCB	1	25	25000	116.5	214.59	182	65.5	16,313.3
A5: AHC	4	25	25000	77.5	322.58	28.5	-4.9	-12,677.4
Total								1,045,374

Table 10: Results of the fixed-percentage allocation evaluation

Stocks	Ranking	Percentage	Budgets	$P_{purchased}$	Units	P_{sold}	Profits	Total
A1: BTS	3	20	20000	0.7	28571	9.65	8.95	798285.7
A2: PTT	5	20	20000	318	62.89	324	6	1363.52
A3: SPALI	2	20	20000	14.3	1398.6	24.18	9.88	35,104.9
A4: SCB	1	20	20000	116.5	171.67	182	65.5	13,050.64
A5: AHC	4	20	20000	77.5	258.06	28.5	-4.9	-10,141.9
Total								837,662.86

To evaluate the performance of the ELECTRE-III method allocation the above algorithms were run 100 times for various simulation parameters and the average values of the profits were calculated. Table 11 shows various parameters used in the experiment.

Table 11: Simulation parameters used in both scenarios

Parameters	Range of values used for simulation
Fixed percentages	20%
Budgets	100000 baths
Years	2 – 4 years
Alternatives	5

In Table 11, there are four parameters. Fixed percentage of a fixed-percentage allocation scheme is 20%. Budget means the budget for all stocks or alternatives. Years mean the number of years that the stocks are hold from 2011 to 2014. Alternatives mean the number of stocks.

B. Results

The above results (Table 7, 8, and 9) show the total profits under the proposed scheme with different top-n choices. Figure 1 compares these results with a fixed-percentage allocation scheme (Table 10). In Figure 1, the vertical axis represents mean of total profit.

In these simulations, the total profits of the alternative ELECTRE-III (3) is higher than of the fixed-percentage allocation. To determine whether a significant difference exists between the total profits of the two groups, an independent t-test and a Wilcoxon signed rank statistic used with a significance level of $\alpha = 0.05$ were applied to the results of these simulations. The test results reject the hypotheses: the total profit of the ELECTRE-III (3) is equal to of the fixed-percentage allocation. It implies that the total profit of the ELECTRE-III (3) is not equal to that of the fixed-percentage allocation comparisons. Additionally, the ELECTRE-III (4) method gives the same results as the ELECTRE-III (3) method. However, the ELECTRE-III (2) method gives the results opposite from the ELECTRE-III (3) because the number of stocks is not much to guarantee a risk from investment.

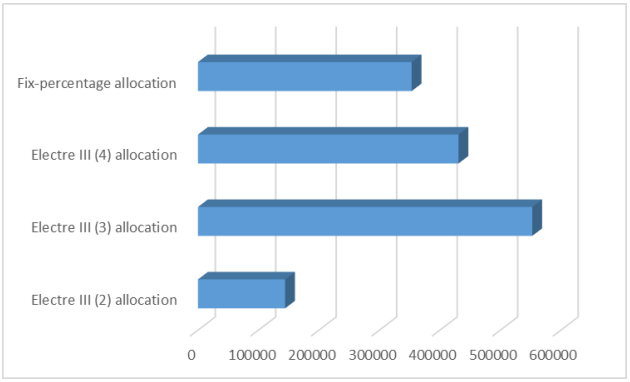


Figure 1: Results of satisfaction evaluation

V. CONCLUSION, LIMITATIONS, AND FUTURE WORKS

Portfolio selection with two significant financial ratios using the ELECTRE III method is proposed for small investors to make trading decisions. We have presented fundamental principles of the ELECTRE III method in detail. In this research, we use a veto indicator deriving from non-veto relations, weak veto relations, and strong veto relations to enhance the mechanism of similarity measure between two cases. 100 times is employed to assess ranking performance of various ELECTRE III methods. From the results of the experiment, we find that the new offer a viable approach for investment advisory ranking. Empirical results show that they offer significantly better ranking performance than the fix-

percentage allocation method. Our proposed prototype using the ELECTRE III method has been successfully validated. This was demonstrated in Figure 1. Such results illustrate that user can get not vague information from application of the ELECTRE III method allocation to help he/she in investment planning and then it could lead to a growing total profits of retail investors in Thailand. Limitations of our study is that it doesn't considered situations with vary alternatives and percentages of ranking stocks.

REFERENCES

[1] Allen F, Karjalainen R. (1999). Using Genetic Algorithms to Find Technical Trading Rules. Journal of Financial Economics 51, 245–271.

[2] Basu S. (1977). Investment performance of common stocks in relation to their Price-Earnings ratios: a test of the efficient market hypothesis, Journal of Finance 32(3), 663-682.

[3] Boonyapatkul P (2011). Investor flows and stock return empirical evidence from stock exchange of Thailand, Master of Science Program in Finance (International Program), Faculty of Commerce and Accountancy Thammasat University, Bangkok, Thailand.

[4] Cao L J, Tay F E H. (2003). Support Vector Machine with Adaptive Parameters in Financial Time Series Forecasting. IEEE Transactions on Neural Networks 14(6), 1506–1518.

[5] Chan L K, Lakonishok J. (2004). Value and growth investing: review and update. Financial Analysts Journal 60(1), 71-86.

[6] Creamer G, Freund Y. (2007). A Boosting Approach for Automated Trading. Journal of Trading 2(3), 84–96.

[7] Creamer G. (2007). Using Boosting for Automated Planning And Trading Systems. Ph.D. Dissertation. Columbia University.

[8] Creamer G. (2012). Model Calibration and Automated Trading Agent for Euro Futures. Quantitative Finance 12(4), 531–545.

[9] Dempster M A H, Payne T W, Romahi Y, Thompson G W T. (2001). Computational Learning Techniques for Intraday FX Trading Using Popular Technical Indicators. IEEE Transactions on Neural Networks 12(4), 744–754.

[10] Fafuła A., Drelczuk K. (2015). Buying stock market winners on Warsaw Stock Exchange - quantitative backtests of a short term trend following strategy. In: Proceedings of Federated Conference Computer Science and Information Systems (FedCSIS), Wrocław, 1361–1366.

[11] Greenblatt J. (2006). The little book that beats the market, John Wiley & Sons, Hoboken.

[12] Kimoto T, Asakawa K, Yoda M, Takeoka M. (2000). Stock Market Prediction System with Modular Neural Networks. Neural Networks in Finance and Investing, 343–357.

[13] Liu J N K, Leung T T S (2001). A Web-based CBR Agent for Financial Forecasting. In: Proceeding of the 4th International Conference on Case-Based Reasoning, 243-253.

[14] Lu C J, Lee T S, Chiu C C. (2009). Financial Time Series Forecasting Using Independent Component Analysis and Support Vector Regression. Decision Support Systems 47(2), 115–125.

[15] Mahfoud S, Mani G. (1996). Financial Forecasting Using Genetic Algorithms. Applied Artificial Intelligence 10, 543–565.

[16] Mandziuk J, Jaruszewicz M. (2011). Neuro-genetic System for Stock Index Prediction. Journal of Intelligent & Fuzzy Systems 22, 93–123.

[17] Moody J, Saffell M. (2001). Learning to Trade via Direct Reinforcement. IEEE Transactions on Neural Networks 12 (4), 875–889.

[18] Moody J, Wu L, Liao Y, Saffell M. (1998). Performance Functions and Reinforcement Learning For Trading Systems and Portfolios. Journal of Forecasting 17, 441–471.

[19] Maneesilasan N. (2011). GARP Investing in Thailand., Working Paper, National Institute of Development Administra-tion, Bangkok.

[20] O J, Lee J W, Zhang B T. (2002). Stock Trading System Using Reinforcement Learning with Cooperative Agents. In Proceedings of the 19th International Conference on Machine Learning, 451–458.

[21] Roy B.: Electre III (1978). Algorithme de classement base sur une representation floue des preferences en presence de criteres multiples, Cahiers du CERO 20(1), 3-24.

[22] Sareewiwatthana P. (2011). Value Investing in Thailand: The Test of Basic Screening Rules, International Review of Business Research Papers 7(4), 1-13.

[23] Tay F E H, Cao L J. (2002). Modified Support Vector Machines in Financial Time Series Forecasting. Neurocomputing 48, 847–861.

[24] Tsang E, Yung P, Li J. (2004). EDDIE-Automation, A Decision Support Tool for Financial Forecasting. Decision Support Systems 37, 559–565. (Periodical style)

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 725–728
DOI: 10.15439/2016F317
ACSIS, Vol. 8. ISSN 2300-5963

Towards solving heterogeneous fleet vehicle routing problem with time windows and additional constraints: real use case study

Krzysztof Bruniecki, Andrzej Chybicki, Marek Moszyński
Gdańsk University of Technology,
Faculty of Electronics,
Telecommunications and Informatics,
ul. Narutowicza 1/2, 80-233 Gdańsk, Poland
Email: Krzysztof.Bruniecki@eti.pg.gda.pl

Mateusz Bonecki, PhD.,
BetterSolutions
(BS Sp. z o. o. Sp. k.),
Gdańsk, Poland
Mateusz.Bonecki@bettersolutions.pl

***Abstract*—In advanced logistic systems, there is a need for a comprehensive optimization of the transport of goods, which would reduce costs. During past decades, several theoretical and practical approaches to solve vehicle routing problems (VRP) were proposed. The problem of optimal fleet management is often transformed to discrete optimization problem that relies on determining the most economical transport routes for a number of vehicles in order to deliver (or pick up) certain amount of goods to geographically distributed set of customers. However, real life problems generally differ from the classical cases because they impose additional constraints to be satisfied. Therefore, research related to developing dedicated theoretical and technological solutions fitted to particular real life use case are very important. In the paper, the particular variant of the VRP problem which is an exemplification of the real-life business case is presented. Authors of the paper proposed an architecture of modular system and an algorithm designed to solve the real-world variant of the VRP which is important from the business perspective.**

I. Introduction

THE fleet management support systems are becoming important tools for various branches of logistics industry. Currently, dedicated software information systems provide tools for optimal resource utilization that in consequence allows for cost reduction related to goods transportation. Therefore, recently significant number of R&D projects related to the problem of optimization of fleet management is being held.

In theoretical terms, the problem of fleet management is often transformed to discrete optimization problem that can be included to the class of so called "NP-hard" problems, which are generally characterized by exponential complexity. Many theoretical approaches to solve vehicle routing problem (VRP) since its first formulation by Dantzig and Ramser 1959 [1] have been studied. Heuristic approaches like Clarke and Wright algorithm [2] based on greedy routes creations or two stages approaches like "route-first cluster-second" (or the opposite) were considered. While such heuristic approaches were mostly proposed for rather simple variants of VRP, recently, local search algorithms with metaheuristic control strategies involved like tabu search [3] or simulated annealing are often considered, especially for solving more complicated variants [4]. Some bio-inspired metaheuristics, including evolutionary approaches like memetic algorithms [5] or ant colony optimization techniques for VRP are also under investigation [6]. Many of the specific variants of VRP are nowadays classical. Some more prominent are Capacitated Vehicle Routing Problem (CVRP) or CVRP with time Windows (CVRPTW). However, real life transportation and logistics problems differ from such classical due to several reasons. The structure and size of real-world input datasets, like fleet heterogeneity, number of customers and many others makes them differ from classical benchmarks and therefore it forces developers to solve business scenarios independently. Also, adding additional constraints to the classical variants of VRP problems caused by business background and clients' requirements is needed. Usually, these conditions are very specific and matched to dedicated implementation. Although many approaches for VRP have been proposed so far, there is still a need to develop dedicated and fitted solutions. Therefore, research and development related VRP that arise from real-world problems is an interesting and important direction [7].

In this paper, the specific variant of VRP is considered. It includes additional constraints mostly related to time windows (TW) and truck/trailer (TT) specifics. Many variants of VRP with time windows were considered in the literature. Relatively simple variants assume only time window and capacity constraints [8]. More complex variants may add additional time dependency of travel time matrix [9] or even introduce classes of customers, e.g., pickup and delivery [10]. The truck and trailer routing problem (TTRP) were proposed and formalized by Chao [11], who also proposed 21 benchmark instances for the problem which are the classical and improved till now [12]. Even a combination of TW and TT were recently considered by Lin et al. [13] as TTRP with Time Windows (TTRPTW). However, even such complicated model is not sufficient for all the real-world demands included in business scenario considered in this article.

In the II section real-world case study was presented. It was performed on the basis of the data retrieved from the existing enterprise resource planning (ERP) system, that is actually used in leading dairy companies from northern Poland. The existing ERP system only partially supports the

This research was sponsored by the National Centre for Research and Development within Applied Research Programme under the project "Multi-criteria routing algorithms for fleet management supporting systems" (contract no. PBS3/B3/37/2015).

vehicle planning process for milk transport industry, so the III section presents the software architecture meant as an lightweight extension of the existing system with dedicated VRP capabilities. The proposed heuristic approach for to the specific variant of the problem and its implementation that uses constraint programming helper library (e.g., google's or-tools library) is presented in the IV section.

II. The problem formulation and analysis

The considered variant of VRP can be considered as a generalization of the classical Truck and Trailer Routing Problem with Time Windows (TTRPTW), therefore it will be called GTTRPTW. Some of the additional restrictions related to GTTRPTW are vehicle heterogeneity and their possible re-use, three types of customers (i.e., not every trailer customer have to provide parking space), maximal vehicle and driver working time and distance as well as additional complexity issues for service time and transshipment time evaluation.

The time interval when the customer is serviced must be within the time window of this customer, and the type of the vehicle to service this particular customer (with or without the trailer) must be acceptable by the customer. Additionally, a list of customers at which it is possible to leave the trailer is given. The vehicle fleet consists of truck units and truck and trailer units. Trailers can be uncoupled en-route at some of the customers where truck sub-tours are built. The truck driving with a trailer can be disconnected, then pick up a load from clients, and then go back to the trailer left and go further.

The vehicle can be assigned to an additional route (or routes) if the previous route is completed. The driver may be exchanged and the vehicle limits renewed. The algorithm should exclude solutions, for which the total time of the designated vehicle route is greater than the maximum total working time of this set or situations where total route distance exceeds specified vehicle maximum route length.

Nine test cases derived from the existing ERP system were converted into GTTRPTW instances. Currently, the system works for the region of northern Poland and allows for manual route planning, that relies on assigning points to route certain vehicles, as well as automatic planning. As specified in Tab. 1, number of clients (nodes in the graph) that goods have to be transported from in 8 of 9 cases exceeds 700, except for test case no. 5. Third row represents total amount of load to be transported in particular problem instance. In this case, it is the number of liters of the milk, since the problem is specific for the dairy transport problems. Some of the node clients allow for entrance of the vehicle with the trailers and, analogically, some of the clients allow for load transshipment at their location. This information is specified in 4th and 5th row of the table respectively. Clients of the system also specify 3 variants of time windows constraints of load pick-up, namely: 24 h windows (no TW restrictions), 6:00 AM -12:00 AM (morning windows) and all-day window (6:00.AM-18:00 PM). Statistical interpretation of the input data characteristics was presented in Tab. 2. In this case, only 30% of clients allow for entrance of the trailer to their localization, and about 25% of clients allow for transshipment. Moreover, if we take under consideration fleet of available vehicles, only 10%-20% of units is equipped with the trailer. The considered GTTRPTW is a combination of classical VRP approach with some particular constraints related to TT and TW. However, TT only applies to 15% of vehicles of the fleet. Moreover, the detail analysis of TW constraints also shows that, for most of the clients (over 80%) no time window is specified. For the rest of the clients, all day window (6AM-18PM) is the most often appearing.

TABLE 1.
SPECIFIED TEST CASES FOR CONSIDERED VRP PROBLEM.

Use case no.	1	2	3	4	5	6	7	8	9
No. of clients (nodes)	736	738	737	737	335	702	783	743	742
Total load ammount	357438	576772	456307	477903	170775	579107	622052	568371	632789
The number of clients allowing for entrance of a vehicle with a trailer	218	234	240	212	98	212	237	218	231
No. of transhipment points	158	167	183	142	76	152	175	165	162
Total no. of vehicles	19	19	19	19	10	22	22	22	22
No. of vehicles with trailers	3	3	3	3	9	2	2	2	2
No time window (0:00-23:59)	596	585	596	585	266	702	783	743	742
Morning window 6:00 AM-12:00PM	16	23	22	17	11	0	0	0	0
All-day window 6:00 AM: -18:00 PM	124	130	119	135	58	0	0	0	0

TABLE 2.
STATISTICAL OVERVIEW OF SPECIFIED TEST CASES FOR CONSIDERED VARIANT OF VRP PROBLEM.

Use case no.	1	2	3	4	5	6	7	8	9
The number of customers allowing a vehicle with a trailer [%]	30%	32%	33%	29%	29%	30%	30%	29%	31%
No. of transhipment points [%]	21%	23%	25%	19%	23%	22%	22%	22%	22%
No. of vehicles with trailers [%]	16%	16%	16%	16%	90%	9%	9%	9%	9%
No time window (0:00-23:59) [%]	81%	79%	81%	79,4%	79,4%	100%	100%	100%	100%
Morning window 6:00 AM-12:00PM [%]	2%	3%	3%	2,3%	3,3%	0%	0%	0%	0%
All-day window 6:00 AM: -18:00 PM [%]	17%	18%	16%	18,3%	17,3%	0%	0%	0%	0%

III. SYSTEM ARCHITECTURE

The aim of the proposed system is to solve or validate different variants of VRP problem instances. The system may be used to find a new solution, as well as it allows to improve any previous solution when provided.

The architecture of the system, presented in Fig. 1, is modular in order to adjust to new formats and variants. The front-end layer module is handling different file formats and VRP variants for solutions and problem instances (e.g., handles proprietary JSON-based format dedicated to describe the GTTRPTW). There are also front-end modules dedicated to classical benchmarks' formats like CVRP, CVRPTW, TTRP and TTRPTW.

Data abstraction layer handles transparently different formats. It's common object-model is used by modules in the solvers' layer. The system uses different VRP solvers and the choose is up to the user. One of the solvers implemented is for GTTRPTW and is described in IV section.

The source code of the system and its modules is implemented in C/C++ language, in order to achieve high efficiency with a reasonable portability. The solvers are implemented separately and may differ from each other.

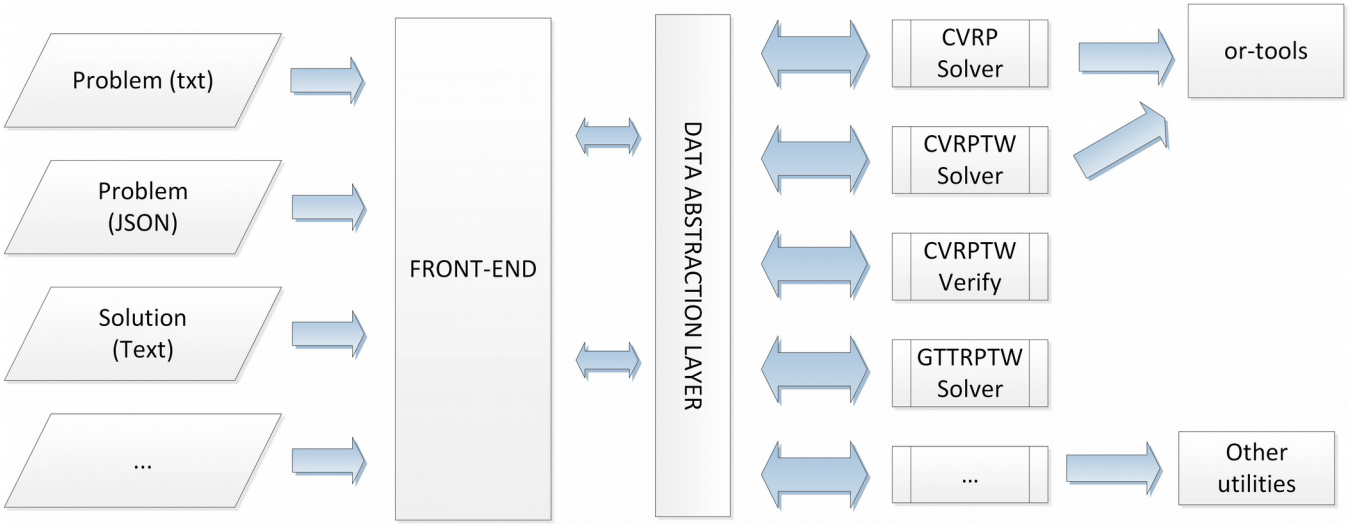


Fig. 1. The architecture of the proposed system

IV. PROPOSED APPROACH

Already implemented solvers including GTTRPTW use constraint programming approach and third party libraries (e.g., or-tools constraint programming library). The proposed approach is based on the observation related to features of input considered test cases. Time windows in the real-world cases is not as restricting. We can distinguish between the time window related to the customer and the time window of the vehicle unit. Typical vehicle unit is often used two times during a single day (different drivers). It means that a single unit will be available within two separate time windows. Most of the customers (as can be seen in the Tab. 1 and Tab. 2) do not specify constraint on the time windows. Although, there is a significant set of customers which specify the time window, most often the time window corresponds to the single shift of the drivers. Therefore we may observe, that if the customer is served on a route by a specific truck and driver it can be often reordered within the route without violating the time window constraint.

Therefore the outline of the strategy used in the GTTRPTW solver is as follows:

1. Find the solution *S* of the relaxed variant (i.e., CVRPTW) of the problem;
2. For each route *r* in the solution *S*;

if the route have truck-only customers and the vehicle is truck and trailer; we need to reorder the route and chose transhipment points;

- 2.1 For each transhipment customer *t* in the route *r* compute the route r_t;
  - 2.1.1 Assume *t* is the transhipment and compute the new route *r′* for all the customers from the original route *r*;
  - 2.1.2. Start the *r′* from depot to *t* and then traverse all customers with only the truck, the number of cycles needed may vary between 1 to 3. Finally, go back from *t* to the depot.
    - 2.1.2.1. The optimal cycles from *t* are computed with use of homogeneous CVRPTW solver with *t* considered to be the depot and the capacity of the fleet equal the capacity of the truck;
    - 2.1.2.2. Optimize the route *r′* greedy by moving some truck and trailer customers from the truck cycle to the segment which represent the starting or ending part of the new route *r′* (the part between original depot and the assumed transhipment *t*).
  - 2.1.3. Remember the route if it has smallest cost among all the previous assumptions for transhipment *t*;
- 2.2. Remember the best r_t as the new route in the solution S';

3. Optimize the solution greedy by moving and exchanging some customers between routes. Must not violate other constraints.

- 3.1. Find a feasible move giving biggest savings,
  - 3.1.1 If there is one go to 3.3;
- 3.2. Find a feasible exchange giving biggest savings
  - 3.2.1. If there is no feasible exchange available go to 4.
- 3.3. Apply the modification found
  - 3.3.1. If the time exceeds the available then go to 4.

makes it its generalization - GTTRPTW. The GTTRPTW was identified during demands analysis of real-life business scenarios from dairy mil companies. Authors proposed an architecture of a lightweight extension to existing ERP system designed to solve many variants of VRP including the one proposed in this paper. Dedicated modules responsible for solving simpler variants (e.g., CVRP, CVRPTW), which are also useful in simpler business scenarios, were already implemented into the system [14]. Beside the formulation, authors presented an algorithm that solves GTTRPTW. The algorithm in its first stage uses constraint programming approach, implemented with use of or-tools library, for solving relaxed variant of problem (e.g., without TT restriction). In the second stage the algorithm continue computation with TT and additional and performs

some local search improvements. Further research and development is planned to finally compare the results obtained by the proposed algorithm with other state of the art approaches. Since the problem variant may be considered as an extension of TTRPTW, the relaxation of the proposed algorithm and its comparison to best known results, for 36 benchmarks from [13], would be valuable. An adaptation of local search strategies like Attribute Based Hill Climbing (ABHC) into more complicated variant of TTRPTW, is also under consideration, especially that ABHC have revealed promising results [15] for classical TTRPTW.

REFERENCES

[1] G. B. Dantzig and J. H. Ramser "The truck dispatching problem," *Management Science*, vol. 6, no. 1, pp. 80–91, 1959.

[2] G. Clarke and J. Wright "Scheduling of vehicles from a central depot to a number of delivery points", *Operations Research*, vol. 12, no. 4, pp. 568-581, 1964.

[3] É. D. Taillard, P. Badeau, M. Gendreau, F. Guertin and J.-Y. Potvin, "A tabu search heuristic for the vehicle routing problem with soft time windows", *Transportation Science,* vol. 31, pp. 170-186, 1997.

[4] S.C. Ho and D. Haugland, "A tabu search heuristic for the vehicle routing problem with time windows and split deliveries", *Computers & Operations Research*, vol. 31, no. 12, pp. 1947-1964, 2004.

[5] E. Osaba and F. Díaz, "Comparison of a memetic algorithm and a tabu search algorithm for the Traveling Salesman Problem," In: FedCSIS. 2012, pp. 131-136.

[6] L.M. Gambardella, E. Taillard and G. Agazzi, "MACS-VRPTW: A Multiple Ant Colony System for Vehicle Routing Problems with Time Windows" , In D. Corne, M. Dorigo and F. Glover, editors, New Ideas in Optimization McGraw-Hill, London, UK, pp. 63-76, 1999.

[7] M. Batsyn and A. Ponomarenko, "Heuristic for a Real-life Truck and Trailer Routing Problem", Procedia Computer Science vol. 31, 2014, Pages 778–792, 2nd International Conference on Information Technology and Quantitative Management, ITQM 2014.

[8] H. Akeb, A. Bouchakhchoukha, M. Hifi, "A beam search based algorithm for the capacitated vehicle routing problem with time windows". In: Computer Science and Information Systems (FedCSIS), 2013 Federated Conference on. IEEE, 2013. pp. 329-336.

[9] J. Hurkała, "Time-Dependent Traveling Salesman Problem with Multiple Time Windows" *Annals of Computer Science and Information Systems*, vol. 6, pp. 71-78, 2015.

[10] R. Bent and P.V. Hentenryck, "A two-stage hybrid algorithm for pickup and delivery vehicle routing problems with time windows," *Computers & Operations Research*, vol.33, no.4, pp.875-893, 2003.

[11] I. M. Chao, "A tabu search method for the truck and trailer routing problem," *Computers & Operations Research*, vol. 29(1), pp. 33-51, 2002.

[12] J. G. Villegasa, C. Prinsb, C. Prodhonb, A. L. Medagliac and N. Velascod, "A matheuristic for the truck and trailer routing problem", *European Journal of Operational Research*, vol. 230, no. 2, pp. 231–244, 2013.

[13] S.-W. Lin, V. F. Yu and C.-C. Lu, "A simulated annealing heuristic for the truck and trailer routing problem with time windows," *Expert Systems with Applications*, vol. 38(12), pp. 15244-15252. 2011.

[14] K. Bruniecki, A. Chybicki, M. Moszyński and M. Bonecki, "Evaluation of vehicle routing problem algorithms for transport logistics using dedicated GIS system," 2016 Baltic Geodetic Congress (BGC Geomatics), Gdansk, pp. 116-121, 2016.

[15] U. Derigs, M. Pullmann and U. Vogel, "Truck and trailer routing-Problems, heuristics and computational experience," *Computers & Operations Research*, vol. 40, no. 2, pp. 536-546, 2013.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 729–732
DOI: 10.15439/2016F323
ACSIS, Vol. 8. ISSN 2300-5963

Risk-based estimation of manufacturing order costs with artificial intelligence

Grzegorz Kłosowski
Lublin University of Technology, Faculty of Management, Department of Enterprise Organization, Lublin, Poland
Email: g.klosowski@pollub.pl

Arkadiusz Gola
Lublin University of Technology, Faculty of Mechanical Engineering, Institute of Technological Systems of Information, Poland
Email: a.gola@pollub.pl

Abstract—The following paper discusses the development of a risk-based cost estimation model for completing non-standard manufacturing orders. The model in question is a hybrid of Monte Carlo Simulation (MCS), which constitutes the main module of the applied model. Vector of order risk probability, which is the input data for the MCS module, is highly difficult to assess and is burdened to a considerable degree with subjectivity, therefore it was resolved that it should be generated with the application of artificial intelligence. Depending on the accessibility of historical data, the model incorporates fuzzy logic or artificial neural networks methods. The presented model could provide support to managers responsible for cost estimation, and moreover, after slight modification also in setting deadlines for non-standard manufacturing orders.

I. Introduction

Analysis of trends in the development and operation of modern manufacturing enterprises indicates that competitiveness of an enterprise can by improved mainly through innovation in the fields of: manufactured products, technology, management and marketing [1]. Introducing new products into the offer [2], improving manufacturing and support processes [3] is becoming increasingly hard owing to rising costs and strong market competition, particularly from big enterprises. An outstanding advantage of small and medium-sized enterprises (SME) consists in their flexibility, which enables them to compete successfully for non-standard and low-batch orders [4]. Realisation of such orders usually requires accepting a project approach, which demands scheduling times, resources and costs with every order [5]. Correct cost and order delivery estimation is imperative, as the contract with the customer must specify such arrangements as order delivery date and cost.

Non-standard manufacturing orders realised by medium-sized enterprises include, for instance:

- special tool orders for big enterprises, used in large-batch production,
- matrices and punches,
- custom machines and appliances (e.g. paper tube and core making machines, CNC nesting machines, storage systems etc).

The literature describes a range of various attempts at classification of risk, which proves it to be complex a and multidimensional phenomenon [6]. Risk modelling is a developing and ongoing process [7], which causes that risk is frequently among the major factors behind miscalculation of non-standard order costs [8]. There is a crucial necessity for cost estimation method that would cover all estimation factors. There are many proposals that lack scientific justification for the produced results, *i.e.* lack technical description of how the results were achieved [9].

In the case of non-standard orders, the risk of untimely delivery or exceeding the budget increases. Efficient cost estimation for diverse manufacturing orders becomes more complicated with the increasing number of factors that remain beyond the control of the manufacturer. These factors include uncertainty of any cooperative tasks, changing currency exchange rates, strict design or material requirements or industrial accidents [10]. Missing deadlines often results in enforcing contractual fines, losing customers and tarnishing the company's reputation, the last one being one of the key assets of any company and a critical success factor

Given the circumstances, it becomes a matter of high importance to develop an efficient method for risk minimisation concerning time and cost estimation of non-standard manufacturing orders [11].

II. Concept

There are three major indicators for the effectiveness of risk-based time-cost estimation of manufacturing orders: accuracy, time effectiveness and applicability. Clearly, solving these problems would not be possible without the application of IT solutions, which can meet the aforementioned criteria. Therefore, enterprises apply expert systems oriented towards aiding the decision-making process in the field of order cost estimation. The market offers numerous software solutions integrating business and production processes, nevertheless, the analysed case is more elaborate and therefore calls for a more versatile tool. Here, the input quality data must be correlated with results of quantitative character, *i.e.* time and cost.

The body of literature points at several methods for estimation of project risk. The most commonly used is the Monte Carlo Simulation (MCS), along with such methods as Artificial Neural Networks (ANN), Fuzzy Logic (FL), and Support Vector Machines (SVM) [12].

What contributes to the popularity of MCS is its uncomplicated applicability. MCS is a quantitative method that consists in building probability distribution for any risk involved. The consequences of unforeseen incidents could lead to unplanned change of costs or order delivery date, which in small-batch/non-standard orders as described here, is understood as a project.

A marked disadvantage of the method is the need for deterministic estimation of particular project risks probability. It is usually carried out by a single expert or a group of experts representing vast knowledge in a given discipline or branch. Eventually, this is still a subjective evaluation of probability, which is exactly one of the key drawbacks of MCS.

One of the most popular ANN methods is multilayer perceptron (MLP). In this variation of ANN it is required that a sufficiently large quantity of suitable historical data is available, which could be then used as input in train, test and validation sets. Another problem is to find cause-and-effect relationship between suitably selected input data sets and order delivery date.

SVM resembles to a certain extent ANN/MLP, inasmuch as it requires a set of historical data to conduct the training process. One advantage over ANN is that SVM training always finds a global minimum and that it mitigates the risk of overtraining, however, the estimation with this method is quite time inefficient.

Fuzzy logic (FL) can be put to use in time-cost risk assessment, particularly when no historical data is available [13]-[14]. The method employs several heuristics, which are developed with *e.g.* Delphi method [15] or Brainstorming. Heuristics are recorded as reasoning rules, which in the next stage provide the core for the fuzzy inference system [16]. Apart from the reasoning rules, this method requires selection of suitable inputs, membership functions and defuzzification.

The short descriptions of each method indicate clearly that each model is burdened with certain limitations of different magnitude, which makes it difficult to apply in estimation of costs and time of individual orders, which can be treated as separate projects.

In order to eliminate the negative impact of subjective estimation, present in the classical MCS, AI can be applied in probability estimation of particular project risks.

Enterprises that do not possess historical data that could be used for training neural network or SVM controller could employ the FL method.

The proposed hypothesis for the application of a hybrid system, incorporating AI for percentage estimation of project risks in Monte Carlo method, will increase the accuracy of the MCS.

The second hypothesis states that determining the probability of project risks with the application of artificial intelligence is more reliable than the deterministic method, based on the subjective opinion of experts.

III. PROPOSED METHODOLOGY

Table 1 shows an exemplary project risk calculation of non-standard manufacturing orders carried out with Monte Carlo simulation. Column 1 contains Risk Breakdown Structure. Column 2 specifies identified risks. Column 3 shows subjectively estimated particular risk probability. Column 4 contains costs of risk that would have to be covered if the risk occurred. An analogical approach could be accepted in determining the risk of untimely delivery dates, by substituting cost with time. In such a case, the set of risks in column 2 would be different as well.

Column 5 shows an expected number of risks, which is a product of columns 4 and 5. The sum of column 5 amounts to 71.50 EUR, which would not cover the potential expenses should the risks R-1, R-2 or R-6 take place. Columns 6 and 7 enable simulation of numerous risk-related variants. Column 6 shows a function randomly generating absolute numbers from the range <0,1). Column 7 contains logical conditional formulas:

$$\text{IF (col.3) >= (col.6) THEN (col.4)} \qquad (1)$$

The sum of column 7 contains the cost of risk in the simulated case. After 1000 simulations with the random number generator and formula (1) cumulative distribution function was obtained, whose part is shown in Fig. 1.

The x-axis shows costs of risk for particular scenarios, the y-axis the number of scenarios population, counted as a percentage share of probable situations. The best-case scenario estimates that the project risk amounts to 50 EUR.

Planning the budget for the realisation of order must account for two opposite goals: cost and risk minimisation. Increasing the budget by 600 EUR for the minimisation of project risk would result in a practically 100% guarantee that the project expenses will not exceed the budget. Nevertheless, excessive costs might not be covered by the customer, which is why, compromise solutions must be sought. It can be assumed that the acceptable compromise is setting the cost risk level at approx. 80%. Broken lines in Fig. 1 show the 84% level of probability, corresponding to 160 EUR of additional risk-related cost it can be observed that the majority of all analyzed scenarios are on the left of that amount.

To remove the element of subjectivity, artificial intelligence can be applied in probability selection of particular risks (Table 1, column 3)

The algorithm for the development of a hybrid system for estimation of project risks is shown in Fig. 2.

TABLE I.
QUANTITATIVE ANALYSIS OF PROJECT RISKS MONTE CARLO SIMULATION

RBS	**Risk description**	**Probability**	**Cost**	**Level**	**RAND()**	**Simulation**
1	*2*	*3*	*4*	*5*	*6*	*7*
R-1	Subcontractors errors	10%	€ 150.00	€ 15.00	60.6 %	
R-2	Currency exchange rate changes	5%	€ 300.00	€ 15.00	1.0 %	€ **300.00**
R-3	Lack of resources	10%	€ 50.00	€ 5.00	85.7 %	
R-4	Order requirements problems	50%	€ 10.00	€ 5.00	22.3 %	€ **10.00**
R-5	Supplier's delay [14]	20%	€ 20.00	€ 4.00	81.3 %	
R-6	Computer network failure	30%	€ 100.00	€ 30.00	96.0 %	
O-1	Supplier discount	5%	-€ 50.00	-€ 2.50	3.1 %	-€ **50.00**
Sum:				**€ 71.50**		€ **260.00**

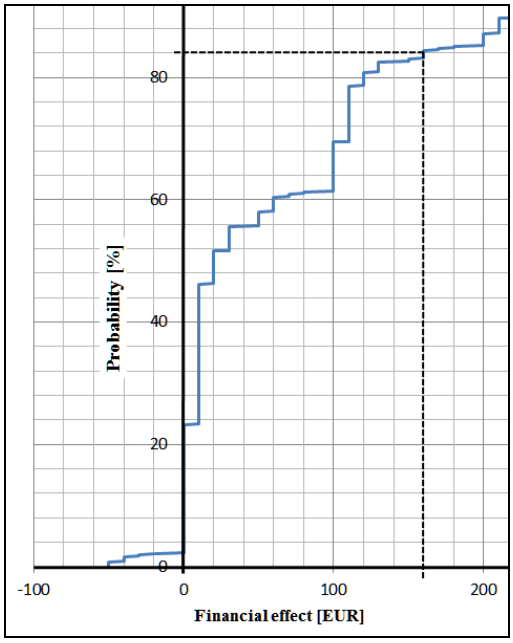


Fig. 1. Probability cumulative distribution function for project risk costs

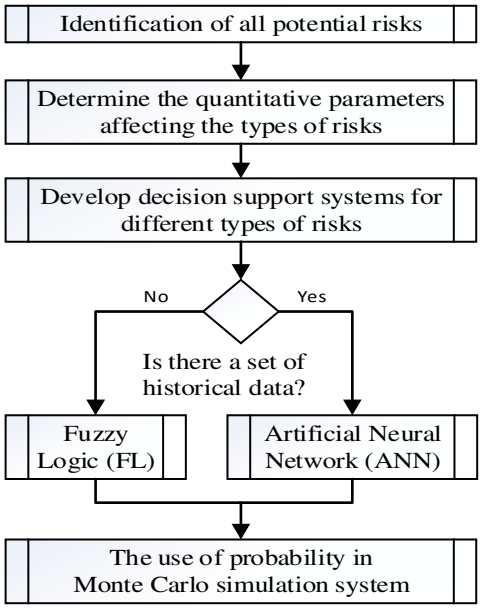


Fig. 2. Algorithm for the development of project risk estimation system for non-standard orders

The first stage of development consists in identification of all potential risks that could affect costs or delivery dates. Next step is to describe the identified risks in quantitative measures. These will provide input data for particular decision-making modules describing the probability of particular risks.

If a system can be fed with tabulated historical data, a decision-making subsystem based on ANN could be created. Otherwise, the application of FL is possible.

SVM was skipped in the present analysis due to extensive computational time, which means it does not meet one of the criterions - time efficiency, and as a result the criterion of applicability.

It ought to be noted that in selection of input vectors for particular modules for determining risk probability it is the historical data accessibility that plays a crucial role. It can be, for instance, assumed that risk R-1 (Subcontractor errors) is influenced by criteria shown in Table 2, corresponding to the intelligent subsystem for project risk estimation shown in the diagram in Fig. 3.

Column 3 Table 2 contains methods for R-1 module input features determination. While to determine feature I-1 is relatively simple, the situation becomes complicated in the case of I-2 and I-5 inputs.

TABLE II. INPUT FEATURES FOR MODULE R-1 (SUBCONTRACTORS ERRORS)

IBS	Input feature name	Measure
1	*2*	*3*
I-1	The number of tasks within the manufacturing order that require hiring subcontractors	[pcs]
I-2	The lowest rating of cooperation history for subcontractor among external service providers cooperating in realisation of order	[%]
I-5	The lowest score among external audits *in situ* at subcontractors'	[1..5]

For instance, determination of a percentage value of feature I-2 requires a prior analysis of timeliness of all subcontractors. This can be calculated from the relationship (2).

$$M_{1-2} = \left(1 - \frac{c}{s_o}\right) \times 100\% \qquad (2)$$

where:
M_{1-2} – feature value I-2
c – total number of claims from given subcontractor
S_o – number of all orders in the past from given subcontractor

Input I-5 can take the value from 1 to 5, where 1 denotes low assessment of quality model. Input feature I-5 requires obtaining the results of audits that would cover data from quality assurance systems of each of the subcontractor. Although I-5 feature values are specified by experts, they can be considered as fully reliable. This is because they were obtained in an analysis of appropriate parameters defined as a part of subcontractor's internal quality management system. If this is a certified system, such as ISO 9001, then evaluation based on the parameters is simple. Otherwise, the evaluation requires defining reliable methods of measurement.

Values of individual project risks probability based on the outputs of intelligent subsystems constitute input data for the system of risk cost calculation, based on Monte Carlo simulation.

Fig. 3. Shows the operation of an intelligent subsystem for project risk estimation. On the left side there is an N-element vector of identified input features, which could have an impact on *n-element set of project risks*; hence, $N \geq n$.

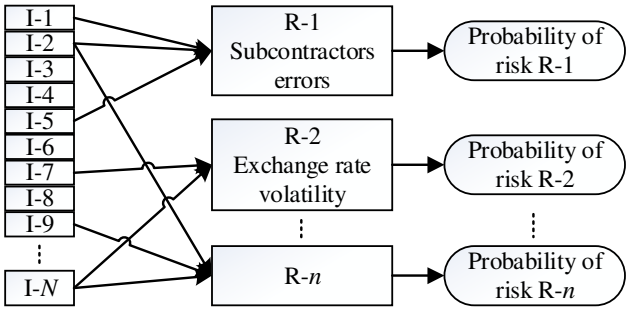


Fig.3. Diagram of an intelligent subsystem for estimation of project risk

Fig. 3 indicates that it is possible for a single output (*e.g.* I-2) to feed two or more module for R-I risk probability evaluation.

Fig. 4 shows a diagram of a complete hybrid system for the calculation of manufacturing order risk. It can be seen that the system is composed of two major subsystems: Artificial Intelligence (AI) and Monte Carlo Simulation (MCS).

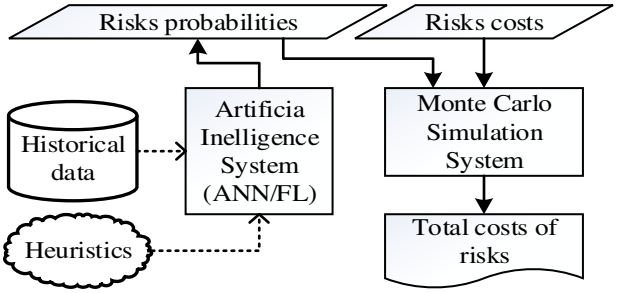


Fig. 4. Diagram of a hybrid system for the calculation of manufacturing order risk

The functioning of AI module is shown in Fig. 3. The process of risk estimation with MCS is shown in Table 1 and Fig. 1. Fig. 4 indicates that the risk probabilities vector, which constitutes the output of AI subsystem, is simultaneously the input of MCS subsystem.

IV. CONCLUSION

This paper presents solutions for the problem of risk calculation in non-standard manufacturing orders in small and medium-sized enterprises. Highly individual character of particular orders permits treating them as separate projects. The two initially formulated, mutually complementing hypotheses stated that employing hybrid systems based on artificial intelligence produces more accurate results of project risk calculations. The truthfulness of the hypotheses was initially confirmed by introduction of logical and coherent vision of reasoning rules, which could replace the subjective, hence imperfect decisions taken by human.

The algorithm for development of a system of risk estimation for non-standard manufacturing orders was created. We proposed an improved Monte Carlo simulation method with an additional subsystem variant, employing fuzzy logic or neural networks, depending on the availability of historical data.

It should be noted that by substituting cost with time, the presented solution could be easily adapted for risk estimation of order delivery delay in single orders.

V. REFERENCES

[1] M. Jasiulewicz-Kaczmarek , “Participatory Ergonomics as a Method of Quality Improvement in Maintenance”, B.-T. Karsh (Ed.): *Ergonomics and Health Aspects*, LNCS 5624, pp. 153–161, 2009.

[2] A. M. Radke, et al. “A risk management-based evaluation of inventory allocations for make-to-order production,” *CIRP Annals-Manufacturing Technology*, pp. 459-462, 2013.

[3] A. Saniuk, et al. “Environmental favourable foundries through maintenance activities”, *METALURGIJA,* Vol. 54 (4), pp. 725-728, 2015.

[4] G. Kłosowski, A. Gola, A. Świć, "Application of Fuzzy Logic Controller for Machine Load Balancing in Discrete Manufacturing System", in: K. Jackowski et al. (Eds): IDEAL 2015, LNCS 9375, pp. 256-264, 2015.

[5] P. Sitek, J. Wikarek, "A hybrid framework for the modelling and optimisation of decision problems in sustainable supply chain management", *International Journal of Production Research*, Vol. 53, Issue 21, 2015.

[6] A. Rudawska, N. Čuboňova, K. Pomarańska, D. Stanečková, A. Gola, "Technical and Organizational Improvements of Packaging Production Processes", *Advances in Science and Technology. Research Journal*, Vol. 10, No. 30, pp. 182-192, 2016.

[7] A. Taroun, J. B. Yang, D. Lowe, “Construction risk modelling and assessment: Insights from a literature review”, *The Built and Human Environment Review*, Vol. 32, Issue 1, pp. 101-115, 2011.

[8] C. Rush, R. Roy, “Analysis of cost estimating processes used within a concurrent engineering environment throughout a product life cycle,” In: *7th ISPE International Conference on Concurrent Engineering*, Lyon, France, July 17th-20th, Pennsylvania USA. pp. 58-67, 2000.

[9] A. O. Elfaki, S. Alatawi, E. Abushandi, “Using Intelligent Techniques in Construction Project Cost Estimation: 10-Year Survey,” *Advances in Civil Engineering*, Vol. 2014, pp. 8, 2014.

[10] N. G. Fragiadakis, V. D. Tsoukalas, V. J. Papazoglou, “An adaptive neuro-fuzzy inference system (ANFIS) model for assessing occupational risk in the shipbuilding industry,” *Safety Science*, Vol. 63, pp. 226–235, 2014.

[11] O. Taylan et al., “Construction projects selection and risk assessment by fuzzy AHP and fuzzy TOPSIS methodologies,” *Applied Soft Computing*, Vol. 17, pp. 105-116, 2014.

[12] I. J. Schwarz, I. P. M. Sánchez, “29 Implementation of artificial intelligence into risk management decision-making processes in construction projects”, 2015, pp. 361-362.

[13] E. E. Ameyaw, A. P. C. Chan, “Evaluation and ranking of risk factors in public–private partnership water supply projects in developing countries using fuzzy synthetic evaluation approach,” *Expert Systems with Applications*, Vol. 42(12), pp. 5102-5116, 2015.

[14] A. Idrus, M. F. Nuruddin, M. A. Rohman, “Development of project cost contingency estimation model using risk analysis and fuzzy expert system,” *Expert Systems with Applications*, Vol. 38, Issue 3, pp. 1501-1508, 2011.

[15] J. Hu, E. Shen, Y. Gu, “Evaluation of Lighting Performance Risk Using Surrogate Model and EnergyPlus,” *Procedia Engineering*, Vol. 118, pp. 522-529, 2015.

[16] P. D. Sentia, M. Mukhtar, S.A. Shukor, “Supply chain information risk management model in Make-to-Order (MTO)”, Procedia Technology, Vol. 11, pp. 403-410, 2013.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 733–740
DOI: 10.15439/2016F38
ACSIS, Vol. 8. ISSN 2300-5963

A declarative decision support framework for scheduling groups of orders

Jarosław Wikarek
Kielce University of Technology Al. 1000-lecia PP 7, 25-314 Kielce, Poland, Institute of Management and Control Systems
e-mail:j.wikarek@tu.kielce.pl

Krzysztof Bzdyra
Koszalin University of Technology, Department of Computer Science and Management, Poland e-mail:
krzysztof.bzdyra@tu.koszalin.pl

***Abstract*—This paper deals with declarative decision support framework for scheduling groups of orders. All orders in a group should be delivered at the same time after processing. The authors present a novel declarative approach to modeling and solving scheduling problems as a declarative decision support framework. The proposed framework makes it possible to ask different types of questions (general, specific, logical, etc.). It also allows, scheduling emerging orders or groups of orders without changing the existing schedules. To implement was used CLP (Constraint Logic Programming) environment. To increase the efficiency of the framework, particularly in the area of optimization made its integration with MP (Mathematical Programming) environment. The paper also presents the implementation of illustrative model, using the proposed framework. In addition, an efficiency analysis of the presented solution in relation to the application of mathematical programming have been conducted.**

I. Introduction

THE proposed research problem (scheduling groups of orders) finds many applications in industrial and services companies, including but not limited to food, textile production industries, distributions, ceramic tile, supply chain, manufacturing of complex devices, fast-foods, restaurants, postal and courier services etc. Assume that each customer has different orders. Each order has a different process function and set of resources, but all items ordered by a customer or group of customers should be delivered at the same time in one package to reduce the transportation costs, subsequent processing steps time and costs, or/and assure proper quality of the product/service and customer satisfaction. In this type scheduling problems, in addition to standard constraints such as precedence or disjunction, new constraints appear related to the given group, concurrent delivery date, etc. In practice also logical constraints may occur, resulting from business, marketing or legal conditions. Therefore the modeling and solving of various constraints in scheduling groups of orders is a key issue. Managers/Decision-makers need to have schedules with defined parameters and/or the knowledge whether the schedule meets the requirements, which may be formulated as simple questions. Good environments for the modeling of constraints, questions and logical conditions include declarative environments, CLP (Constraint Logic Programming) in particular.

Our motivation was to develop an environment for the modeling and decision support of the problem for scheduling groups of orders. The use of this framework would help obtain quick answers to key questions (Is it possible…?, What If…?, What is the minimum/maximum..?) asked by managers/decision makers.

This paper proposes the concept of a declarative decision support framework for scheduling groups of orders and presents its implementation in the CLP environment. The illustrative example shows the potential of the framework.

The remainder of the article is organized as follows. Section 2 presents problem statement, research methodology, contribution etc. The concept and implementation aspects of a declarative decision support framework are provided in Section 3. Computational examples, tests of the implementation framework and discussion are presented in Section 4. Possible extensions of the proposed approach as well as the conclusions are included in Section 5.

II. Problem statement and methodology

Scheduling methods for optimal and simultaneous service to groups of orders are proposed most often in the flexible flow-shop system (FFS). In the FFS system, processing is divided into several stages with parallel resources/machines at least in one stage. All of the orders should pass through all stages in the same order [1,2]. The objectives of the problem [2] are minimizing the total amount of time required to complete a group of orders and minimizing the sum of differences between the completion time of a particular order in the group and the delivery time of this group containing that order (waiting period). In practical applications, flexible flow-shop system is insufficient since the sequence of operations/tasks in the orders from different users is rarely the same.

The majority of models for scheduling of group orders presented in the literature refer to a single problem and optimization according to single criterion. Fewer studies are devoted to multiple-criteria [2]. Most of them are modeled and solved by operations research (OR) methods. Declarative environments such as CLP facilitate problem

modeling and introduction of logical and symbolic constraints [3,4,5]. Unfortunately, high complexity and the multiple types of constraints of decision-making models as well as combinatorial nature contribute to poor efficiency of modeling in OR methods and inefficient optimization in CLP. Therefore, a new approach to modeling and solving such problems was developed [6,7,8]. A declarative environment was chosen as the best structure for this approach especially in modeling [3,5,9,10]. Mathematical programming environment was used for problem optimization [11]. This integrated approach is the basis for the creation of the implementation environment to support managers. In addition to optimizing particular decision making problems connected with groups of orders, such environment allows asking various questions while processing the orders.

A. Problem description –illustrative example

This problem can be stated as follows. Orders Z_i for different group of product p enter the system in groups at different periods v. Each order consists of a set of operations and should be processed with specific set of parallel resources. It is assumed that there are no gaps between the operations of the order. The orders in each group Z_i should be delivered at the same time. Special points a at which orders are submitted and then delivered are introduced. The problem does not cover the configuration of the points but relates to handling orders, as many orders may come from one customer. Each order may be processed by a different resource set in any order.

In this case, decision support is to respond to the questions asked, which in general can be: specific questions, general questions, logic questions etc.

Possible questions (Q) for such problem are (including but not limited to):

- What is the minimum makespan for groups of orders $Z_1,..,Z_n$ entering in period $v_1,..,v_n$ at the point $a_1,..,a_n$? (Q1)
- Is it possible to execute the new group of orders $Z_{n+1},..,Z_m$ from the period v_k with existing resources at specified period T? (No change orders that are in progress.) (Q2)
- Is it possible to execute group of orders $Z_1,..,Z_n$ in time T with use of the resource $k=N$? (Q3)
- What is the minimum use of resource k to execute orders $Z_1,..,Z_n$ in time T? (Q4)
- Is it possible to execute the new group of orders $Z_{n+1},..,Z_m$ from the period v_k with existing resources at specified period T and the use of resource $k=N$? (Q5)
- What is the minimum makespan for groups of orders/tasks $Z_1,..,Z_n$, entering in period $v_1,..,v_n$ at the point $a_1,..,a_n$? (with all the resources k reduced by C%) (Q6)
- Is it possible to execute the groups of orders $Z_1,..,Z_n$ in time T entering in period $v_1,..,v_n$ at the point $a_1,..,a_n$? with exclusively use resources k_i and k_j? (Q7)

Decision variables of this problem are shown in Table I.

TABLE I.
DECISION VARIABLES

Decision variables	
Calculated number of periods g delivery of all orders for point a.	Tkp_a
If at a given point a ordered product p then $Xzk_{a,p}=1$, otherwise $Xzk_{a,p}=0$	$Xzk_{a,p}$
Number of period g in which operation o can be started for product p ordered at point a	$B_{a,p,o}$
If the execution of operation o for product p ordered at point a uses resource k in period g then $X_{a,p,o,k,g}=1$, otherwise $X_{a,p,o,k,g}=0$	$X_{a,p,o,k,g}$
If the execution of operation o for product p ordered at point a uses resource k in period g then $Xo_{a,p,o,k,g}=zk_{a,p}$, otherwise $Xo_{a,p,o,k,g}=0$	$Xo_{a,p,o,k,g}$
If g is the last period in which resource k is used in the execution of operation o for product p at point a then $Y_{a,p,o,k,g}=1$, otherwise $Y_{a,p,o,k,g}$=0	$Y_{a,p,o,k,g}$
If g is the last period in which orders are executed for point a then $W_{a,g}=1$, otherwise $W_{a,g}=0$	$W_{a,g}$
Number of period g from resource k can be used for operation o of product p ordered at point a	$S_{a,p,o,k}$

The set of reference constraints for the problem was created and its mathematical/formal notation is included in Appendix A.

Constraint (1) determines whether in a given point a product p has been ordered (setting the value of variable $Xzk_{a,p}$). Constraints (2) ensures the order execution of operations for the product p (precedence constraint). Constraint (3) specifies the moment (period) from which resource k is needed to execute product p. Constraint (4) states no start is possible before orders appear. The term of delivery to the point a defines the constraint (5). Constraint (6) ensures that the number of available resources k in period g is not exceeded. Constraint (7) provides resource occupancy for the time of the order execution. Operations are not interrupted during their execution (8). Simultaneous completion of orders for product p from the given point a is ensured by constraints (9,10). Constraint (11) is responsible for the binarity of selected decision variables.

III. A DECLARATIVE DECISION SUPPORT FRAMEWORK FOR SCHEDULING GROUPS OF ORDERS-CONCEPT AND IMPLEMENTATION

The declarative decision support framework was proposed for scheduling groups of orders. The concept is based on the declarative programming paradigm, which allows high level programming with the use of predicates and facts. Due to the character of problems in the scheduling of groups of orders, CLP (Constraint Logic Programming) was selected from among many declarative options. The implementation of the framework was performed with the use of ECL^{i}PS^e [12].

The following general assumptions were applied:

- possibility of modeling constraints of any type;
- automatic generation of implementation models in the form of MILP models;
- data recorded as facts;

- problem dynamic taken into account (possibility of introducing new orders or groups of orders).

Figure 1 presents the general concept of the framework. The framework comprises several phases: modeling, presolving, generating and solving. It has two inputs and uses the set of facts. Inputs are the set of questions and the set of constraints to the reference model of a given problem. Based on them, the primary model of the problem is generated as a CLP model, which is then presolved. The built-in CLP method (constraint propagation [13]) and the method of problem transformation designed by the authors [6,9] (Section 3A) are used for this purpose. Presolving procedure results on the transformed model CLP^T. This model is the basis for the automatic generation of the MILP (Mixed Integer Linear Programming) model, which is solved in MP (with the use of an external Solver or CLP as a library). The general concept of the framework consists in modeling and presolving of a problem in the CLP environment with the final solution (including optimization) found in the MP environment. This approach is the result of experience as well as extensive research devoted to both environments [6,9,10] and their integration[6,14]. In all its phases, the framework uses the set of facts having the structure appropriate for the problem being modeled and solved (Fig. 2). The set of facts is the informational layer of the framework, which can be implemented as a relational database, XML database, etc.

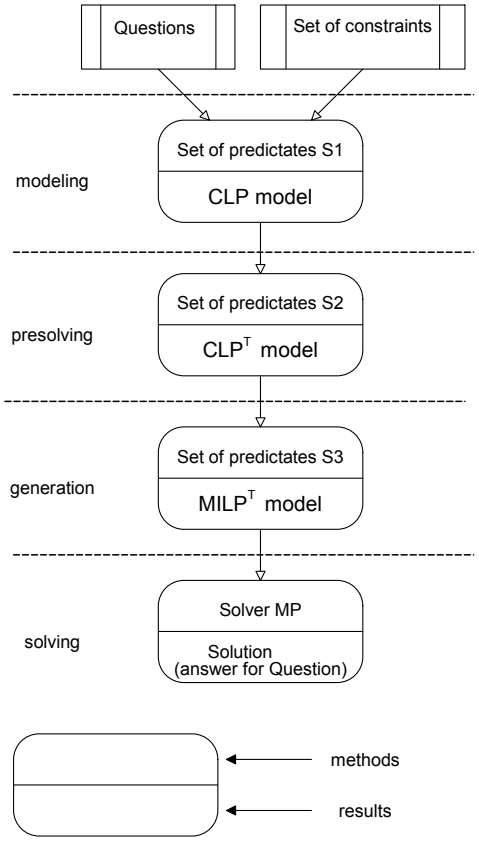


Fig. 1 A concept of a declarative decision support framework

In order to take into account the dynamics resulting, for example, from new orders, the $MILP^T$ model is solved iterative way with the use of the algorithm [15]. The main characteristic of this algorithm is iterative activation of MP solver and update of resource availability [15].

A. Presolving

The presolving phase is an important element of the framework as it makes it possible to simplify the model for the problem being solved and to reduce the combinatorial search space.

For the presolving phase to be effective, unfeasible combinations of model dimensions have to occur. In practice, unfeasible combinations of the index of decision variables and/or facts occur.

The proposed framework uses constraint propagation and transformation for the presolving procedure. Constraint propagation is a concept and method that appears in constrained-based environments. Constraint propagation embeds any reasoning which consists in explicity forbiding values from some varable domain of a problem, because all constraints can not be satisfied otherwise. Transformation transforms decision variables of the problem along with constraints and facts. The transformation method for the illustrative example is shown in Fig. 2, and the post-transformation variables are compiled in Table AII. For the problem presented, the transformation consisted in the change from the problem's operational representation into the resource representation. This resulted in the removal of all decision variables, parameters, etc. From the operation index, thereby reducing their numbers. The new set of decision variable, constraints and facts was the basis for creating the CLP^T model.

IV. COMPUTATION EXAMPLES FOR ILLUSTRATIVE MODEL

In order to verify and evaluate the proposed framework, many numerical experiments were performed for the illustrative example. In the first phase, all the experiments relate to the system with five points ($a=1..5$), eight order types (products) ($p=1..8$), eight resource types ($k=1..8$), thirty time periods ($g=1..30$) and eleven orders $zg_{v,p.}$ *(five groups of orders* Z_i in three periods*)*

All data instances for these experiments were recorded in the form of facts and included Appendix B.

Computational experiments consisted in asking questions Q1..Q7 to illustrative example. For each question was generated and solved suitable implementation model using declarative decision support framework. Orders are placed in groups for v_1=1, v_2=2 and v_3=5 (only for Q2 and Q5) periods. The answers to these questions are shown in Table II. Figure 3 shows the implementation schedule of all group of orders for question Q1 (minimizing makespan). A proper schedule utilization of resources corresponding to the schedule of Figure 3 is shown in Figure 4. By contrast, Figure 5 shows the implementation schedule of all group of orders for question Q6 (with all the resources k reduced by

50%). In analogy to the previous question, a proper schedule utilization of resources corresponding to the schedule of Figure 5 is shown in Figure 6.

The answer to the question Q4 determines the minimum resource requirements ($k_1..k_8$) necessary to complete all group of orders within T (Table II).

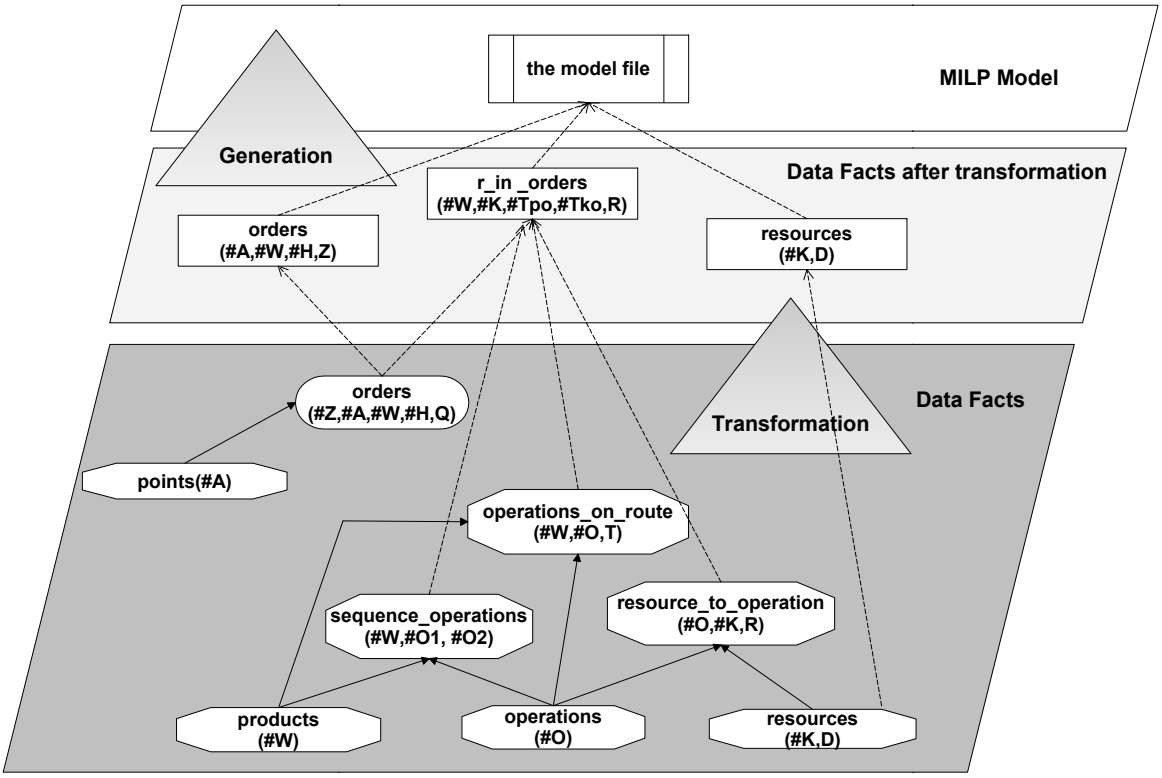


Fig. 2 Information layer of the framework (data fact before and after transformation, MILP model file)

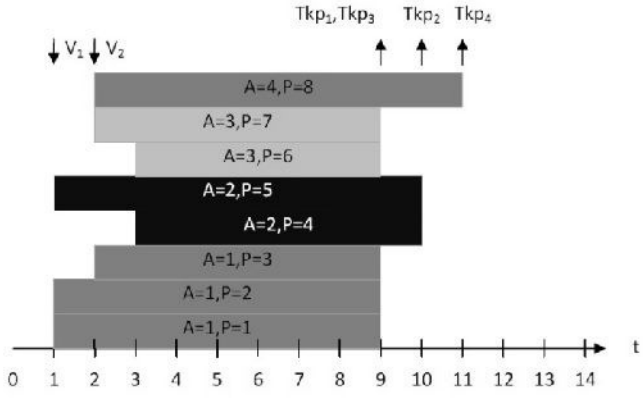


Fig. 3 Gantt chart for illustrative question Q1 (V_1=1, V_2=2, Tkp_1=Tpk_3=9, Tkp_2=10, Tkp_4=11, C_{max}=11)

The answers to the remaining questions confirm feasibility or unfeasibility of the schedule in the set conditions and at defined parameters (Table II). Information about the feasibility of the schedule and about available resources is especially useful when a new group of orders (Q2,Q5) appears. The last question from the set (Q7) is the question with a logical condition relating to disjoint use of resources.

In the second phase of the experiments, a comparative analysis was performed for questions Q1 and Q6 (most compute-intensive of all) in two environments (declarative decision support framework and MP) to evaluate the effectiveness and efficiency of the proposed framework relative to the classical MP environment.

Obtained more than 400-fold reduction of time searching for solutions (Table III). This is due to the fact that the application of the framework has allowed the reduction of decision variables from 134616 to 13873 (10-fold) and constraints from 252541 to 31239 (8-fold) (Table III).

TABLE II.
ANSWERS TO QUESTION FOR ILLUSTRATIVE EXAMPLE

Question	Parameters	Answer
Q1	---	C_{max}=11
$Q2_A$	T=12	NO
$Q2_B$	T=14	YES
$Q3_A$	T=12, k_1=k_2..= k_8=k=12	NO
$Q3_B$	T=13,k_1=k_2..= k_8=k=12	YES
Q4	T=20	k_1=8, k_2=8, k_3=6, k_4=7, k_5=4, k_6=5, k_7=1 k_8=6,
$Q5_A$	Z_3, v_3=5, k=12, T=12	NO
$Q5_B$	Z_3, v_3=5, k=12, T=13	YES
$Q6_A$	15%	C_{max}=11
$Q6_B$	50%	C_{max}=12
$Q7_A$	T=20, k_5 i k_7	YES
$Q7_B$	T=20, k_5 i k_8	NO

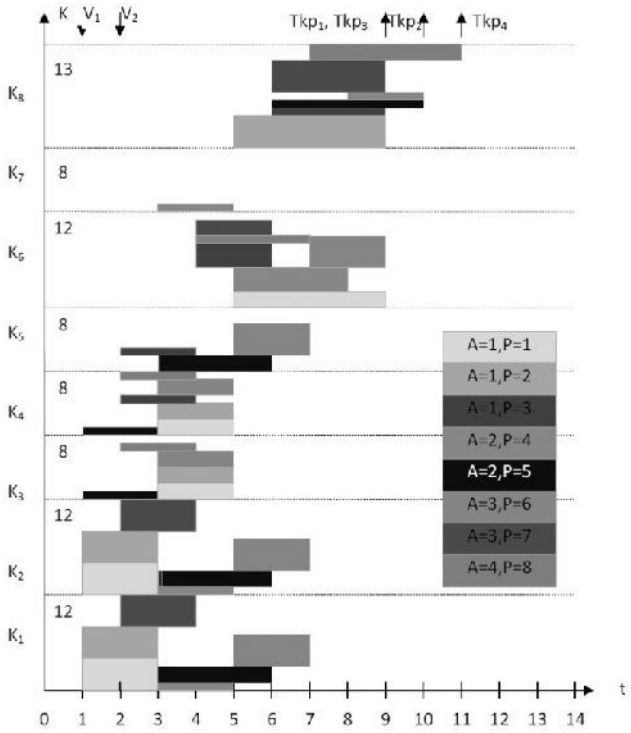


Fig. 4 Gantt chart for illustrative question Q1

TABLE III.
NUMERICAL EXPERIMENTS ON THE EFFICIENCY

Model	V(Vint)	C	Answer	T
Q1				
MILP	134616(134435)	252541	11	867
MILP^T	13873 (13690)	31239	11	2
Q6 (50%)				
MILP	134616(134435)	252541	12	856
MILP^T	13873 (13690)	31239	12	2

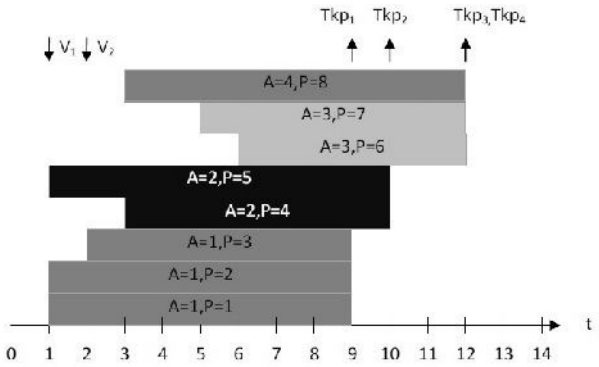


Fig. 5 Gantt chart for illustrative question $Q6_B$ (V_1=1, V_2=2, Tkp_1=9 Tpk_2=10, Tkp_3=Tkp_4=12, C_{max}=12)

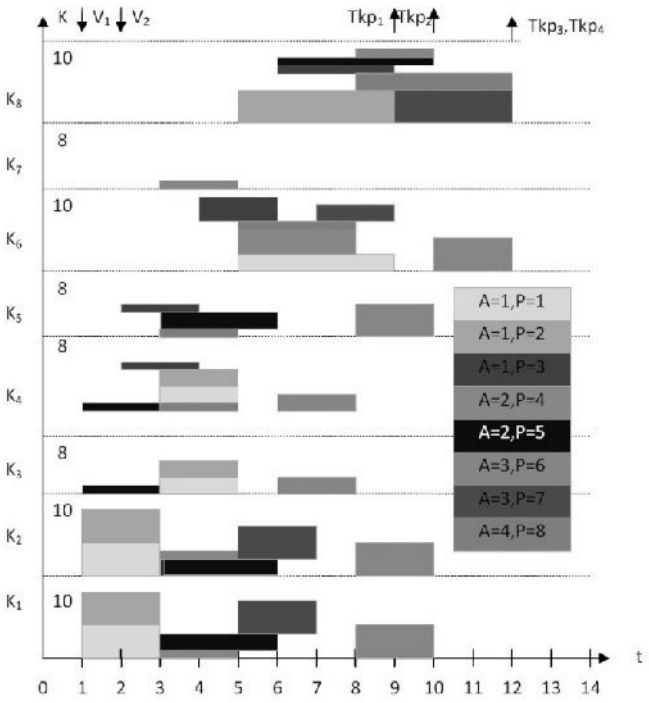


Fig. 6 Gantt chart for illustrative question $Q6_B$

TABLE IV.
STRUCTURE OF THE FACTS

Fact	Description
points (#A)	points of ordering and delivery A – point id
products (#W)	products, services, etc. W – product id
operations (#O).	technological operations, tasks, etc. O – operation id
resources (#K,D).	resources (machines, tools, peoples, etc.) K – resource id, D - the number of available resources
resource_to_operation (#O,#K,R)	the resources necessary to execute the operations O – operation id, K – resource id, R – the number of resource *k* needed for the operation
operations_on_route (#W,#O,T)	a set of operations for the product (route), W – product id, O – operation id, T – duration of the operation
sequence_operations (#W,#O,#O)	the order of operations for the product, W – product id, O – operation id
orders (#Z,#A,#W,V,Q)	orders, Z – orders id, A – point id, W – product id, V – order period, Q – order size
r_in _orders (#W,#K,#Tpo,#Tko,R)	the allocation of resources to products, W – product id, K – resource id, Tpo – the beginning of the use of the resource, Tko – the end of the use of the resource, R – the number of resource *k* needed for the operation

V. CONCLUSIONS

Two types of questions can be asked in the proposed declarative decision support framework.

General questions may require domain solution, which in practice determines the availability of resources to execute orders, timely execution etc. The specific wh-questions will in practice define the best, fastest, cheapest, or the most

expensive of the possible solutions. To obtain answers to these questions, optimization is necessary.

Both question types can contain logical conditions relating, for example, to the disjoint use of resources, etc.

The illustrative example shows only part of potential of the framework designed to increase both the speed and the size of the problems solved.

This is particularly evident if we compare the possibilities of the framework in relation to the classical approach based on mathematical programming (Table III).

Further work will consist in the implementation of more complex models, uncertainty, product demand interdependencies [16], fuzzy logic [17,18] etc., and as a cloud internet application [19]. New questions will be implemented to broaden the scope of decision support.

APPENDIX A

TABLE AI.
SUMMARY INDICES, PARAMETERS

Sets	
Set of points (tables)	LA
Set of resources	LK
Number of periods	LG
Number of periods in which orders can be entered	LV
The set of operations	LO
The set of products (orders)	LP
Indices	
Points (tables)	a=1..LA
Resources	k=1..LK
Period	g=1..LG
Period in which orders can be entered	v=1..LV
Operation	o=1..LO
Products	p=1..LP
Parameters	
Number used to convert periods to moments (for connecting index *g* with variable)	pp_g
The number of available resources *k* in the period *g*	$dkg_{k,g}$
Duration of operation *o* for product *p*	$t_{p,o}$
Time to complete all operations *o* for the product *p*	$t1_{p,o}$
If the operation o_1 precedes o_2 for product *p* than $kol_{p,o1,o2}$=1 otherwise $kol_{p,o1,o2}$=0	$kol_{p,o1,o2}$
If the operation *o* uses resource *k* than $zas_{o,k}$=1 otherwise $zas_{o,k}$=0	$zas_{o,k}$
Number of *k* resources needed for execution operation *o*	$r_{o,k}$
How much of the product *p* ordered at the point *a*.	$Zk_{a,p}$
The number of period in which new orders appeared.	ts
Inputs	
How much of the product *p* ordered at the point *a* in the period *v*.	$zgp_{v,a,p}$

TABLE AII.
DECISION VARIABLES AFTER TRANSFORMATION

Decision variables	
Calculated number of periods *g* delivery of all orders for point *a*.	Tkp_a
If at a given point *a* ordered product *p* then $Xzk_{a,p}=1$, otherwise $Xzk_{a,p}=0$	$Xzk_{a,p}$
If the execution of product *p* ordered at point *a* uses resource *k* in period *g* then $X_{a,p,k,g}$=1, otherwise $X_{a,p,k,g}$=0	$X_{a,p,k,g}$
If the execution of product *p* ordered at point *a* uses resource *k* in period *g* then $Xo_{a,p,k,g}=zk_{a,p}$, otherwise $Xo_{a,p,k,g}=0$	$Xo_{a,p,k,g}$
If *g* is the last period in which resource *k* is used in the execution of product *p* at point *a* then $Y_{a,p,k,g}=1$, otherwise $Y_{a,p,k,g}=0$	$Y_{a,p,k,g}$
If *g* is the last period in which orders are executed for point *a* then $W_{a,g}=1$, otherwise $W_{a,g}=0$	$W_{a,g}$
Number of period *g* from resource *k* can be used for product *p* ordered at point *a*	$S_{a,p,k}$

$$\begin{aligned} &Xzk_{a,p}\cdot LG \ge Zk_{a,p}\ \forall\ a=1..LA,\ p=1..LP \\ &Xzk_{a,p}\le zk_{a,p}\forall\ \ a=1..LA,\ p=1..LP \end{aligned} \tag{1}$$

$$\begin{aligned} &B_{a,p,o_1}+t_{p,o_1}\cdot Xzk_{a,p}=B_{a,p,o_2} \\ &\forall a=1..LA,\ p=1..LP,\ o_1,o_2=1..LO{:}kol_{p,o_1,o_2}=1 \end{aligned} \tag{2}$$

$$\begin{aligned} &S_{a,p,o,k}=B_{a,p,o}\forall a=1..LA,\ p=1..LP,o=1..LO,\ k=1..LK{:}\ r_{o,k}>0 \\ &S_{a,p,o,k}=0\forall a=1..LA,\ p=1..LP,o=1..LO,\ k=1..LK{:}r_{o,k}=0 \end{aligned} \tag{3}$$

$$B_{a,p,o}\ge ts\cdot Xzk_{a,p}\forall\ a=1..LA,\ p=1..LP,\ o=1..LO{:}t_{p,o}>0 \tag{4}$$

$$Tkp_a\ge B_{a,p,o}+(t1_{p,o}+t_{p,o})\cdot Xzk_{a,p}\forall a=1..LA,p=1..LP \tag{5}$$

$$\sum_{a=1}^{LA}\sum_{p=1}^{LP}\sum_{o=1}^{LO}(Xo_{a,p,o,k,g}\cdot r_{o,k})\le dkg_{k,g}\forall k=1..LK,g=1..LG \tag{6}$$

$$\begin{aligned} &\sum_{g=1}^{LG}X_{a,p,o,k,g}=t_{p,o}\cdot Xzk_{a,p} \\ &\forall a=1..LA,p=1..LP,o=1..LO,k=1..LK \\ &\sum_{g=1}^{LG}Xo_{a,p,o,k,g}=t_{p,o}\cdot Zk_{a,p} \\ &\forall a=1..LA,p=1..LP,o=1..LO,k=1..LK \end{aligned} \tag{7}$$

$$\begin{aligned} &X_{a,p,o,k,g-1}-X_{a,p,o,k,g}\le Y_{a,p,o,k,g-1} \\ &\forall a=1..LA,p=1..LP,o=1..LO,k=1..LK,g=2..LG \\ &\sum_{g=1}^{LG}Y_{a,p,o,k,g}\le Xzk_{a,p} \\ &\forall a=1..LA,\ p=1..LP,\ o=1..LO,\ k=1..LK \end{aligned} \tag{8}$$

$$Y_{a,p,o,k1,g}=Y_{a,p,o,k2,g}\ \forall a=1..LA,p=1..LP,o=1..LO,k1,k2=1..K,\ g=1..LG:zas_{o,k1}=1\wedge zas_{o,k2}=1 \tag{9}$$

$$\begin{aligned} &Tkp_a=\sum_{g=1}^{LG}pp_g\cdot W_{a,g}\ \forall\ a=1..LA \\ &Y_{a,p,o,k,g-t1_{m,o}}\le W_{a,g} \\ &\forall\ a=1..LA,p=1..LP,o=1..LO,k=1..LK,g-t1_{p,o}\ge 0 \\ &Y_{a,p,o,k,g}=W_{a,g+t1_{m,o}} \\ &\forall\ a=p..LP,o=1..LO,k=1..LK,g=1..LG+t1_{p,o}\le LG \end{aligned} \tag{10}$$

$$
\begin{gathered}
X_{a,p,o,k,g} \in \{0,1\} \\
\forall\, a = 1..LA, p = 1..LP, o = 1..LO, k = 1..LK, g = 1..LG \\
Xo_{a,p,o,k,g} \in C \\
\forall\, a = 1..LA, p = 1..LP, o = 1..LO, k = 1..LK, g = 1..LG \\
Y_{a,p,o,k,g} \in \{0,1\} \\
\forall\, a = 1..LA, p = 1..LP, o = 1..LO, k = 1..LK, g = 1..LG \\
S_{a,p,o,k} \in C \;\forall a = 1..LA, p = 1..LP, o = 1..LO, k = 1..LK \\
Tkp_a \in C \;\forall\, a = 1..LA \\
B_{a,p,o} \in C \;\forall\, a = 1..LA, p = 1..LP, o = 1..LO \\
W_{a,g} \in C \forall a = 1..LA, g = 1..LG \\
Xzk_{a,p} \in \{0,1\} \forall a = 1..LA, p = 1..LP
\end{gathered}
\tag{11}
$$

APPENDIX B

```
points('A1'). points('A2'). points('A3').
points('A4'). points('A5').
products('W1'). products('W2'). products('W3').
products('W4'). products('W5'). products('W6').
products('W7'). products('W8').
operations('O1'). operations('O2').
operations('O3'). operations('O4').
operations('O5'). operations('O6').
operations('O7'). operations('O8').
operations('O9'). operations('O10').
resources('K1',20). resources('K2',20).
resources('K3',20). resources('K4',20).
resources('K5',20). resources('K6',20).
resources('K7',20). resources('K8',20).
resource_to_operation('O1','K1',2).
resource_to_operation('O1','K2',2).
resource_to_operation('O2','K3',1).
resource_to_operation('O2','K4',1).
resource_to_operation('O3','K1',2).
resource_to_operation('O3','K2',2).
resource_to_operation('O3','K5',2).
resource_to_operation('O4','K1',1).
resource_to_operation('O4','K2',1).
resource_to_operation('O4','K7',1).
resource_to_operation('O5','K4',1).
resource_to_operation('O5','K5',1).
resource_to_operation('O6','K6',3).
resource_to_operation('O7','K8',1).
resource_to_operation('O8','K6',2).
resource_to_operation('O9','K6',1).
resource_to_operation('O9','K6',1).
resource_to_operation('O10','K8',2).
operations_on_route('W1','O1',2).
operations_on_route('W1','O2',2).
operations_on_route('W1','O9',4).
operations_on_route('W2','O1',2).
operations_on_route('W2','O2',2).
operations_on_route('W2','O10',4).
operations_on_route('W3','O5',2).
operations_on_route('W3','O6',2).
operations_on_route('W3','O7',3).
operations_on_route('W4','O4',2).
operations_on_route('W4','O6',3).
operations_on_route('W4','O7',2).
operations_on_route('W5','O2',2).
operations_on_route('W5','O3',3).
operations_on_route('W5','O7',4).
operations_on_route('W6','O2',2).
operations_on_route('W6','O3',2).
operations_on_route('W6','O8',2).
operations_on_route('W7','O1',2).
operations_on_route('W7','O9',2).
operations_on_route('W7','O10',3).
operations_on_route('W8','O2',2).
operations_on_route('W8','O9',3).
operations_on_route('W8','O10',4).
sequence_operations('W1','O1','O2').
sequence_operations('W1','O2','O9').
sequence_operations('W2','O1','O2').
sequence_operations('W2','O2','O10').
sequence_operations('W3','O5','O6').
sequence_operations('W3','O6','O7').
sequence_operations('W4','O4','O6').
sequence_operations('W4','O6','O7').
sequence_operations('W5','O2','O3').
sequence_operations('W5','O3','O7').
sequence_operations('W6','O2','O3').
sequence_operations('W6','O3','O8').
sequence_operations('W7','O1','O9').
sequence_operations('W7','O9','O10').
sequence_operations('W8','O2','O9').
sequence_operations('W8','O9','O10').
orders('Z1','A1','W1',1,2).
orders('Z1','A1','W2',1,2).
orders('Z1','A1','W3',1,1).
orders('Z2','A2','W4',1,1).
orders('Z2','A2','W5',1,1).
orders('Z3','A3','W6',2,2).
orders('Z3','A3','W7',2,2).
orders('Z4','A4','W8',2,1).
orders('Z5','A5','W1',5,1).
orders('Z5','A5','W7',5,2).
orders('Z5','A5','W8',5,1).
```

REFERENCES

[1] I. Ribas, R. Leisten, J.M. Framinan, "Review and classification of hybrid flow shop scheduling problems from a production system and a solutions procedure perspective", in: *Computer Operation Research,* 37, pp.1439–1454, 2010.

[2] B. Tadayon, N. Salmasi, "A two-criteria objective function flexible flowshop scheduling problem with machine eligibility constraint", in: *The International Journal of Advanced Manufacturing Technology,* 64(5-8), pp. 1001-1015, 2013.

[3] K. Apt, M. Wallace, "Constraint Logic Programming using Eclipse", Cambridge: Cambridge University Press, 2006.

[4] G. Bocewicz, I. Nielsen, Z. Banaszak, "Iterative multimodal processes scheduling", in: *Annual Reviews in Control,* 38(1), pp. 113-132, 2014.

[5] F. Rossi, P. Van Beek, T. Walsh, "Handbook of Constraint Programming", New York: Elsevier Sc. Inc, 2006.

[6] P. Sitek J. Wikarek, "A hybrid method for modeling and solving constrained search problems", in: *Federated Conference on Computer Science and Information Systems (FedCSIS 2013)*, pp. 385-392, 2013.

[7] P. Sitek, J. Wikarek, "A Hybrid Programming Framework for Modeling and Solving Constraint Satisfaction and Optimization Problems", in: *Scientific Programming*, vol. 2016, Article ID 5102616, 2016. doi:10.1155/2016/5102616.

[8] P. Sitek, J. Wikarek, "A hybrid framework for the modelling and optimisation of decision problems in sustainable supply chain management", in: *International Journal of Production Research*, pp. 6611-6628, 2015. doi:10.1080/00207543.2015.1005762.

[9] P. Sitek, "A hybrid CP/MP approach to supply chain modelling, optimization and analysis", in: *Federated Conference on Computer Science and Information Systems (FedCSIS)*, pp. 1345-1352, 2014. doi:10.15439/2014F89

[10] P. Sitek, "A hybrid approach to the two-echelon capacitated vehicle routing problem (2E-CVRP)", in: *Advances in Intelligent Systems and Computing,* 267**,** pp. 251–263, 2104. doi:10.1007/978-3-319-05353-0_25.

[11] A. Schrijver, "Theory of Linear and Integer Programming", John Wiley & Sons, New York, NY, USA, 1998.

[12] Eclipse, 2015, Eclipse - The Eclipse Foundation open source community website, Accessed August 12, www.eclipse.org.

[13] B.M.W. Cheng, K.M.F. Choi, J.H.M. Lee, J.C.K. Wu, "Increasing Constraint Propagation by Redundant Modeling: an Experience Report", Constraints May 1999, Volume 4, Issue2, pp 167-192, 1999.

[14] M. Milano, M. Wallace M., "Integrating Operation Research in Constraint Programming", in *Annals of Operation Research*, 175(1), 2010, pp. 37-76, DOI:10.1007/s10479-009-0654-9.

[15] P. Sitek, J. Wikarek, "A novel approach to decision support and optimization of group job handling for multimodal processes in manufacturing and services" in *15th IFAC/IEEE/IFIP/IFORS Symposium on Information Control Problems in Manufacturing*, Ottawa, Kanada,2015, pp 2183-2188,doi:10.1016/j.ifacol.2015.06.401

[16] P. Nielsen, I. Nielsen, K. Steger-Jensen, Analyzing and evaluating product demand interdependencies in *Computers in Industry*, 61 (9), 2010, 869-876, doi:10.1016/j.compind.2010.07.012.

[17] M. Relich, W. Muszynski, The use of intelligent systems for planning and scheduling of product development projects in *Procedia Computer Science*, vol. 35, 2014, pp. 1586–1595.

[18] G. Kłosowski, A. Gola, A. Świć, Application of Fuzzy Logic Controler for Machine Load Balancing in *Discrete Manufacturing System,* [in:]. K. Jackowski et. al. (Eds.): IDEAL 2015, LNCS 9375, 2015, pp. 256-263.

[19] S. Bak, R. Czarnecki, S. Deniziak "Synthesis of Real-Time Cloud Applications for Internet of Things", in *Turkish Journal of Electrical Engineering &Computer Sciences,* 2013, DOI: 10.3906/elk-1302-178.

3rd International Workshop on Cyber-Physical Systems

PROLIFERATION of computers in everyday life requires cautious investigation of approaches related to the specification, design, implementation, testing, and use of modern computer systems interfacing with real world and controlling their environment. Cyber-Physical Systems (CPS) are physical and engineering systems closely integrated with their typically networked environment. Modern airplanes, automobiles, or medical devices are practically networks of computers. Sensors, robots, and intelligent devices are abundant. Human life depends on them. Cyber-physical systems transform how people interact with the physical world just like the Internet transformed how people interact with one another.

The event is a continuation and extension of 2006-2010 Real-Time Software FedCSIS workshops as well as 2013 and 2015 IWCPS. The objective of the workshop is to assemble and develop a community with main interest in cyber-physical systems.

Due to an extensive scope of the topics, the workshop will accept papers in the following areas:

- Control Systems
  - embedded/networked/intelligent
  - wireless sensing/actuation
  - adaptive/predictive
- Scalability/Complexity
  - modularity
  - design methodology
  - legacy systems
  - tools
- Interoperability
  - concurrency
  - models of computation
  - networking
  - heterogeneity
- Validation and Verification
  - assurance
  - certification
  - simulation
- Cyber-security
  - intrusion detection
  - resilience
  - privacy
  - attack vectors
- Applications of CPS
  - robotics
  - transportation
  - military
  - medical
  - consumer
  - manufacturing
  - power systems
- CPS Education
  - curriculum development
  - web-based laboratories
  - academic courses
  - pedagogy issues

EVENT CHAIRS

- **Grega, Wojciech,** AGH University of Science and Technology, Poland
- **Kornecki, Andrew J.,** Embry Riddle Aeronautical University, United States
- **Nigro, Libero,** Università della Calabria, Italy
- **Szmuc, Tomasz,** AGH University of Science and Technology, Poland
- **Zalewski, Janusz,** Florida Gulf Coast University, United States

PROGRAM COMMITTEE

- **Babiceanu, Radu,** Embry Riddle Aeronautical University, United States
- **Ehrenberger, Wolfgang,** University of Applied Science Fulda, Germany
- **Golatowski, Frank,** University of Rostock, Germany
- **Gomes, Luis,** Universidade Nova de Lisboa, Portugal
- **Halang, Wolfgang A.,** Fernuniversitaet, Germany
- **Letia, Tiberiu,** Technical University of Cluj-Napoca, Romania
- **Malec, Jacek,** Lund University, Sweden
- **Marwedel, Peter,** Technische Universität Dortmund, Germany
- **Motus, Leo,** Tallinn University of Technology, Estonia
- **Saglietti, Francesca,** University of Erlangen-Nuremberg, Germany
- **Sanden, Bo,** Colorado Technical University, United States
- **Trybus, Leszek,** Rzeszow University of Technology, Poland
- **Vardanega, Tullio,** University of Padova - Dept. of Maths, Italy
- **Villa, Tiziano,** Università di Verona, Italy
- **Zoebel, Dieter,** University Koblenz-Landau, Germany

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 743–749
DOI: 10.15439/2016F50
ACSIS, Vol. 8. ISSN 2300-5963

Situational Awareness Network for the Electric Power System: the Architecture and Testing Metrics

Damiano Bolzoni
SecurityMatters BV
Eindhoven, The Netherlands
Email: damiano.bolzoni@secmatters.com

Rafał Leszczyna
Gdańsk University of Technology
Faculty of Management and Economics
Narutowicza 11/12, Gdańsk, Poland
Email: rle@zie.pg.gda.pl

Michał R. Wróbel
Gdańsk University of Technology
Faculty of Electronics,
Telecommunications and Informatics
Narutowicza 11/12, Gdańsk, Poland
Email: wrobel@eti.pg.gda.pl

***Abstract*—The contemporary electric power system is highly dependent on Information and Communication Technologies which results in its exposure to new types of threats, such as Advanced Persistent Threats (APT) or Distributed-Denial-of-Service (DDoS) attacks. The most exposed components are Industrial Control Systems in substations and Distributed Control Systems in power plants. Therefore, it is necessary to ensure the cyber security of this critical infrastructure and develop new cyber security technologies able to protect from advanced cyber threats. In this paper a pioneering Situation Awareness Network for the electric power system is presented together with a set of metrics for its testing.**

I. Introduction

MODERN energy infrastructures aim at reducing peak demand, shifting usage to off-peak hours, lowering total energy consumption and carbon dioxide footprint [1] or enabling consumers to control their power consumption based on local needs and real-time electricity price rates [2].

To meet these requirements it is necessary to ensure the continuous exchange of data between all points of the network. Although the communication infrastructure may partially exist, it is necessary to facilitate its vast expansion by increasing bandwidth (among the others due to the introduction of two-way communication as an inherent component of the new energy infrastructure and smart grid) and connecting consumers (residential, commercial, industrial, etc.). To reduce the costs which are incurred by this process, the Internet is often used as communication backbone for the energy management systems [1].

However, such an approach exposes the power system to a great security breach. Every network layer and technology used in the new energy infrastructure represents a potential target of a cyber-attack. This in particular refers to Industrial Control Systems (including SCADA) in substations and Distributed Control Systems (DCS) in power plants. Moreover in the recent years wireless networks have been widely employed as part of many industrial communication systems, which exposes the entire network to even greater risk [3].

Advanced Persistent Threats (APT) are dedicated attacks able to persistently target a specific entity and to cause an intended effect, such as an interruption to the power supply [4], [5]. DDoS attacks, on the other hand, attempt to delay, block or corrupt the communication in the grid [6].

Stuxnet [7] was the first wide manifestation of malware that was specifically designed to attack networked industrial control systems used in the power system. Detected for the first time in 2010, Stuxnet is a cyber worm able to infect process servers and Programmable Logic Controllers (PLCs) and alter physical processes. The ultimate goal of Stuxnet is to sabotage the attacked facility by reprogramming programmable logic controllers (PLCs) to render them operating out of their specified boundaries. Later studies revealed that Stuxnet was not the first threat of that type. In fact that it had its precursor called Flame that was undetected [8]. Flame is a large complex malware designed to aggressively gather information from its target systems. Apart of conventional information stealing methods it is able to capture Skype calls and record audio [9].

Since the manifestation of Stuxnet both information security experts and hackers have shown a much greater level of interest in this area. As a result, 64 ICS vulnerabilities were discovered in 2011 and 98 additional ones were announced in the first eight months of 2012 alone – more than the total number for the preceding seven years combined [10]. In parallel sophisticated attacks have been appearing – Duqu, Red October, Gauss and Black Energy – each of them more complex and advanced than its predecessor [9], [11]. Duqu was designed to steal information in preparation for a Stuxnet-like attack and it used new techniques never previously noted [9]. Red Dragon and Gauss utilise encryption in order to effectively penetrate the infiltrated information systems [11], [9]. Black Energy is the most recently discovered malware which aims at Industrial Control Systems used in critical infrastructures [12].

Taking into account all these threats and the attacks already carried out, it is necessary to take countermeasures. Standard cyber security technologies and best practices – such as access control, anti-malware, firewalls, intrusion detection and prevention systems, defence in depth, and system hardening

The study presented in this paper is based on work carried out in the DEnSeK (Distributed Energy Security Knowledge) project founded by the European Commission, Directorate-General for Home Affairs (Programme „Prevention, Preparedness and Consequence Management of Terrorism and other Security-related Risks" – CIPS, Project Reference: HOME/2012/CIPS/AG/ 4000003772) and partially supported from the project funds. It is also supported by the DS Programs of Faculty of Management and Economics and Faculty of Electronics, Telecommunications and Informatics of Gdańsk University of Technology.

– are indispensable in protecting the power system. However, they are only a partial solution [4], [13], [14], [15].

To counter the evolved, highly sophisticated threats, advanced cyber security technologies are required, such as Security Information and Event Management (SIEM) systems, application whitelisting, and Trusted Platform Modules (TPM) [4], [13], [16] together with an efficient and effective risk assessment and management [17]. Developing and deploying Situation Awareness Networks (SANs) with SIEM software will improve situational awareness and will allow for better control and faster response to threats [18].

Such a Situation Awareness Network (SAN) has been developed in the project DEnSeK (Distributed Energy Security Knowledge) [19]. The project aimed at improving the security and resilience of the new energy infrastructure against cyber-threats by providing a platform for the security knowledge exchange between companies of the European energy sector and establishing a European Energy ISAC (Information Sharing and Analysis Centre) which enables interactive and real-time knowledge and information sharing between all involved parties [19].

In this paper the SAN architecture is presented along with the set of metrics to be used for its evaluation. To the best of authors' knowledge such a dedicated set of metrics for Situation Awareness Networks (SANs) has not been proposed so far, most probably because the concept of SAN is relatively new. It must be underlined that evaluation of Situation Awareness Networks is an area distinct from the quantitative assessment of the level of situational awareness. For the latter several approaches exist [20].

II. Situational awareness network architecture

The Situational Awareness Network encompasses and combines a number of diverse network-based sensors, which facilitate network traffic and data monitoring and detection of various events. Collected and processed data sets are visualised to a SAN operator who responds to emerging threats.

The need for combining together multiple sensors stems from the observation that in the past half-decade monitoring tools have become more specialised and now they focus on specific threat vectors and/or analysis approaches. Hence, in order to offer a broad overview of network activities and potential issues, it is crucial to combine diverse monitoring engines.

The purpose of the visualization is two-fold. First, operators can spot anomalies that the automatic systems might not be able to detect or might not be configured to detect. In this case, a visualisation dashboard supports the analysis of a large amount of data as it reduces it significantly focusing on key parameters for detecting anomalies.

Secondly, once an event is reported by one of the automatic systems (for instance, a malware spread is detected), operators can leverage the visualisation dashboard to observe the way network traffic evolves and either confirm or reject the alert previously raised.

In the DEnSeK project a three-tier architecture of the Situational Awareness Network, presented in Fig. 1, was proposed. The lowest, data tier consists of sensors which collect network data. In the logic tier, Security Information and Event Management (SIEM) software processes data from sensors and transmits them to the top layer. Finally in the presentation tier, the dashboard visualises the data by a user-friendly operator interface.

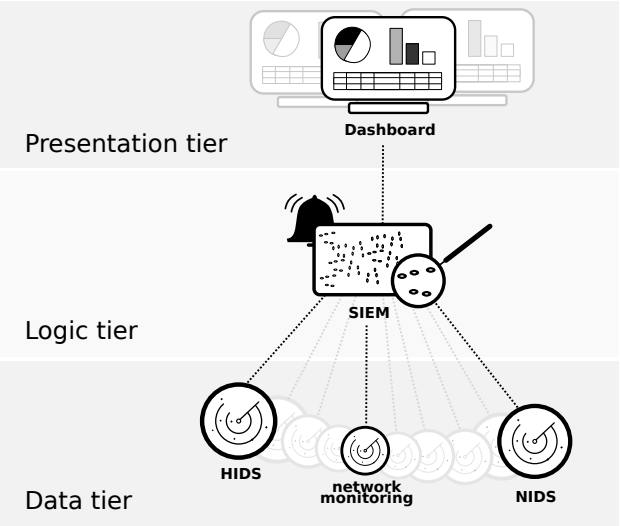


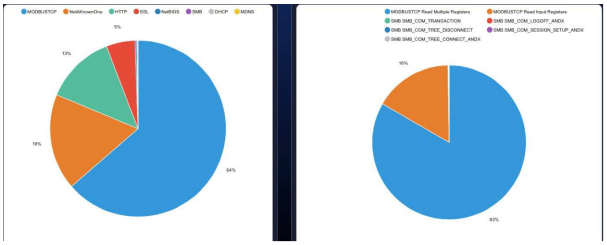
Fig. 1. SAN three-tier architecture

In the approach presented in this paper, SAN needs a signature-based NIDS (such as Snort [21], [22] and Suricata [23]), to detect well-known attack payloads, and several behavioural-based engines to analyse both payloads and flows for anomalies. As it relies on open-source/freely available tools, currently exist few alternatives for behavioural-based systems that can be used in production environments. One of them is Bro [24], [25], which can be used to code any type of algorithm on top of its protocol parsers.

On top of regular network monitoring tools and SIEMs that are available off-the-shelf, a visualisation dashboard is located. Its main role is to allow operators to observe the behaviour of the underlying industrial network. The dashboard provides several widgets, presented in Fig. 2 that can be instantiated to present various dimensions of network traffic (IP addresses, TCP ports, protocols, etc.) using different metrics (bytes, packets, protocol messages, etc.).

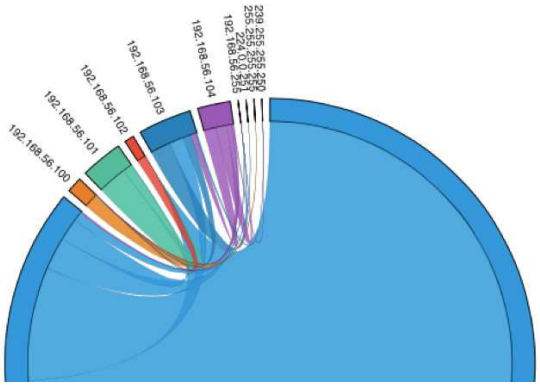
The central SIEM node is provided with Syslog (system logging) messages by various network-based sensors. This is a standard practice that enables required flexibility while providing all the necessary information. The visualisation dashboard leverages diverse software and components in order to deliver the extracted metadata to a central repository.

At the bottom of the architecture, a Linux-based computer equipped with Argos (a Linux-based SCADA alternative [26]) analyses network streams and extracts relevant metadata. Such data are sent via Apache Kafka [27] publish-subscribe messag-



(a) Network protocols and commands used in the communication

(b) Bandwidth by protocol and destination



(c) Connections between network hosts

Fig. 2. SAN Dashboard widgets

ing service to a central repository based on Druid [28]. This is a real-time data store that takes advantage of an in-memory architecture to facilitate data aggregation and fast querying. Data processed by Druid are queried via graphical web widgets based on the D3JS framework.

The visual analytics component does not fulfil only the task of depicting network data flows and interactions in real-time, but can be also applied to guiding the development of specific controls and checks. Every organisation in fact will exhibit a slightly different network layout and configuration, even those within the same industrial sector and/or running the same software package. This diversity requires a certain degree of customisation of controls, to tackle the specificity of a certain environment.

End users can perform an assessment of their network through the visual analytics component to baseline network behaviour, discover misconfigurations and assess issues. After this initial assessment, end users can select key indicators that can point at operational issues or cybersecurity breaches. To provide some examples, RTUs used in the field typically exchange data with the SCADA master via long-lasting connections that could be running for weeks or months. In case an RTU loses too often connectivity and re-establishes connectivity frequently, this could indicate an issue with the device itself (for instance, end of life) or with the network infrastructure (because of a wireless link). Key indicators can be enforced by writing a specific script in Bro, or a signature in Snort and/or Suricata.

III. Metrics in the Testing and Evaluation Process

Testing is an integral part of software development process. In the document „Standard Glossary of Software Engineering Terminology" IEEE defines *testing* as the „process of operating a system or a component under specified conditions, observing or recording the result and making an evaluation of some aspect of the system or component" [29].

The goal of testing is to detect the difference between existing and required conditions and to evaluate the features of the software items [30]. Currently, testing is a mature and well-defined area of software engineering. Good testing process design should ensure the repeatability, manageability and measurability [31].

As a part of the development of the Situational Awareness Network for the DEnSeK project, integration tests have been carried out. Their aim was to validate a selection of SAN components and check their operational capability in a complex test environment. During the tests appropriate interaction between the components was verified.

The tests were performed in the cyber security laboratory of one of the largest European electricity companies. They proved that the architecture and system components were properly selected and the system operates as intended [32].

Despite the positive results of the tests, the lack of quantitative indicators made it difficult to objectively assess the results. It was only possible to grade binary – it works or does not work. The extent to which the requirements are met, however, could not be determined in a measurable way. As far as only integration tests are concerned, this binary evaluation is sufficient. Nevertheless the majority of evaluations require higher precision.

Thus in order to enable objective evaluation of a software product and its development process *software metrics* were introduced. A *software metric* is a „quantitative measure of the degree to which a system, component or process has given attribute" [29]. The knowledge gathered on the basis of metrics should lead to an improved process and products [33]. The metrics can be divided into two groups: product metrics and metrics for testing process [34].

Metrics from the first group are used to provide information about the quality and maturity of the tested product. They

facilitate early detection of product flaws and related problems and enable their more accurate correction or elimination. In addition, metrics provide quantitative criteria which may be used in the process of acceptance of the final product.

The latter group contains metrics that allow for monitoring of the progress of the testing process and its results after the execution. They are used on one hand, to evaluate the effectiveness of the testing process. On the other hand, they provide test termination criteria.

The use of software metrics as objective evaluation criteria is extremely important in the management of the software development process [35].

Software metrics have been developed practically for all application domains. However, to the best of authors' knowledge the metrics for Situational Awareness Platforms have not been proposed so far. This is most probably due to the fact that the concept of Situation Awareness Network and the implementing it platforms are relatively recent.

In order to fill this gap the relevant research studies have been investigated to provide a comprehensive set of metrics for testing SANs. The metrics' proposals in several fields have been identified and analysed, including Intrusion Detection and Prevention Systems, Security Information and Event Management systems as well as general domains such as software engineering, testing or cyber security. The data collected allowed for selecting the relevant metrics.

When choosing the metrics, the Jaquith's [36] recommendations were taken into account. According to him, good metrics should be [36]:

- consistently measured,
- expressed as cardinal number or percentage,
- expressed using unit of measure,
- contextually specific,
- possible to obtain at reasonable cost.

In order to design the test procedure for the SAN system, a set of metrics was selected. Metrics described in Section IV regard the testing process. They facilitate the control and management of the testing process, as well as deciding when to end it.

Other metrics are related to the product. In the evaluation, the product is understood as a complete SAN system. Given the characteristics of the system the metrics are also divided into two groups. In Section V cyber security metrics are presented. They allow for evaluating the core SAN functionality answering the question of how the system copes with the detection of security threats. The last group of metrics, presented in Section VI, refer to system usability. As one of the functions of the SAN, provided by the Dashboard, is the visualisation of security threats to an operator, the quality of user interface is very important.

The main criterion taken into account when selecting the metrics was the possibility of their straightforward implementation to assess SAN platform at every stage of development. In addition, a set of metrics was chosen to cover as widely as possible all aspects of the testing process.

IV. Metrics for testing process

Testing metrics are widely used in the field of software testing. Their aim is to „provide information about the testing status of a software product" [34].

Quadri and Farooq [34] divided testing metrics into several groups. First of all they highlighted the metrics related to measuring time, such as time required to run a test, time interval between failures or number of failures in specific time interval. After that the metrics for evaluating test efficiency, source code coverage and quality were described. Finally metrics related to defect identification and fixing were presented.

Chen et al. [37] conducted an in-depth analysis of software metrics, examining the effectiveness of a set of complementary metrics for cost, time, and quality to measure the quality of test process. Based on the result they proposed four new testing metrics: two related to product improvements and two related to costs.

Kaur et al. [35] surveyed, classified and systematically analysed the metrics proposed in the previous decades. They discussed advantages or disadvantages for each product metric along with its need and purpose. The suitability, effect, data calibration and interpretation of metrics was also evaluated.

Based on the studies as well as the specificity of the DEnSeK project four testing metrics were selected.

A. Source code coverage

The source code coverage metric enables evaluating the confidence in the effectiveness of a test suite. The metric is defined as follows:

$$SC = \frac{St_t}{St} \tag{1}$$

where:

- SC – source code coverage,
- St_t – number of statements of a source code covered by test suite
- St – number of statements of a source code,

The metric shows what part of the source code has been covered with tests. If the value is too low, there should be written additional test cases for uncovered source code.

B. Test case defect density

Test case defect density metric indicates whether the test cases are effective and efficient in their ability to detect a larger number of defects. It is defined as:

$$DD = \frac{F}{TE} \times 100\% \tag{2}$$

where:

- DD – test case defect density,
- F – failures detected,
- TE – number of executed test cases.

C. Failures detection rate

Failures detection rate metric test indicates whether the prepared tests are time effective in terms of the number of detected defects per unit time. The metric is defined by the following formula:

$$FD = \frac{F_T}{T} \tag{3}$$

where:

- FD – failures detection rate
- F_T – failures detected in T time
- T – number of business days used for testing

D. Test improvement in product quality

Test improvement in product quality metric shows the relation between the number of weighted defects detected and the size of the product release. It is defined as:

$$TI = \frac{W_p}{KCSI} \tag{4}$$

where:

- TI – test improvement in product quality,
- W_p – number of weighted defects found in one specific test phase,
- $KCSI$ – number of new or changed source lines of code in thousands.

The higher this number, the higher is the improvement of the quality of the product contributed during this test phase.

V. Cyber security metrics

Cyber security metrics are strictly related to the functional operation of the Situational Awareness Network. The selection was made among the metrics defined for security systems such as Intrusion Detection Systems. One of the main problems that SAN operators would face is the reliability of the threats detection. There are two main aspects to be taken into consideration: false positives and true negatives. [38]

A *true positive* is when SAN informs about threat that really exists. This is the desired situation. A *false positive* takes place when SAN informs about threat that does not occur. A *true negative* refers to the situation when SAN does not inform about threat that really occurs.

Using these terms three metrics have been defined. Additionally two metrics based on the research conducted by Bayuk and Mostashari [39] were proposed.

A. Accuracy

Accuracy metric describes the proportion of true results (both true positives and true negatives) in the population of all network events. It is defined as:

$$A = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

where:

- A – detection accuracy
- TP – number of true positives
- TN – number of true negatives
- FP – number of false positives
- FN – number of false negatives

A higher value indicates a more reliable system operation.

B. Detection rate

Detection rate determines the effectiveness of threats' detection. When the metric value is closer to 1, the system is more effective. A value of 1 means that each threat has been detected.

$$DR = \frac{TP}{TP + FN} \tag{6}$$

where:

- DR – detection rate
- TP – number of true positives
- FN – number of false negatives

C. False positive rate

False positives are one of SAN biggest issues. Their frequent occurrence significantly undermines the effectiveness of the SAN. Efforts should be made to the lowest value of this indicator.

$$FPR = \frac{FP}{FP + TP} \tag{7}$$

where:

- FPR – detection accuracy
- FP – number of false positives
- TP – number of true positives

D. Mean Time Between Failures

Mean time between failures (MTBF) is a standard metric that describes reliability of the system. In the case of SAN failure is defined by the occurrence of either false positive or true negative. The metric is defined as:

$$MTFB = \frac{\sum_2^{NF}(B_n - E_{n-1})}{NF - 1} \tag{8}$$

where:

- $MTFB$ – Mean Time Between Failures
- B_n – beginning of *n*-th failure
- E_n – end of *n*-th failure
- NF – number of failures

E. Time To Protect

The metric is defined as the mean time between the detection of the threat and noticing it by the operator. In this way, both the effectiveness and efficiency of the system, as well as the legibility of the information about the threat on the dashboard, are evaluated.

$$TTP = \frac{\sum_1^{NT}(A_n - D_n)}{NT} \tag{9}$$

where:

- TTP – Time To Protect
- A_n – time of *n*-th threat notice
- D_n – time of *n*-th threat detection
- NT – number of threat detections

VI. User experience metrics

In addition to testing against the above criteria, software should be evaluated in terms of usability. This is particularly relevant to software interface, but not exclusively. As far as the DEnSeK SAN is concerned, the Dashboard requires special attention in regard to usability. Based on the metrics proposed by Tullis and Albert [40], the following metrics are proposed for evaluation of the SAN usability.

A. Task success

This metric enables measuring the extent to which a user is able to perform a given task. Task success can be measured binary (succeed/failed), or as a level of success. The tasks with a lower coefficient of success must be analysed to detect the elements of the user interface which cause problems.

B. Time-on-task

Time-on-task allows for measuring the time required to complete a specific task. The faster a user can complete a task, the experience is better. In the DEnSeK project the metric serves for evaluating the efficiency of the Dashboard.

C. Efficiency

In contrast to the previous metric, which concerned time, the efficiency metric enables measuring the amount of work required to complete a task. For instance such an effort can be expressed by means of the number of mouse clicks or keystrokes.

D. Errors

This metric allows for detecting improperly designed user interface elements that cause users' confusion. It is measured as the number of user errors when performing a task. Errors may be related to spelling, pressing a wrong key, etc.

E. Learnability

The learnability metric supports examining whether and how user productivity increases with the better knowledge of the system. Measuring learnability requires intense studies spanning a long period of time. For this reason it is often left out.

VII. Conclusion

The metrics described in the paper are used to evaluate Situational Awareness Network (SAN) system developed in the DEnSeK project. The SAN was designed as a three-tier architecture. The lowest tier encompasses a number of sensors for network monitoring. In the middle tier, the SIEM software collects and processes the data from the sensors. Finally, the dashboard on the top tier visualizes information about the threats.

In order to select the appropriate set of metrics a thorough literature analysis was conducted. To the best of authors' knowledge the metrics for SAN have not been proposed so far. Therefore software metrics developed for several related fields, including cyber security, Intrusion Detection and Prevention Systems, SIEM systems, software engineering and testing have been analysed. The study made it possible to derive a set of metrics for testing Situational Awareness Networks.

The selected metrics were divided into three groups. The first group contains metrics related to the evaluation of the testing process, the second – to the effectiveness of threat detection, and the last – to the usability of the dashboard. The metrics are used at each stage of the SAN development. In addition they will be applied during final product evaluation and acceptance process.

References

[1] R. Kyusakov, J. Eliasson, J. Van Deventer, J. Delsing, and R. Cragie, "Emerging energy management standards and technologies - Challenges and application prospects," in *IEEE International Conference on Emerging Technologies and Factory Automation, ETFA*, 2012. doi: 10.1109/ETFA.2012.6489674. ISBN 9781467347372

[2] F. Maturana, R. Staron, K. Loparo, R. Ambre, and D. Carnahan, "Simulation-based environment for modeling distributed agents for smart grid energy management," in *IEEE International Conference on Emerging Technologies and Factory Automation, ETFA 2011*, 2011. doi: 10.1109/ETFA.2011.6059124. ISBN 9781457700187. ISSN 1946-0740

[3] G. Dini and M. Tiloca, "On simulative analysis of attack impact in Wireless Sensor Networks," in *IEEE International Conference on Emerging Technologies and Factory Automation, ETFA*, 2013. doi: 10.1109/ETFA.2013.6648059. ISBN 9781479908622. ISSN 19460740

[4] Y. Aillerie, S. Kayal, J.-p. Mennella, R. Samani, S. Sauty, and L. Schmitt, "Smart Grid Cyber Security," 2013.

[5] Y. Yan, Y. Qian, H. Sharif, and D. Tipper, "A Survey on Cyber Security for Smart Grid Communications," *IEEE Communications Surveys & Tutorials*, vol. 14, no. 4, pp. 998–1010, 2012. doi: 10.1109/SURV.2012.010912.00035. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6141833

[6] W. Wang and Z. Lu, "Cyber security in the Smart Grid: Survey and challenges," *Computer Networks*, vol. 57, no. 5, pp. 1344–1371, apr 2013. doi: 10.1016/j.comnet.2012.12.017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1389128613000042

[7] N. Falliere, L. O. Murchu, and E. Chien, "W32.Stuxnet Dossier," Symantec Security Response, Tech. Rep., 2011.

[8] D. Kushner, "The real story of stuxnet," *IEEE Spectrum*, vol. 50, pp. 48–53, 2013. doi: 10.1109/MSPEC.2013.6471059

[9] P. Shakarian, J. Shakarian, and A. Ruef, *Introduction to Cyber-warfare*. Elsevier, 2013. ISBN 9780124078147. [Online]. Available: http://www.sciencedirect.com/science/article/pii/B9780124078147000087

[10] P. Technologies, "SCADA Safety in Numbers," Tech. Rep., 2012.

[11] N. Virvilis and D. Gritzalis, "The Big Four - What We Did Wrong in Advanced Persistent Threat Detection?" in *2013 International Conference on Availability, Reliability and Security*. IEEE, sep 2013. doi: 10.1109/ARES.2013.32. ISBN 978-0-7695-5008-4 pp. 248–254. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6657248

[12] ICS-CERT, "Alert (ICS-ALERT-14-281-01B) Ongoing Sophisticated Malware Campaign Compromising ICS (Update B)," 2014.

[13] A. Carcano, A. Coletta, M. Guglielmi, M. Masera, I. N. Fovino, and A. Trombetta, "A Multidimensional Critical State Analysis for Detecting Intrusions in SCADA Systems," *Industrial Informatics, IEEE Transactions on*, vol. 7, no. 2, pp. 179–186, 2011. doi: 10.1109/TII.2010.2099234

[14] A. Felkner and A. Kozakiewicz, "More Practical Application of Trust Management Credentials," in *Proceedings of the 2015 Federated Conference on Computer Science and Information Systems*, ser. Annals of Computer Science and Information Systems, M. Ganzha, L. Maciaszek, and M. Paprzycki, Eds., vol. 5. IEEE, 2015. doi: 10.15439/2015F95 pp. 1125–1134. [Online]. Available: http://dx.doi.org/10.15439/2015F95

[15] O. Rysavy, J. Rab, and M. Sveda, "Improving security in SCADA systems through firewall policy analysis," in *Proceedings of the 2013 Federated Conference on Computer Science and Information Systems*, M. P. M. Ganzha L. Maciaszek, Ed. IEEE, 2013, pp. pages 1423–1428.

[16] M. Chakraborty, N. Chaki, and A. Cortesi, "A New Intrusion Prevention System for Protecting Smart Grids from ICMPv6 Vulnerabilities," in *Proceedings of the 2014 Federated Conference on Computer Science and Information Systems*, ser. Annals of Computer Science and Information Systems, M. P. M. Ganzha L. Maciaszek, Ed., vol. 2. IEEE, 2014. doi: 10.15439/2014F287 pp. pages 1539–1547. [Online]. Available: http://dx.doi.org/10.15439/2014F287

[17] A. Bialas, "Experimentation tool for critical infrastructures risk management," in *Proceedings of the 2015 Federated Conference on Computer Science and Information Systems*, ser. Annals of Computer Science and Information Systems, M. Ganzha, L. Maciaszek, and M. Paprzycki, Eds., vol. 5. IEEE, 2015. doi: 10.15439/2015F77 pp. 1099–1106. [Online]. Available: http://dx.doi.org/10.15439/2015F77

[18] H. Khurana, M. Hadley, and D. Frincke, "Smart-grid security issues," *IEEE Security & Privacy Magazine*, vol. 8, no. 1, pp. 81–85, jan 2010. doi: 10.1109/MSP.2010.49. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5403159

[19] "DEnSeK (Distributed Energy Security Knowledge) - project website." [Online]. Available: http://www.densek.eu/

[20] M. R. Endsley and D. J. Garland, *Situation Awareness Analysis and Measurement*. CRC Press, Inc., 2000.

[21] "Snort Home Page." [Online]. Available: http://www.snort.org/

[22] Z. Zhou, "The study on network intrusion detection system of Snort," in *2010 International Conference on Networking and Digital Society*, vol. 2. IEEE, may 2010. doi: 10.1109/ICNDS.2010.5479341. ISBN 978-1-4244-5162-3 pp. 194–196.

[23] OISF, "Suricata - Open Source IDS / IPS / NSM engine." [Online]. Available: http://suricata-ids.org/

[24] "The Bro Network Security Monitor," 2016. [Online]. Available: https://www.bro.org/

[25] G. K. Varadarajan, "Web Application Attack Analysis Using Bro IDS," 2012. [Online]. Available: http://www.sans.org/reading-room/whitepapers/detection/web-application-attack-analysis-bro-ids-34042

[26] "Argos," 2016. [Online]. Available: https://sourceforge.net/projects/argos-scada-en/

[27] "Apache Kafka: a high-throughput distributed messaging system," 2016. [Online]. Available: http://kafka.apache.org/

[28] "Druid." [Online]. Available: http://druid.io/

[29] A. September, "IEEE Standard Glossary of Software Engineering Terminology/IEEE Std 610.12-1990," p. 96, 1990. [Online]. Available: http://www.amazon.com/Standard-Glossary-Engineering-Terminology-610-12-1990/dp/155937067X

[30] J. Radatz, A. Geraci, and F. Katki, "IEEE standard glossary of software engineering terminology," *IEEE Std*, vol. 610121990, p. 121990, 1990.

[31] I. Burnstein, T. Suwanassart, and R. Carlson, "Developing a testing maturity model for sofiware test process evaluation," in *Test Conference, 1996*, 1996. ISBN 0780335406 pp. 581–589.

[32] R. Leszczyna, R. Małkowski, and M. R. Wróbel, "Testing Situation Awareness Network for the Electrical Power Infrastructure," *Acta Energetica*, vol. 1, pp. 270–276, 2015.

[33] P. Goodman, *The Practical Implementation of Software Metrics*. McGraw-Hill, Inc., 1993.

[34] S. Quadri and S. Farooq, "Notable Metrics in Software Testing," *5th National Conference on Computing For Nation Development - INDIACom-2011*, pp. 273–276, 2011.

[35] A. Kaur, B. Suri, and A. Sharma, "Software testing product metrics-A Survey," in *National Conference on Challenges & Opportunities in Information Technology*, 2007, pp. 1–6.

[36] A. Jaquith, *Security Metrics, Replacing Fear, Uncertainty, and Doubt*. Addison-Wesley Professional, 2007.

[37] Y. Chen, R. L. Probert, and K. Robeson, "Effective test metrics for test strategy evolution," pp. 111–123, 2004.

[38] D. Kang, D. Fuller, and V. Honavar, "Learning classifiers for misuse and anomaly detection using a bag of system calls representation," *Sixth Annual IEEE SMC Information Assurance Workshop*, 2005. doi: 10.1109/IAW.2005.1495942

[39] J. L. Bayuk and A. Mostashari, "Measuring cyber security in intelligent urban infrastructure systems," in *2011 8th International Conference & Expo on Emerging Technologies for a Smarter World*. Ieee, nov 2011. doi: 10.1109/CEWIT.2011.6135873. ISBN 978-1-4577-1591-4 pp. 1–6.

[40] W. Albert and T. Tullis, *Measuring the user experience: collecting, analyzing, and presenting usability metrics*. Newnes, 2013. ISBN 9780124157811

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 751–755
DOI: 10.15439/2016F481
ACSIS, Vol. 8. ISSN 2300-5963

Comparison of Network Architectures for a Telemetry System in the Solar Car Project

Cody R. Barnes, Ethan G. Toney, Jerzy W. Jaromczyk
University of Kentucky
Department of Computer Science
Lexington, KY 40506, USA
Email: cody.barnes@uky.edu, egto222@uky.edu, jurek@cs.uky.edu

***Abstract*—A solar car is an electric vehicle that runs entirely on solar energy. Designing, building and racing solar cars has been a longstanding worldwide challenge for engineering and computer science students, with the overarching goal being to design devices that use sustainable energy sources. This article describes our experience and educational outcomes (*the modeling and design of computer-based systems in a way that demonstrates comprehension of the trade-offs involved in design choice*) attained while designing the network architecture for a solar car project.**

The computer science members of the University of Kentucky Solar Car Strategy Team are tasked to reliably collect and analyze car data in real-time, both to assist in the car development process, and then to provide important sensor readings to the driver during racing. The challenges in designing the architecture and protocols for the computer system that supports the solar car are to ensure that: (1) energy consumption is minimal, (2) data collection is reliable, (3) the network system is secure, and (4) the implementation of the system is not overly complex.

Our computer system supporting the telemetry tasks uses three micro-controllers to collect and send data over serial communications to a master micro-controller (the Raspberry Pi), that parses and stores data in an on-board database.

We compare three protocols: a simple USB-based protocol, and two protocols used in traditional non-solar cars: CAN and Ethernet. We analyze (1) their energy consumption over a period of time, (2) their reliability (by performing stress tests such as disconnecting devices and driving over bumpy terrain), (3) their security (by attempting to compromise the system by remotely sending data over communication lines), and (4) their complexity in terms of time and effort for implementation and development.

I. Introduction

SOLAR Car racing (see Figure 1), the competitive racing of fully electric vehicles using solar energy, has existed since 1985. Universities and businesses around the globe participate with different goals in mind. Universities typically participate in order to improve engineering and technical knowledge and skills of the students, while businesses typically participate to develop renewable energy technologies. Traditionally, a solar car requires a collaboration of electrical, mechanical, and computer engineers (see [6]). However, with data storage becoming cheaper, and micro-controllers becoming more abundant and inexpensive (such as the Arduino and Raspberry Pi) these make programming embedded systems just as easy as programming an application for a typical computer. In other words there is a growing need for computer science students to be a part of the team. Solar car teams can easily capture, store, and analyze data in order to improve the overall performance of the car.

In the context of power consumption, we analyzed a telemetry system for the solar car project. An automated communications process collected and transmitted car performance data to receiving equipment and personnel for monitoring, processing, testing and decision making. When choosing the network architecture to support telemetry to achieve the aforementioned capturing, storing, and analysis of data, we must determine trade-offs that are critical to the car's functionality. It is also important to carefully consider and design the underlying software architecture that is the driving force behind the network architecture.

Minimizing the power consumption, and comparing various solutions and their trade-offs, are some of the contemporary challenges, addressed in many projects, including the past workshops of this conference. See for example, [1], [2], [4], and [5].

This paper reports on experimental results of a student team, for whom the Solar Car project serves as an attractive way of learning engineering topics. Clearly, tools, instruments, and methods are unequal to sophisticated, advanced, and likely expensive telemetry systems developed for professional motor racing, such as in Formula One. However, the experience with a real-life project of designing the network architecture for a solar car project, contributes to the important computer science student outcome: *the modeling and design of computer-based systems in a way that demonstrates comprehension of the trade-offs involved in design choice*. See also [3] for a discussion on using real-life projects for "developing skills needed for the proper formulation of system visions and requirements specifications."

The subsequent sections describe and compare three different network solutions, the architecture of our telemetry system, and finally, provide our conclusions.

II. Network Comparison

We now discuss the trade-offs of using USB [7], CAN [8], and Ethernet while trying to meet these four challenges: low power-consumption, reliability, security, and simplicity.

A. Low Power-Consumption

Low power-consumption is by far one of the most critical requirements in the development of a solar car. Naturally, the

Fig. 1. Solar Car – team UK

architectural design of the network was not exempt. Since we are racing the car, keeping the power consumption down allows us to allocate more power towards other critical parts of the car. Saving this power for the motor has the potential to result in more mileage out of the total amount of power. We hypothesized that USB and CAN would require about the same amount of power and that Ethernet would require much more than the previous two. Consequently, we chose to compare USB and CAN first and then talk about Ethernet afterwards.

1) USB versus CAN: USB and CAN are known to be very low power networks as long as they are not being used to power a device. To accurately compare these two approaches, we chose to measure the amount of power required to power a CAN chip and a USB chip with no data being sent over them. From experiments, we found that the CAN chip required 11% less power than the USB chip. Although 11% may appear big, in practice the difference is negligible due to the measured values being low.

2) Ethernet: Compared to USB and CAN, Ethernet is known to use significantly more power. This is because of larger hardware requirements to allow for an Ethernet network. Not only would one need to put a router or switch on the car, one would also have to add more hardware on the actual micro-controllers to be able to communicate with Ethernet.

B. Reliability

The reliability of each of the protocols is important because it ultimately determines whether or not data are received. When discussing the reliability of the three network architectures, we focus on the hardware aspects.

1) USB: USB provides data integrity within the cables, meaning that there is a high degree of confidence that the data sent over the transmitting end will make it to the receiving end. Any unrecoverable errors will likely be noticed and reported to the appropriate end of the cable. Data integrity is provided through USB's self-recovery system. This approach guarantees that a message will be resent at least three times before reporting an error to the client software, and will throw time-outs for lost or invalid packets.

Unrelated to the actual data integrity within the cable, but still considered in our evaluation, is the observation that USB has a high likelihood of becoming mechanically disconnected from its transmitting or receiving end due to jostling that occurs while the car is in motion.

2) CAN: CAN is similar to USB in that it also provides data integrity. The CAN protocol defines no less than five different ways to detect error:

- Cyclic Redundancy Check (CRC) acts similar to a checksum by doing polynomial division on the bits and comparing the end result to the 15 bit CRC field located within the packet.
- Acknowledgment (ACK) Check - The node that transmitted a message has essentially sent a recessive level and would expect to receive a dominant ACK message. If it does not, then it acts as if the previously sent message was lost and responds accordingly.
- Form Error Check - If there is a dominant bit in the CRC field delimiter, ACK field delimiter, or EOF (end of frame) then the message is re-sent because, per protocol definitions, there must not be a dominant bit in these fields.
- Bit Stuffing - If six consecutive bits with the same polarity occur between the SOF (start of frame), then the CRC field throws an error. The EOF field should be the only field with six consecutive bits of the same polarity.
- Bit error occurs when the transmitter reads a signal that is opposite of what it sent except during arbitration and in the ACK field. Arbitration is a method that provides a bitwise losslessness if all of the nodes on the CAN network are synchronized to allow for every bit on the CAN network to be sampled at the same time.

In contrast to USB, the few devices on the car that utilize CAN have never had any issues with disconnectivity due to the jostling of wires.

3) Ethernet: Ethernet in itself is not reliable. It does not support retransmission, it does not provide acknowledgment of successful frame delivery, and if a frame were to become corrupt it simply drops it without letting the transmitting or receiving end know. Due to this characteristic, we consider the use of Ethernet with TCP, Transmission Control Protocol. TCP is a protocol that is often overlaid on top of Ethernet. TCP supports detection of duplicate data, retransmission, and sequencing. Because Ethernet is usually used to send packets over long distances, and because multiple Ethernet switches are used, packets can be duplicated. TCP detects duplicate packets and drops the unnecessary ones. In conjunction with this duplication, packets can get out of order. Therefore TCP

supports sequencing - placing packets in the order that they were sent. Additionally, if a packet gets lost or corrupted, TCP provides retransmission of the affected packet.

C. Security

The UK solar car was designed just for races, and not intended for mass production. Nevertheless, network security is becoming a general concern and should always be accounted for. Especially considering the recent cyber-security incidents related to private cars being hacked ([10]). Even though our specific network application represents a minimal security threat, security was not ignored.

1) USB: From a security standpoint USB is actually a very good option. It only allows for peer-to-peer communication, so there is essentially no network to abuse. However, due to the nature of peer-to-peer, there is no form of authentication. Simply, each computer does not have a way of being completely sure who it is talking to. Given this situation, the only realistic way to attack a USB-based system would be to plug in directly to the device you wanted to lie to.

2) CAN: A CAN network is riddled with security problems stemming from the underlying protocol. Much like the IP protocol, it relies on the sender to accurately choose a message ID. From this message ID you can tell what kind of data are being sent, as well as what device is sending it. This is done by a device choosing a range of IDs, and making sure that every message it sends is inside of this range. Given this type of protocol, it is trivial to impersonate another device on the network. A person could just send a message onto the network with an ID that is not their own, and the receiving device would not know the difference. As a proof of concept, we simulated this vulnaribility and were able to get the Raspberry Pi on our CAN network to tell the motor to accelerate even though the pedal was not being pressed. It is understandable why the developers of this network originally left this vulnearibility, because CAN was always intended to be an internal network. Getting on to the network to abuse its security faults is not easy. This vulnerability is also why there has been so much news about cars being hacked lately; see [10]. Since CAN is the industry standard, it is not too difficult for a hacker to use a laptop with a CAN cord and plug into the network manually to break in. In some cases, where the car has an Internet access, there may be ways to get from the Internet network to the CAN network and remotely hack a car.

3) Ethernet: Ethernet has almost the same set of vulnerabilities as CAN. However, it is substantially easier to abuse a CAN network than it is to abuse the security vulnerabilities in Ethernet. Numerous protocols built on top of Ethernet reduce its vulnerability. Clearly, it is still possible for hackers to abuse Ethernet network by misrepresenting the identity and impersonating devices on the network.

D. Simplicity

Simplicity, a universal value and software development practice, is commonly used by student teams. These teams operate on schedules dictated by the academic calendar. In addition, once they graduate, the code base is picked up by future Computer Science students to continue assisting the Solar Car project; see [11] for the repository with our code.

1) USB: On the software side, USB is relatively simple. The majority of languages that would be commonly used in these type of applications (Python, Java, C, C++, etc.) support serial communications. One must simply identify the port to communicate with and open a serial connection with said port.

On the hardware side, USB actually is not native to the majority of the solar car peripherals except for the micro-controllers themselves. As a result, USB use can cause problems for the developers with the devices on the car that communicate through CAN. To alleviate this difficulty, the team utilizes CAN-to-USB converters in order to receive data.

2) CAN: Even though, communicating with CAN through software is not as common, it can be done in a relatively standard way. In fact, communication support is implemented in `Python`, the main language used on the Raspberry Pi. However, CAN is more complex to use rather than USB or Ethernet. In fact, even the documentation about the implementation through `python-can` library suggests that working with `python-can` is much more complex than the implementation of USB through the `pyserial` library. Thus, in our solution, we actually use a CAN to USB converter between the CAN network and the Raspberry Pi, to keep the software implementation as simple as possible.

The hardware side for CAN is the upside, however, with CAN being the industry standard for regular cars due to its reliability, upgradeability, performance, and cost. As the benefit of standardization, one can expect that the majority of the car peripherals communicate over CAN. Among them is the Motor Controller that drives the rear third wheel of the solar car.

3) Ethernet: Ethernet is very similar to USB when it comes to software aspects. Almost all languages have some kind of built in library that handles communication over sockets, which are the Ethernet equivalent to ports in USB. This gives Ethernet a distinct advantage over CAN when it comes to software implementations.

For the hardware side, Ethernet is in the same category as USB. Ethernet is not common among any of the car peripherals besides the micro-controllers. In effect, Ethernet switches and CAN to Ethernet converters have to be used.

Table I summarizes the above evaluation.

TABLE I
SUMMARY OF THE COMPARISON FOR THE SOLAR CAR PROJECT

	Power usage	Reliability	Security	Simplicity
USB	low	good	good	poor hardware support
CAN	low	good	poor	nontrivial software support
Ethernet	high	good (with TCP)	poor	poor hardware support

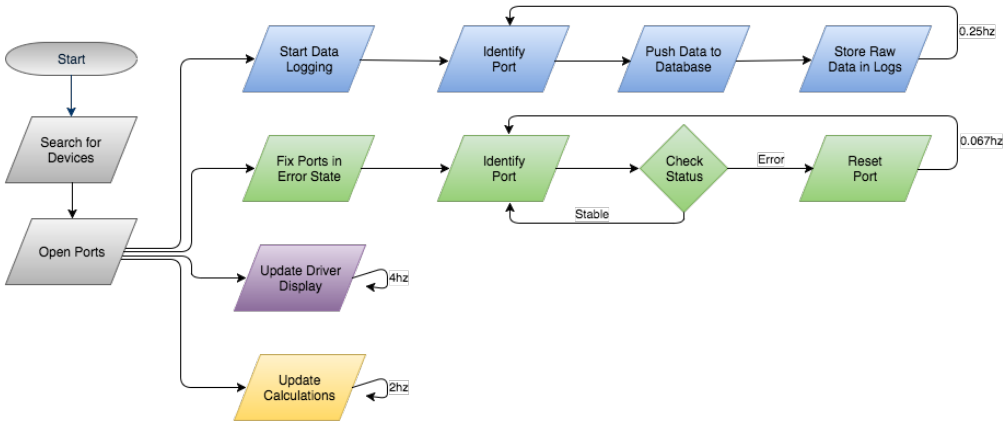


Fig. 2. High-level overview of the software architecture

III. Software Architecture

There are multiple choices for the network architecture and the final decision on whether or not CAN, USB, or Ethernet would be better choice for us is greatly influenced by the current software architecture.

Depicted in Figure 2 is a high-level overview of our software architecture. When the program is started on boot, it immediately opens the ports where data are expected, and then connects to the corresponding devices for those ports.

Rules in UDEV - a device manager for the Linux kernel - are used so that each device is assigned a unique identifier based off of the product id on the USB chip.

The flow of control is as follows. The program starts off by initially searching for any connected devices and opening up a connection to them. After this point four threads are started up and read from a global state containing of a list with each device. The threads run concurrently, and access the global state, but do not communicate directly with one another. The first thread, *the logging thread*, starts off by ensuring that the device has been identified. Then it logs the device buffer to the database on the car, and stores the raw data logs for the *post mortem* analysis.

The second thread, *error correcting thread*, iterates through all of the devices and identifies their kind. Then, this thread looks at the statistics of each device to determine whether the device is in an error state.

The third thread, *update drive display*, sends the vehicle speed and battery current measurements to the attached Arduino. The Arduino then takes these values and displays them on two seven segment displays.

The last thread, *update calculations*, does the calculations that would otherwise require too much bandwidth to send the sufficient amount of information for the other end of the telemetry system to calculate.

The responsibilities of the threads described above, and presented in Figure 2, are:

1) Thread Data Logging: This thread loops through every device (port), logs any data that have been saved up in the device buffer, pushes the data to the SQL database and writes the data to raw text files as a backup.
2) Thread - Port Fixing: This thread also loops through every device, but instead of logging data it checking to ensure that the device is still connected and that we are receiving data. If the device is in an error state then this thread will attempt to fix it until it attains a stable state again. This thread is also the only thread with the capabilities to significantly change the global state, so its actions are heavily synchronized with the other threads.
3) Thread - Updating Driver Display: This thread updates the driver display with the current car speed and voltage usage.
4) Thread - Update Calculations: This thread updates any calculations that are being sent over the radio telemetry system in the trail (chase) car. As an alternative solution, these calculations could be moved to the Java application on the receiving end to lessen the amount of work that the program on the Raspberry Pi has to do. The tradeoff would be in using more bandwidth to send the data.

The diagram in Figure 3 shows the general layout for the current USB based telemetry system for our solar car. There are multiple devices that are plugged into a powered hub: two battery boxes, the motor CAN network, an Arduino, and a telemetry box. Also you may note that the Arduino controls many devices: two seven-segment displays (used to show speed and current), gyro, accelerometer, and GPS. These extra devices data are eventually sent on to the Raspberry PI for processing. There is also a small CAN network between the drivers control box and motor controller that must be there regardless of the choice of architecture. Regarding the powered hub, on the other end there is a Raspberry PI micro-computer that uses this powered hub to create separate connections to each device and manage each of them individually. The micro-computer then reads and interprets data from each device, and then sends them to the telemetry box; a matching telemetry box in the chase car receives these data. During the interpretation of data on the Raspberry

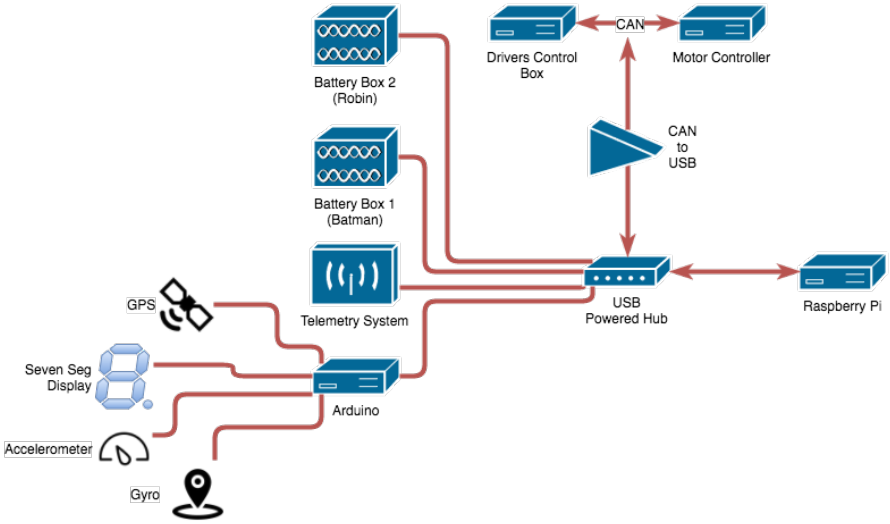


Fig. 3. Architecture of our telemetry system

Pi that data are permanently stored in an on-board MySQL database.

IV. Conclusion

By analyzing various network solutions, we have determined trade-offs in Solar Car network architectures that try to meet the four challenges: low power consumption, reliability, security, and simplicity. In comparing these solutions, we have found that USB and Ethernet are very respectable options. USB provides a quick starting point. For example, if a project involves a micro-controller then it is very likely that the micro-controller contains, at the least, a USB port. Furthermore, communications over a serial connection, which is what USB is, are very easy to implement. On the other hand, Ethernet, being a class D network, can handle very high data rates (up to 100 mb/s). If the Solar car needed to transfer this much data then Ethernet high energy usage could be countered by the trade-off for high data rates and easy implementation. Even with USB and Ethernet being respectable options, CAN still has been found to be the best overall option. The hardware is easy to implement, cheap, and upgradeable. It uses a low amount of energy, which is highly critical to our application, and although it is not as common to implement in software, there exists workarounds. Working on a solar car design, along with implementing and testing solutions, has provided us with incomparable learning outcomes in networking, security and teamwork, and have allowed all the participating students to clearly see design trade-offs, even when some of them are subtle.

References

[1] R. Banach, P. Van Schaik, and E. Verhulst. Simulation and formal modelling of yaw control in a drive-by-wire application. In M. Ganzha, L. Maciaszek, and M. Paprzycki, editors, *Proceedings of the 2015 Federated Conference on Computer Science and Information Systems*, volume 5 of *Annals of Computer Science and Information Systems*, pages 731–742. IEEE, 2015.

[2] M. Eshaftri, A. Al-Dubai, I. Romdhani, and M. Bani Yassein. A new energy efficient cluster based protocol for wireless sensor networks. In M. Ganzha, L. Maciaszek, and M. Paprzycki, editors, *Proceedings of the 2015 Federated Conference on Computer Science and Information Systems*, volume 5 of *Annals of Computer Science and Information Systems*, pages 1209–1214. IEEE, 2015.

[3] J. Král and M. Žemlička. Experience with real-life students' projects. In M. Paprzycki M. Ganzha, L. Maciaszek, editor, *Proceedings of the 2014 Federated Conference on Computer Science and Information Systems*, volume 2 of *Annals of Computer Science and Information Systems*, pages pages 827–833. IEEE, 2014.

[4] C. Panait and D. Dragomir. Measuring the performance and energy consumption of AES in wireless sensor networks. In M. Ganzha, L. Maciaszek, and M. Paprzycki, editors, *Proceedings of the 2015 Federated Conference on Computer Science and Information Systems*, volume 5 of *Annals of Computer Science and Information Systems*, pages 1261–1266. IEEE, 2015.

[5] T. Szydlo, P. Nawrocki, R. Brzoza-Woch, and K. Zieliński. Power aware MOM for telemetry-oriented applications using gprs-enabled embedded devices - levee monitoring use case. In M. Paprzycki M. Ganzha, L. Maciaszek, editor, *Proceedings of the 2014 Federated Conference on Computer Science and Information Systems*, volume 2 of *Annals of Computer Science and Information Systems*, pages pages 1059–1064. IEEE, 2014.

[6] R. Mangu, K. Prayaga, B. Nadimpally and S. Nicaise, Design, Development and Optimization of Highly Efficient Solar Cars: Gato del Sol I-IV, in *2010 IEEE Green Technologies Conference*, Grapevine, TX, 2010, pp. 1-6. doi: 10.1109/GREEN.2010.5453800, http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5453800&isnumber=5453775

[7] USB Specifications http://www.usb.org/developers/docs/. Online; accessed 9-May-2016

[8] CAN Specifications http://www.usb.org/developers/docs/. Online; accessed 9-May-2016.

[9] M. Faezipour, N. Faezipour, M. Adnan, S. Adnan, S. Addepall., Progress and challenges in intelligent vehicle area networks, in *Communications of the ACM*, 55 (2), 2012, pp. 90–100,

[10] J. Vanian. Hacking Cars Is Easy, http://fortune.com/2016/01/26/security-experts-hack-cars/, Online: accessed 10-July-2016.

[11] E. Toney. https://github.com/KentuckySolarCar/RaspberryPi

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 757–766
DOI: 10.15439/2016F24
ACSIS, Vol. 8. ISSN 2300-5963

Method for Approaching the Cyber-Physical Systems

Tiberiu S. Letia
Dept. of Automation
Technical University of Cluj-Napoca
Romania
Email: Tiberiu.Letia@aut.utcluj.ro

Attila O. Kilyen
Dept. of Automation
Technical University of Cluj-Napoca
Romania
Email: ors.kilyen@accenture.com

***Abstract*—The main topics of the Cyber-Physical Systems (CPSs) cover the specification, modeling, control, design, verification and testing. The CPSs implementation consists of reactive programs conceived using models that are capable to sustain the mentioned activities. Component diagrams (introduced by Unified Modeling Language) are used here for the architecture design, with the goal to split the CPS complexity into smaller entities that are easier to tackle. All the components are modeled by Fuzzy Logic Enhanced Time Petri Nets (FLETPNs) that can simultaneously describe the discrete event and the time discrete features. This unique and compact approach facilitates the control synthesis, the software design, the verification and the testing.**

An example of application to the control of a system composed of a wind turbine generator, a photo-voltaic generator and loads is used to show the model utilization and its benefits.

I. Approaches of the Cyber-Physical Systems

CPSs integrate the dynamics of the physical processes with those of the software and communication. There are some surveys that present the main characteristics, the main domains where they are applied and the main topics of the CPSs [10], [11], [12]. The main topics of the CPSs cover the specification, the modeling, the control, the design, the verification and the testing. The CPSs implementation consists of reactive programs that are based on models that are capable to support the mentioned activities.

The main goal of the current research is to conceive a control system that concurrently reacts to discrete events and continuous modifications of plant state. The target is a set of interacting dynamic models capable to approach the following specification:

- the reaction to synchronous and asynchronous (plant) events that are signaled by continuous variables (instead of single level events)
- the continuous time reaction to modification of some (plant output) variables
- the reaction control signals that belong to continuous domains (the discrete domains, as the binary set, should be particular cases)

Some reactions require the execution of activities involving non-ignorable durations and could have real-time constraints that have to be fulfilled. This requirement leads to the conclusion that the target model has to be capable to describe concurrent behavior.

A relevant issue is to conceive a model that is capable to describe the controller behavior and its structure. A practical goal is to make possible the verification that the implemented model fulfills the specified requirements.

The implementation of controllers on digital computers supposes that the information of continuous variables can be represented with a limited and tolerated accuracy (due to the limited length of the number representation) and the calculus can add other losses of the precision. On the other hand, the continuous time reactions are not possible to be implemented on digital computers. For this reason, instead of continuous time models, the discrete time models are used. The loss of accuracy due to the conversion of the continuous time models into discrete time models is supposed to be tolerated too.

The OMG (Object Management Group) Unified Modeling Language could successfully fulfill the requirements using a set of state machines, but these dynamic models have to be endowed with many variables, equations and condition expressions to completely describe the desired behavior [1]. The verification that the obtained models fulfill the requirements needs the use of other complex methods (such as different kinds of Petri nets) or simulation tools.

Many authors emphasize that CPSs are hybrid systems [2], [10], [11]. A hybrid system is composed of a discrete event side and a continuous time side in an interaction that provides a complex behavior. The control of a hybrid system is a challenge due to the requirements of asynchronous reactions to the discrete events as well as to the continuous adjustment of some controlled outputs. The CPSs involve interdependencies between physical behavior and digital control [13]. The control system implementation should be based on asynchronous interrupts and synchronous discrete time reactions.

The controller asynchronous reaction involves the execution of rules of the form:

$$ON\ event\ IF\ condition\ THEN$$
$$action_1 \wedge action_2 \wedge \cdots \wedge action_k \quad (1)$$

The ordinary Petri nets can model the handling of events, the binary conditions, the concurrency and the controller structure. These models are not capable to model the cases when the

978-83-60810-90-3/$25.00

involved reactions require input of continuous variables and outputs that signal continuous variables. These models are not appropriate to model continuous type operations.

The fuzzy logic controller (FLC) based on fuzzy logic provides a means of converting a linguistic control strategy based on expert knowledge into automatic control strategies. This approach was chosen (in the current research) for its capability to conceive controllers that tackle, beside the synchronous reactions (i. e. the periodic discrete time feature), the asynchronous reactions for the cases that require variable output control signals.

An overview of the possibility of implementing the fuzzy control systems as fuzzy rule-base systems is contained in [6]. Here it is justified that the conventional methods are good for simpler problems, while the fuzzy systems are suitable for complex problems or control applications that involve human descriptions or intuitive thinking. Lee presents a survey of the general methodology for constructing an FLC and the assessing of its performance [7].

In [9] another model that links the Petri nets with FLC is introduced.

II. Fuzzy Logic Enhanced Time Petri Net Models

A. Low Level Petri Nets

As it is well known, a Petri Net (PN) is a directed graph with two kinds of nodes. An *ordinary PN* is a 5-tuple

$$PN = (P, T, pre, post, \mathbf{M}) \tag{2}$$

with:

- a finite place set $P = \{p_1, p_2, ..., p_m\}, (m \geq 0)$
- a finite transition set $T = \{t_1, t_2, ..., t_n\}, (n \geq 0)$
- $pre : P \times T \rightarrow \mathbf{N}$ (natural number set) is the backward incidence function:
- $post : P \times T \rightarrow \mathbf{N}$ is the forward incidence function

In the current approach

- $pre(p,t) = 0$, if there is not an arc from p to t and $pre(p,t) = 1$, if there is an arc from p to t,
- $post(p,t) = 0$, if there is not an arc from t to p and $post(p,t) = 1$, if there is an arc from t to p.

$N = (P, T, pre, post)$ describes the structure without marking. $PN = (N, \mathbf{M}_0)$ is the structure with a marking $\mathbf{M}$ where $M : P \rightarrow \mathbf{N}$ is the marking specifying the number of tokens of each place. The marking $\mathbf{M} = [M(p_1), M(p_2), ..., M(p_m)]^T$ describes the PN state.

The lack of the PN capability to handle the time is removed in the models *Time Petri Nets (TPNs)*. The TPNs are suited for modeling the time-dependent systems with timing constraints [3] A timed Petri net can be defined with delayed transitions, or delayed tokens [4],[5]. The current approach uses the timed transitions. A TPN is a PN with each transition t_i delayed by an assigned delay d_i from a set of non-negative integers $D = \{d_1, d_2, ..., d_n\}$. That means, each transition t_i is delayed with d_i time units from the moment of time when it becomes enabled.

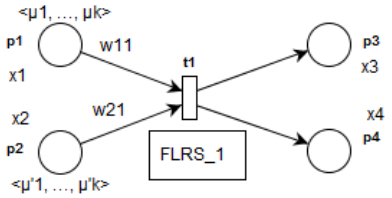


Fig. 1. Example of a FLETPN

The definition of TPN is:

$$TPN = (P, T, pre, post, D, \mathbf{M}) \tag{3}$$

where $P, T, pre, post$ and $\mathbf{M}$ have the previous meanings. $D : T \rightarrow \mathbf{N}$ is a mapping that assigns to each transition a delay.

An Enhanced Time Petri Net (ETPN) is a TPN endowed with an input place set *Inp* and an output place set *Out* [5]. In ETPN only the transitions with single input places can be delayed. The input places (Inp) are loaded with tokens by the plant. The ETPN injects tokens in the output places (Out) and these tokens are extracted immediately by the plant.

All these kinds of Petri nets have a single type of tokens.

B. High Level Petri Nets

Unlike the above defined Petri nets, the high level Petri nets have distinct tokens. The current approach is based on a particular case of high level Petri nets. There are some kinds of Petri Nets endowed with fuzzy features. A relevant review of fuzzy Petri nets and industrial applications can be found in [8].

A FLETPN is an ETPN extended with fuzzy logic rules that is capable of processing fuzzy information. Each place has a distinct token and its capacity is equal to one. Each place of the ETPN is assigned a variable and each transition has assigned a fuzzy logic rule set, but one fuzzy logic rule set could be assigned to more than one transition. A token injected into a place expresses the membership degrees of the (assigned) variable to the fuzzy sets. Figure 1 shows a FLETPN that has a transition with two input places and two output places. Each place p_i has assigned a variable x_i.

The definition of a FLETPN is:

$$FLETPN = (P, T, pre, post, D, W, X, EFS, \mathbf{FLRS}, \alpha, \beta, \mathbf{M}) \tag{4}$$

where $P, T, pre, post$ and D have the previous meanings. $X = \{x_1, x_2, \cdots, x_m\}$ is a set of variables with $x_i \in \mathbf{R}$ (with $\mathbf{R}$ a domain in the real number set). α is a bijective mapping $\alpha : P \rightarrow X$ that assigns to each place a variable from the set X. EFS is an extended fuzzy set of the fuzzy set $FS = \{A_1, A_2, \cdots, A_k\}$, $EFS = FS \bigcup \{\Phi\}$. The statement $x\, is\, \Phi$ means there is no information about the value of the variable x at the current moment of time.

The marking $M(p_i)$ of a place p_i is the vector of the membership degrees of the assigned variable to the fuzzy set

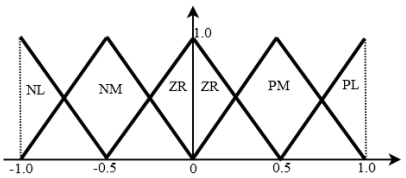


Fig. 2. Fuzzy logic membership functions.

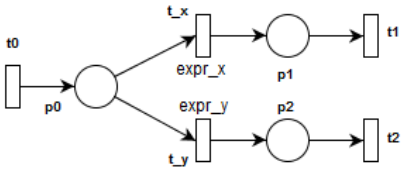


Fig. 3. Selection by expressions.

TABLE I
Fuzzy logic rules $FLRSx, y$ in Fig. 4

NL	NM	ZR	PM	PL
Φ,ZR	Φ,ZR	PL, PL	ZR,Φ	ZR,Φ

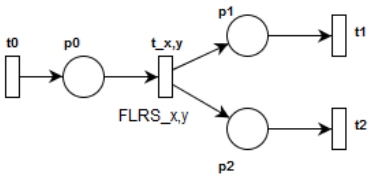


Fig. 4. Selection by FLRS.

FS. Any distinct token of the form

$$\mu =< \mu_1, \mu_2, \cdots, \mu_k > \tag{5}$$

inserted into a place p_i expresses the membership degree of the variable x_i to the fuzzy set FS. In the current case, a place corresponds to a set of statements and the information is available only when a token μ is contained.

Each input arc of a transition is endowed with a weighting coefficient: $W : P \times T \rightarrow \mathbf{R}$ (with **R** a domain in the real number set), $W(p_i, t_j) = w_{ij} \in \mathbf{R}$.

FLRS is a set of fuzzy rule sets. β is a mapping that assigns to each transition a fuzzy logic rule set $\beta : T \rightarrow$ **FLRS**. The fuzzy logic rules considered here have the form:

$$IF\ x_1 isA_1 \wedge x_2 isA_2 \wedge \cdots \wedge x_k isA_k\ THEN \\ x'_1 isA_1 \wedge x'_2 isA_2 \wedge \cdots \wedge x'_k isA_k \tag{6}$$

with $x'_1, x'_2, \cdots, x'_k$ belonging to the same set X and representing the consequences of the inference rules.

An example is $FS = \{NL, NM, ZR, PM, PL\}$ where the elements mean negative large, negative medium, zero, positive medium and positive large respectively. For simplicity reasons, the membership functions used for fuzzification and defuzzification are those presented in Figure 2. For practical reasons the values of the variables $x_i \in X$ were bounded to the real number set [-1,1].

An example of rule using FS is:

$$IF\ x_1 isZR \wedge x_2 isNM\ THEN\ x_3 isPM \wedge x_4 isPL \tag{7}$$

In an earlier release of FLETPN model (see [9]) the selection of alternatives was implemented based on logical expressions assigned to transitions. For example, the selection to continue the execution with transitions t_1 or t_2 included in the partial FLETPN model represented in Figure 3 was chosen using the expressions $expr_x$ and $expr_y$. In the current release the logical expressions were removed and the selection is performed by appropriately conceiving the $FLRS_{x,y}$, as shown in Figure 4 and in Table I.

Supposing that all the rules have the same two inputs and two outputs (i. e. consequences) the fuzzy logic rule set can be described in a table such as Table II. There are represented the following rules:

$$IF\ x_1 isNL \wedge x_2 isNL\ THEN\ x_3 isNL \tag{8}$$

$$IF\ x_1 isPL \wedge x_2 isPL\ THEN\ x_4 isPM \tag{9}$$

The consequence of rule (7) injects a token into the place p_3 and another one into the place p_4. The consequence of rule (8) means that if only this rule is activated, the execution of the transition leads to a token in the place p_3 and no token in the place p_4. Unlike rule (8), rule (9) leads to a token in the place p_4 and no token in the place p_3. This manner allows the selection to continue the execution from the place p_3, the place p_4 or from the both of them.

An input x_i can belong to the fuzzy set A_j with a membership degree $\mu_j(x_i)$. For the given example, if the variable x_i is assigned to a place p_i a token injected into this place can be $< \mu_{NL}, \mu_{NM}, \mu_{ZR}, \mu_{PM}, \mu_{PL} >$ and it describes the membership degree of the variable x_i to the fuzzy set FS. All the rules included into a fuzzy rule set have the same inputs and outputs. The dimension (cardinal) of a fuzzy rule set of a transition t_i is $|FS|^l$ where $|FS|$ is the cardinal of the fuzzy set and l the number of the input places of the current transition.

The fuzzy rule set provides an output vector of the dimension equal to the cardinal of the transition output place set. The elements of this output vector are fuzzy sets.

The *execution of an enabled transition* t_j involves:

- the extracting of the tokens from the transition input places (denoted by ${}^o t_i$);

TABLE II
Example of inference rules

$x_1 \backslash x_2$	**NL**	**NM**	**ZR**	**PM**	**PL**
NL	NL, ϕ	NL, Φ	NL,PL	ZR,PL	PL,PL
NM	NM, Φ	PL, Φ	PL,PL	PL,PL	NL,PL
ZR	PM,PL	PM,PL	ZR,PL	NL,PL	PM,PL
PM	ZR,PL	PL,PL	NL,PL	ϕ, NM	ϕ, PL
PL	NL,PL	ZR,PL	ZR,PL	ϕ, PM	ϕ, PM

- the defuzzification of all input variables x_i;
- the multiplication of the variables with the corresponding weighting coefficients $x'_{ij} = w_{ij}x_i$;
- the fuzzification of the variables x'_{ij};
- the use of the FLRS with x'_{ij} as inputs;
- the normalization operation that reduces the previous consequences to a single one and leads to injection of a single token into the output places;
- the injecting of the resulted tokens into the transition output places (denoted by t_i^o) when the delay elapses.

Due to the fact that a variable number of rules can be involved (activated) for a transition execution, this could inject a variable number of tokens into the output places. To avoid this, a *normalization* operation is required. Let $r_l, l = 1, \cdots$ be the rules that are activated. The *strength* s_l of a rule is calculated with $s_l = \mu_1 \cdots \mu_k$ where μ_i is the membership of the input variable. Let z_l be the crude value provided as a consequence by the rule r_l. The value of the transition output variable x' is:

$$x' = \frac{\sum\limits_{l} z_l \cdot s_l}{\sum\limits_{l} s_l} \tag{10}$$

The regular fuzzification of the variable x provides the token that is injected into the output place. As a consequence, the execution of every transition leads to a single token or no token in each output place. The result of the normalization operation leads to a token the fulfills the relation:

$$\sum_{i} \mu_i = 1 \tag{11}$$

A FLETPN can model the synchronous and asynchronous reaction to a signal belonging to a continuous domain. The handling of the discrete events can be implemented by constraining the membership degree. For example, if a variable x_i assigned to a place p_i is of the discrete event type, the variable belongs (by convention) to PL set. That means all the tokens injected into the place p_i have the form $< 0, 0, 0, 0, 1 >$.

A discrete event variable is a particular case of a continuous variable. All the discrete event variables belong to the same set (could be a fuzzy set) with a membership degree $\mu = 1$. A transition could have input places corresponding to discrete event variables and places corresponding to continuous variables. The assigned fuzzy rule set has to be constructed according to this structure.

In conclusion, a FLETPN model can mix the continuous type tokens with the discrete event type tokens, but every place can contain only one type tokens. Using Petri nets with tokens integrating higher complex information simplifies the program structure, while the program functionality is moved to the associated FLRSs. A program with a simpler structure is easier to be synthesized and to be tested for fulfilling the real-time features. The FLRSs have to be found such that the program fulfills the functional requirements.

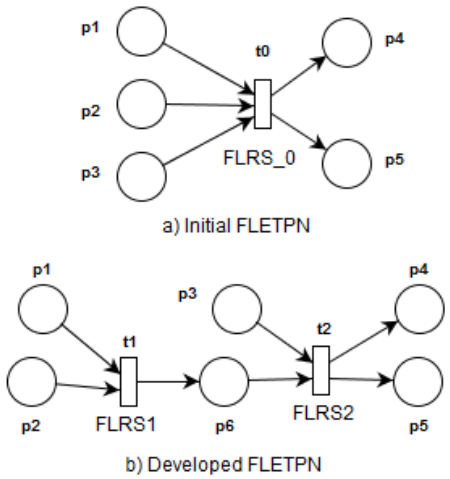


Fig. 5. Reducing the number of a transition's input places.

The FLETPN model should be free of conflicts. If conflicts exist in the ETPN model, the executor grants the execution to the transitions with shorter delays, and if multiple transitions with the same delays are simultaneously enabled, the transitions with lower indexes are chosen for fire. Even if an ETPN model has conflicts, these can be removed by the appropriate conceiving of the FLRSs using the method shown in Figure 4 and Table I.

According to [8] the reasoning process by using fuzzy PN can be implemented by algorithms involving reachability trees, algebra forms and high level PNs. The current approach concerns the modeling of dynamic control systems implemented by reactive programs. The TPNs describe the program structures, while the FLRSs implement their functionalities.

C. FLRS construction

For practical reasons it is convenient to have transitions with maximum two input places. In this case all the fuzzy logic rules have maximum two premises. If there are requirements to have transitions with more than two input places, each of them can be replaced by two transitions as shown in Figure 5. Consequently, all the FLRSs have maximum two premises, but the number of consequences of a rule is not limited.

Supposing that the control system synthesizer constructed the FLETPN, a remained relevant task for the current method consists of the construction of the fuzzy logic rule sets that are assigned to transitions. The proposed method uses the Genetic Algorithm (GA) to search a FLRS that is capable to control the given plant with a specified competence. The control system fulfills the competence requirements if the assessment of the system behavior exceeds a specified threshold. The control system synthesizer has to provide the set of relevant tests used for system evaluation.

The genome is composed of genes coding (by non-negative integers) the consequences of the all FLRSs and genes coding (by real numbers) the weighting coefficients. Table III shows

TABLE III
Example of a genotype.

$c_{1,1}$	$c_{1,2}$	...	$c_{4,5}$	$c_{5,5}$	w_{11}	w_{21}
000010	010100	...	100100	001100	$2,35$	$-3,28$

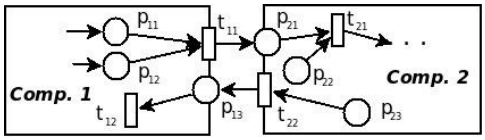


Fig. 6. Example of CPS component diagram.

an example of a genotype for the FLETPN presented in Figure 1. The notations $c_{i,j}(i = 1,\ldots,5; j = 1,\ldots,5)$ represent the pairs of consequences included in a table (e. g. Table II) considered as matrix.

Two kinds of mutation operators were used: one acting on the integer number part and another on the real number part. The synthesizer has to provide the domains of the parameters w_{ij}. The crossover operator splits the genotypes in one point. The selections are performed using the classical performance functions taking into account the plant set points and the constraints.

III. CPS Component Diagram

The design of a control system is based on a set of components that contain discrete event, discrete time or hybrid models. They are included into a component diagram. The models of the current approach are FLETPN.

Fig. 6 shows an example of a CPS component diagram. The two components Comp.1 and Comp.2 are connected through the ports represented on the frontiers. The component ports on the frontiers contain transitions and places. The transitions on the frontier describe outgoing actions (send tokens), and the places receive the tokens.

All the components have input and output ports implementing interfaces. In the given example the component Comp. 1 has the interfaces $Out = \{t_{11}\}$ and $Inp = \{p_{13}\}$. In a well designed component diagram the transitions of a set Out have input places only from the places included in the FLETPN model of the current component, and the places of a set Inp have output transitions only from the current component.

Each component has its own thread of execution. An interface Inp is implemented by an input port and an interface Out by an output port.When a transition of the set Out is executed, the component (output port) sends the corresponding tokens to the components that have in their input interfaces places of the current transition output set via linked component ports. This is denoted by $send(t_i^o, X_{i,out})$ where $X_{i,out}$ corresponds to the tokens (i.e. the variable values of the marking) that have to be sent and included in the transition output places. The destination component executes $receive(Inp, X_i)$ and updates its marking.

The component thread is executed cyclically or when an external event is signaled. The thread is awoke by the input port when a new token arrives or when the clock tic occurs. If the signaled event is tic, the delays of the activated transitions are decreased. If the signaled event is *new token*, the cycle of the thread starts with updating the information of its input place set. The FLETPN marking is updated with the newly received information.

Complex applications can be conceived including components in other components. The links between a component and its included components implement the same protocol outgoing port-ingoing port based on the transition-place connection.The proposed models partition the program structure and functionality at the component level in a compact manner.

IV. FLETPN executor

The executor algorithm of FLETPN is executed with the period of 1 time unit (t.u.) or when an external event is signaled loading an input place with a token. The algorithm updates the places of the input set and determines the transitions that are enabled taking into account the markings of the transition's input place set. If a transition is chosen to be executed, the tokens from its input place set are removed and injected into a temporal marking vector M_t. A time counter $Delay[t_i]$ is loaded with the transition assigned delay if it has any or zero. If the time counter is zero or it reaches the value 0 (after decreasing), the execution of the transition is finished. If a transition belongs to the output set *Out*, its execution is signaled to the linked output place and the corresponding token is loaded.

The counters of all the started transitions are decreased after each sample period.

FLETPN executor algorithm:
Input: ***Pre***, ***Post***, $X, M_0, P, T, D,$ ***FLRS***, Out, Inp;
Initialization: $M = M_0$, *execList* $= empty$;
* reorder the transition set T according to their delays;
repeat
  wait(event);
  if event is tic **then**
    * decrease the Delays of the transitions in *execList*;
  else
    $receive($*Inp*$, X_i)$;
    * update M;
  end if;
  repeat
    for all $t_i \in T$ **do**
      if all $p \in {}^o t_i$, $M(p) \neq \Phi$ **then**
        * move the tokens of ${}^o t_i$ from M to M_t;
        * add t_i to *execList*;
        $Delay[t_i] = d_i$;
      end if;
    end for;
    for all t_i in *execList* **do**
      if ($Delay[t_i]$ is 0) **then**
        * remove t_i from *execList*;
        * calculate and inject the tokens in M for all t_i^o;

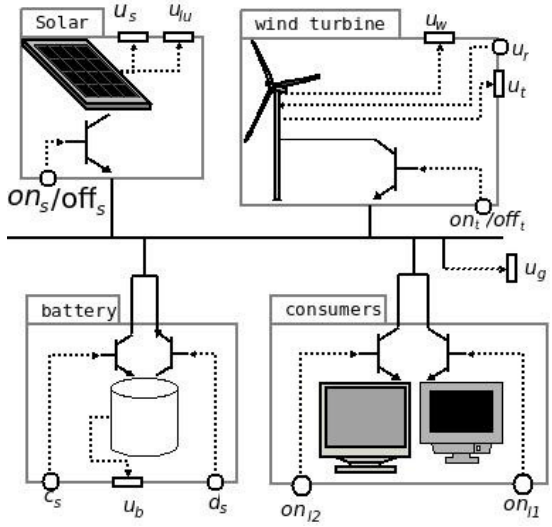


Fig. 7. Microgrid architecture.

* remove the tokens from M_t for all ${}^o t_i$;
end if
if $t_i \in Out$ **then**
$send(Out, X_{i,out})$;
end if
end for
until there is no transition that can be executed;
until the time horizon;
END algorithm;

V. Examples of Application

A. Energy microgrid specifications

The plant represented in Fig. 7 concerns an energy micro grid composed of a solar cell, a wind turbine, a battery and two loads. The plant has to be controlled according to the specification. The control system has to maintain the voltage (u_g) of the main bus between u_m and u_M. The turbine and solar generators asynchronously inject energy into the system. The consumers asynchronously demand the use of the energy as well. When the produced energy exceeds the demands, the surplus has to be discharged on the battery. If the level of the generated energy is lower than the demand, the battery should be used to increase the voltage to guaranty that the bus voltage level remains between the specified limits. The control system is composed of load controller (L-Controller), turbine controller (T-Controller), solar cell controller (S-Controller) and battery controller (B-Controller) as can be seen in Fig. 8.

The following notations are used:

- u_t the turbine output voltage
- u_w the wind force applied to the turbine
- u_n the main bus nominal voltage
- u_g the voltage of the grid main bus
- u_s the solar cell output voltage
- u_{lu} the luminosity on the solar cell
- on_s/off_s control signal to connect or disconnect the solar cell
- on_t/off_t control signal to connect or disconnect the turbine
- d_s control signal to connect the battery to increase the voltage
- c_s control signal to connect the battery to charge it
- u_b the battery level
- on_{l1} control signal to connect the first load
- on_{l2} control signal to connect the second load

The control system requirements are:

repeat
$u_g = read(u_{bus})$;
if $(u_g \geq u_M)$ **then**
* connect the battery to discharge energy;
end if;
if $(u_g \leq u_m)$ **then**
$u_b = read(u_{battery})$;
if $u_b > u_{minim}$ **then**
* connect the battery to increase the voltage;
end if;
end if;
if $(u_m \leq u_g \leq u_M)$ **then**
* disconnect the battery;
end if;
$u_t = read(u_w)$;
if $(u_t < u_n)$ **then**
* stop(turbine);
else
*start(turbine);
end if;
$u_s = read(u_{lu})$;
if $(u_s < u_n)$ **then**
* stop(solarCell);
else
* start(solarCell);
end if
wait(1 t.u.);
$u_g = read(u_{bus})$;
if $(u_g < u_m)$ **then**
* stop(load 2);
else
*allow(load 2);
end if;
wait(1 t.u.);
$u_g = read(u_{bus})$;
if $(u_g < u_m)$ **then**
* stop(load 1);
else
*allow(load 1);
end if;
wait(1 t.u.);
until (the time horizon);

B. Plant model

The loads are considered pure resistances. The photo voltaic solar cells produce energy proportionally with the environment luminosity. The most complex is the wind turbine model.

The discretization (by approximation) of the wind turbine model constructed of differential equations [14] leads to:

$$X_1(k+1) = A_1 \cdot X_1(k) + B_1 \cdot u_r(k) \tag{12}$$

$$Y_1(k+1) = C_1 \cdot X_1(k) \tag{13}$$

$$U(k) = u_w(k) \cdot \cos(Y_1(k)) \tag{14}$$

$$X_2(k+1) = A_2 \cdot X_2(k) + B_2 \cdot U(k) \tag{15}$$

$$u_t(k+1) = C_2 \cdot X_2(k) \tag{16}$$

The notations are:

- X_1 and X_2 are 3 dimensional state vectors,
- Y_1 is a mono dimensional output vector,
- u_r is the input control signal used for the positioning of the turbine,
- U is a combination of the output Y_1 and the wind force u_w.
- u_t is the turbine output voltage.

The corresponding matrix are:

$$A_1 = \begin{pmatrix} 0.050558 & 2.6979e-10 & 5.9355e-05 \\ 0.029825 & 1 & 0.77505 \\ 0.034992 & -4.1333e-06 & 0.58666 \end{pmatrix}$$

$$B_1 = \begin{pmatrix} 0.052746 \\ 0.0020924 \\ 0.0049453 \end{pmatrix}$$

$$C_1 = \begin{pmatrix} 0 & 1 & 0 \end{pmatrix}$$

$$A_2 = \begin{pmatrix} 0.0021802 & -5.8872e-08 & 0.0062 \\ 0.029825 & 1 & 0.77505 \\ 0.034992 & -4.1333e-06 & 0.58666 \end{pmatrix}$$

$$B_2 = \begin{pmatrix} 0.0019624 \\ 0.010389 \\ 0.18356 \end{pmatrix}$$

$$C_2 = \begin{pmatrix} 300 & 0 & 0 \end{pmatrix}$$

The discretization of the continuous model of the wind turbine was performed for the aim of reducing the calculus volume involved by the GA.

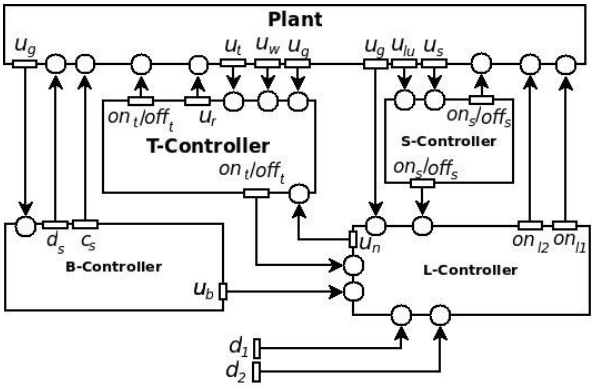


Fig. 8. Control component diagram.

TABLE IV
COEFFICIENTS OF THE FLETPN MODEL

w_1	w_2	w_3	w_4	w_5	w_6
-0.1829	-2.5079	-7.7209	0.1332	4.7337	0.0076

C. Control system architecture

The control system can be conceived as:

- independent controllers
- coordinated controllers or
- cooperative controllers.

Figure 8 shows the proposed component diagram for the control system. The component T-Controller solves the control problem related to the turbine, the component S-Controller concerns the solar cell and the component B-Controller connects or disconnects the battery. The L-Controller has the role to receive the user demands to connect *load 1* and *load 2*. The L-Controller decides to perform these actions taking into account the current energy produced and accumulated. The L-Controller can work independently, to coordinate the other controllers, or to cooperate with them according to the specifications.

D. Wind turbine control component

Figure 9 shows the FLETPN model synthesized for the turbine control. The zero delays of the transitions are not represented on the FLETPN for simplicity reason. The coefficients of the FLETPN are given in Table IV. Some transitions have associated FLRSs as they are mentioned in Tables V, VI, VII, VIII and IX. Other transitions perform simple transformations or store operations.

The T-Controller achieves a kind of fuzzy logic PID (Proportional Integrative Derivative) control function. The place p_0 is loaded with a token corresponding to u_n (nominal voltage, i.e. set point) and the place p_1 with a token corresponding to u_t (turbine output voltage; when it works, it is equally to the main bus voltage). The transition t_0 calculates (using the FLRS assigned to the current transition) the error $e(k) = u_n(k) - u_t(k)$. The resulted tokens are injected into the places

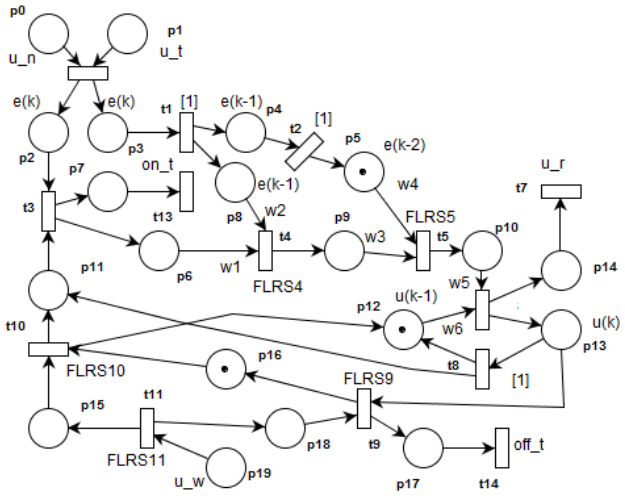


Fig. 9. Turbine control FLETPN.

TABLE V
Fuzzy logic rules of transition t_4 in Fig. 9

$x_6 \backslash x_8$	NL	NM	ZR	PM	PL
NL	NM,PM	PM,ZR	NL,NM	NL,ZR	NL,NM
NM	NL,ZR	NM,NL	NM,ZR	ZR,ZR	PL,PM
ZR	NL.PM	NL,NM	NL,PM	NL,NL	NL,ZR
PM	ZR,NL	PM,NM	ZR,NM	ZR,PM	PL,PM
PL	PM,PM	NM,ZR	NL,NM	PM,NL	NM,NM

TABLE VI
Fuzzy logic rules of transition t_5 in Fig. 9

$x_9 \backslash x_5$	NL	NM	ZR	PM	PL
NL	PM,ZR	PL,ZR	ZR,PL	ZR,NM	PM,PL
NM	PL,NL	NM,ZR	NM,PM	NL,PL	ZR,PM
ZR	PL,NM	PL,NL	NM,PL	NL,NL	PM,NL
PM	PM,PM	NL,NM	PM,NM	PM,NM	PL,ZR
PL	NL,NL	NL,PM	ZR,ZR	PL,ZR	NL,ZR

TABLE VII
Fuzzy logic rules of transition t_9 in Fig. 9

$x_{13} \backslash x_{18}$	NL	NM	ZR	PM	PL
NL	ZR,NL	ZR,NL	Φ, Φ	Φ, Φ	Φ, Φ
NM	ZR,NM	ZR,NM	Φ, Φ	Φ, Φ	Φ, Φ
ZR	ZR,ZR	ZR,ZR	Φ, Φ	Φ, Φ	Φ, Φ
PM	ZR,PM	ZR,PM	Φ, Φ	Φ, Φ	Φ, Φ
PL	ZR,ZR	ZR,ZR	Φ, Φ	Φ, Φ	Φ, Φ

TABLE VIII
Fuzzy logic rules of transition t_{10} in Fig. 9

$x_{15} \backslash x_{16}$	NL	NM	ZR	PM	PL
NL	Φ, Φ	Φ, Φ	Φ, Φ	ZR,PM	ZR,PL
NM	Φ, Φ	Φ, Φ	Φ, Φ	ZR,PM	ZR,PL
ZR	Φ, Φ	Φ, Φ	Φ, Φ	ZR,PM	ZR,PL
PM	Φ, Φ	Φ, Φ	Φ, Φ	ZR,PM	ZR,PL
PL	Φ, Φ	Φ, Φ	Φ, Φ	ZR,PM	ZR,PL

TABLE IX
Fuzzy logic rules of transition t_{11} in Fig. 9

x_{19}	NL	NM	ZR	PM	PL
	Φ,ZR	Φ,ZR	Φ, Φ	ZR,Φ	ZR,Φ

p_2 and p_3. The transition t_1 injects tokens corresponding to the variable e(k-1) into the places p_4 and p_8 after 1 t.u. (time unit) delay. Similar injection performs the transition t_2 for the value e(k-2) into the place p_5. The transition t_3 calculates if the wind turbine can work properly and inject into the output places p_7 the token on_t. The transition t_4 calculates a function of the type $f(w_1e(k), w_2e(k-1))$ using the $FLRS_4$ presented in the Table V. A similar function is performed by the transition t_5 using the $FLRS_5$ given in Table VI. The place p_{10} contains the current variation $\Delta u(k)$ of the control signal. The transition t_6 calculates the current control signal, using the values $w_5\Delta u(k)$ and $w_6u(k-1)$, and injects it into the places p_{13} and p_{14}. The transition t_7 sends the control signal u_r to the turbine. The transition t_8 reloads the place p_{12} with the previous value of the control signal and permits a new execution loading the place p_{11}.

The input place p_{19} is loaded with the wind force u_w (speed) value. If the u_w is lower than a specified value, the turbine is stopped by injecting a token off_t into the place p_{18}. This allows the execution of the transition t_9. If the turbine was stopped (p_{16} has a token) and the wind speed is according to specification, the transition t_{11} allows the turbine to start injecting a token into the place p_{15}. The $FLRS_{11}$ assigned to the transition t_{11} discerns if the wind turbine can work properly injecting a token into the place p_{15} or not and as a consequence it injects a token into the place p_{18}. Table IX contains the $FLRS_{11}$ assigned to transition t_{11}.

E. Load control component

Figure 10 presents the FLETPN model of an independent L-Controller component. This receives in place p_1 as a continuous variable the bus voltage u_g and transforms it into a fuzzy logic value. The L-Controller receives the user's demand d_1 to connect *load 1* as a discrete input $< 0, 0, 0, 0, 1 >$ or disconnect as the value $< 1, 0, 0, 0, 0 >$. The controller uses these two pieces of information to accept or not the demand using the $FLRS_1$ and signals this by the port (transition) t_4 with the values $< 0, 0, 0, 0, 1 >$ or $< 1, 0, 0, 0, 0 >$. The information is passed further to the place p_4. The transition t_2 takes the user demand d_2 to connect or not *load 2*, calculates the controller behavior using the $FLRS_2$ and signals this by the transition t_5.

Figure 11 presents a FLETPN that correspond to a cooperative L-Controller. It added the information on_t and on_s to determine the connection or disconnection of the *load 1* and *load 2*. The L-Controller also sets the reference point u_n for the T-Controller to a better adjustment of the bus voltage u_g. Unlike the previous L-controller, the cooperative controller uses the information $E(k)$ denoting the current power (energy)

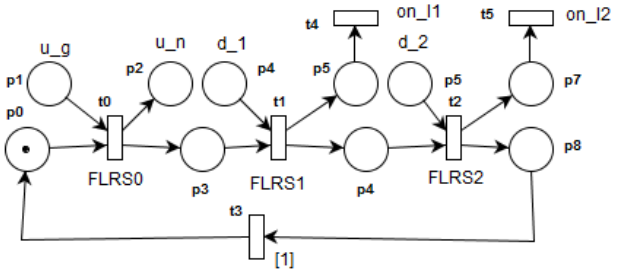


Fig. 10. FLETPN of the independent L-Controller.

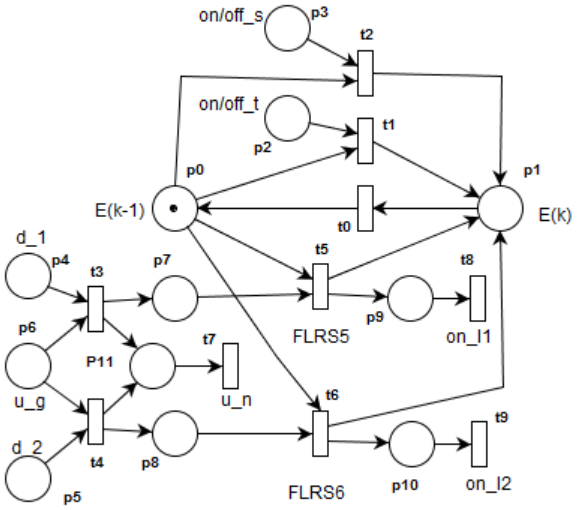


Fig. 11. FLETPN of the cooperative L-Controller.

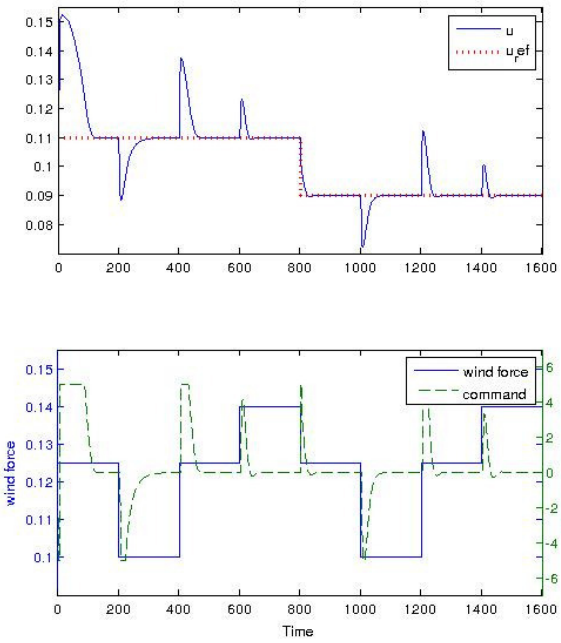


Fig. 12. Turbine signals.

TABLE X
Fuzzy logic rules of transitions t_5 and t_6 in Fig. 11

$x_0 \backslash x_5, x_6$	**NL**	**NM**	**ZR**	**PM**	**PL**
NL	NL, Φ	NM, Φ	ZR, Φ	PM, Φ	PL, Φ
NM	NL, Φ	NM, Φ	ZR, Φ	PM, Φ	PL, Φ
ZR	NL, Φ	NM, Φ	ZR, Φ	PM, Φ	PL, Φ
PM	NL, Φ	NM, Φ	ZR, Φ	ZR, ZR	PM, ZR
PL	NL, Φ	NM, Φ	ZR, Φ	ZR, ZR	PM, ZR

introduced into the system. $E(k-1)$ stores the power available at the previous clock tic. The input place p_2 is injected with a token on/off_t signaling the event that the wind turbine is working or not working respectively. The transition t_1 is used to increase or decrease the information about the current power level. Similar function performs the transition t_2 for the solar cell using the token on/off_s for that purpose. The demand for connecting *load 1* or *load 2* is granted according to the current available power and the voltage u_g. The transitions t_5 and t_6 permit or not the connections and modify the power level. The transitions $t_1, t_2, ..., t_6$ have assigned the necessary FLRSs. Table X shows $FLRS_5$ and $FLRS_6$ assigned to transitions t_5 and t_6.

VI. Tests and Results

All the tests were performed by simulations using standard Java language. Figure 12 presents the test results for the turbine generator. The weighting coefficients $w_i, i = 1, 2, \cdots, 6$ and the FLRSs are calculated using a genetic algorithm. The genome contains the rows of the FLRSs and the weighting coefficients. The fitness function assesses the response to perturbations as shown in Figure 12. The searching process was stopped when a competent solution was obtained, that is the control performances exceed the specified values. The FLRSs obtained by GA are given in Table V, Table VI, ... and Table IX.

Adding the empty set Φ to the fuzzy logic set permits the deterministic selection of the execution on the different paths as can be seen in the FLETPN presented in Figure 9. The conflict between the transitions t_8 and t_9 was solved by the rule *execute the earliest possible transition*. The conflict between the transitions t_9 and t_{10} is solved by transition t_{11} injecting tokens into the places p_{15} or p_{18} according to the token introduced into the input place p_{19}.

In Figure 11 the conflict between the transitions t_3 and t_4 is solved by the rule *execute the transition with the lowest index*. The conflict between t_5 and t_6 is solved by the previous selection.

VII. Conclusions

The proposed method can be easily used to conceive the hybrid control system for different kinds of hybrid plants. It needs the use of knowledge from the same field combining the Petri nets capabilities to implement the discrete event systems

requirements with fuzzy logic models suitable for continuous systems.

There are some benefits of the proposed method: Constructing the tokens with the membership degrees of a variable to all the fuzzy set and assigning to any transition an entire fuzzy rule set leads to a smaller Petri net, and this increases the capability of the model to be used in more complex applications.

The FLETPN models are capable to include the discrete event part and discrete time part. The distinct tokens injected into the corresponding discrete event type places and continuous type places make possible to comprise in the same model the discrete event and discrete time behavior. The FLETPN models can describe the concurrent, synchronous and asynchronous behavior. The reactions to asynchronous events are taken into account when the event occur. These models can easily be implemented, and if a TPN executor is used, the need of a real-time operating system can be avoided.

The structure of the model can be verified using the TPN analyses methods. The proposed method can be used for the verification of the discrete event behavior.

The verification (i. e. the performance evaluation) of the continuous side behavior can be performed by simulation. The weighting coefficients added to input arcs increase the continuous control capabilities enhancing the fuzzy logic rules with the possibility to amplify the relative significance of some variables.

REFERENCES

[1] OMG Unified Modeling Language V2.5, 2015, http://www.omg.org/spec/UML/2.5/PDF/

[2] R. Banach, M. Butler, S. Qin, N. Verma and H. Zhu. Core Hybrid Event-B Machine, Science of Computer Programming, Elsevier, vol. 105, pp. 92-123, 2015.

[3] F. Cicirelli, C. Nigro and L. Nigro, "Qualitative and Quantitative Evaluation of Stochastic Time Petri Nets", Proc. of the IEEE Federated Conference on Computer Science and Information Systems, Vol. 5, 2015, pp. 763–772, doi: 10.15439/2015F69

[4] E.Y.T. Juan, J.P. Tsai and T. Murata, Y. Zhou. Reduction methods for real-time systems using delay time Petri nets, IEEE Transactions on Software and Engineering, volume: 27, no. 5, pp. 422-448, 2001.

[5] T. S. Letia and A. O. Kilyen, Enhancing the time-Petri nets for automatic hybrid control synthesis. Proc. of ICSTCC IEEE Conference, Sinaia, Romania, pp. 627-633, 2014, doi:10.1109/ICSTCC.2014.6982486.

[6] T. Munakata and Y. Jani. Fuzzy Systems: An Overview, Comm. ACM, vol. 37, no. 3, pp. 69-76, Mar. 1994.

[7] C.C. Lee. Fuzzy Logic in Control Systems: Fuzzy Logic Controller, Part I1, Trans. on Systems, Man and Cybernetics, vol. 20, no. 2, pp. 419-435, 1990.

[8] K.-Q. Zhou and A. M. Zain, "Fuzzy Petri nets and industrial applications: a review", Artificial Intelligence Review, Springer, 2016, doi: 10.1007/s10462-015-9451-9.

[9] T. S. Letia and A. O. Kilyen, Fuzzy Logic Enhanced Time Petri Net Models for Hybrid Control Systems . Proc. of AQTR 2016 IEEE Conference, Cluj-Napoca, Romania, 2016, doi: 10.1109/AQTR.2016.7501322.

[10] J. Shi, J. Wan, H. Yan and Hui Suo. A Survey of Cyber-Physical Systems, WCSP - International Conference on Wireless Communications and Signal Processing, pp. 1-6, 2011.

[11] I. Horvath and B. H. M. Gerritsen. Cyber-Physical Systems: Concepts, Technologies and Implementation Principles, Proceedings of TMCE 2012, Karlsruhe, pp.19-36, 2012.

[12] S. N. Krishna and A. Trivedi. Hybrid Automata for Formal Modeling and Verification of Cyber-Physical Systems, Journal of the Indian Institute of Science VOL 93:3 Jul.Sep. 2013

[13] R. Banach, P. Van Schaik and E. Verhulst, "Simulation and Formal Modelling of Yaw Control in a Drive-by-Wire Application ", Proc. of the IEEE 2015 Federated Conference on Computer Science and Information Systems, 731–742, 2015, doi:10.15439/2015F132,

[14] J. de J. Rubio, L. A. Soriano and W. Yu, Dynamic Model of a Wind Turbine for the Electric Energy Generation, Mathematical Problems in Engineering, vol. 2014, Article ID 409268, 8 pages, 2014. doi:10.1155/2014/409268

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 767–773
DOI: 10.15439/2016F164
ACSIS, Vol. 8. ISSN 2300-5963

Cyber Security Impact on Power Grid Including Nuclear Plant

Yannis Soupionis, Roberta Piccinelli and Thierry Benoist
European Commission, Joint Research Centre (JRC)
Institute for the Protection and Security of the Citizen (IPSC)
Security Technology Assessment Unit (STA)
Via E. Fermi, 2749, 21027 Ispra, Italy
Email: yannis.soupionis, roberta.piccinelli, thierry.benoist @jrc.ec.europa.eu

***Abstract*—Decentralized Critical infrastructure management systems will play a key role in reducing costs and improving the quality of service of industrial processes, such as electricity production. The recent malwares (e.g. Stuxnet) revealed several vulnerabilities in today's Distributed Control Systems (DCS), but most importantly they highlighted the lack of an efficient scientific approach to conduct experiments that measure the impact of cyber threats on both the physical and the cyber parts of Networked Critical Infrastructures (NCIs). The study of those complex systems, either physical or cyber, could be carried out by experimenting with real systems, software simulators or emulators. Experimentation with production systems suffers from the inability to control the experiment environment. On the other hand the development of a dedicated experimentation infrastructure with real components is often economically prohibitive and disruptive experiments on top of it could be a risk to safety. In this paper, we focus on the implementation of a Cyber-Physical (CP) testbed which includes physical equipment. We illustrate and the cyber security issues on the communication channel between the Critical Infrastructures(CIs), such as a power grid, a nuclear plant and the energy market. We simulate the power grid network (including nuclear plant), but we emulate the Information and Communications Technology (ICT) part which is the focus of our work. Within this context we assume that we are able to implement scenarios, which produce consequences on the normal operation of the power power grid and the financial area.**

***Index Terms*—Networked Industrial Control Systems; Cyber security; Cyber physical system; power grid; power market; Nuclear plant;**

I. Introduction

EUROPEAN security, both physical and economic, rests upon a foundation of highly interdependent critical infrastructures. A critical infrastructure [1] refers to an asset, system or part thereof located in Member States that is essential for the maintenance of vital societal functions, such as health, safety, security, economic or social well-being of people. The disruption or destruction of such infrastructures would have a significant impact on a Member State as a result of the failure to maintain these functions. The damage to a critical infrastructure, its destruction or disruption by natural disasters, terrorism, criminal activity or malicious behaviour, may have a significant negative impact on the security of the EU and the well-being of its citizens [2].

Given the lack of practical experience with massive infrastructure failures, modeling a multi-CI testbed is highly crucial. Interdependencies between CIs are similarly highlighted in numerous technical publications [3][4][19][20]. The underlying technical theme is that modeling and designing of critical infrastructures must take a holistic, systemic perspective and incorporate interdependencies. In this paper, we examine the complexity of the infrastructure interdependency and highlight the relevant cyber security impact.

In the past, CIs were isolated environments and used proprietary hardware and protocols, thus limiting the threats that could affect them. Nowadays, CIs or more accurately Distributed Control Systems (DCS) are exposed to significant cyber-threats; a fact that has been highlighted by many studies on the security of Supervisory Control And Data Acquisition (SCADA) systems [5], [6], [7].

In this paper, we explore the complexity of the infrastructure interdependency security issues and we design a testbed (Fig. 1), which combines:

- simulated physical infrastructures (e.g. power grid and nuclear plant),
- simulated power stock market,
- emulated ICT controlling infrastructure, and
- real physical equipments, i.e. Programmable Logical Controllers (PLCs).

Based on this implementation, the Network & Information Security Laboratory (NIS Lab) created a cyber-physical system/prototype in order to underline and motivate the need for modeling multiple interconnected critical infrastructures, since the behavior of an interconnected one can be propagated. We discuss the implementation of the testbed and illustrate a possible attack scenario, which shows that network anomalies can produce financial and power disturbances.

The paper is structured as follows. Our study is presented in the context of other related approaches in Section II. In section III we show in detail our experimentation infrastructure and its elements. The experimental scenarios and setup are presented in Section IV. Conclusions are presented in Section V.

II. Related Work

Recent events such as Stuxnet [8], Duqu [9] and Flame [10], caused the scientific community to address cyber security concerns regarding CP systems. In this section we provide a brief presentation of the most relevant approaches addressing

978-83-60810-90-3/$25.00

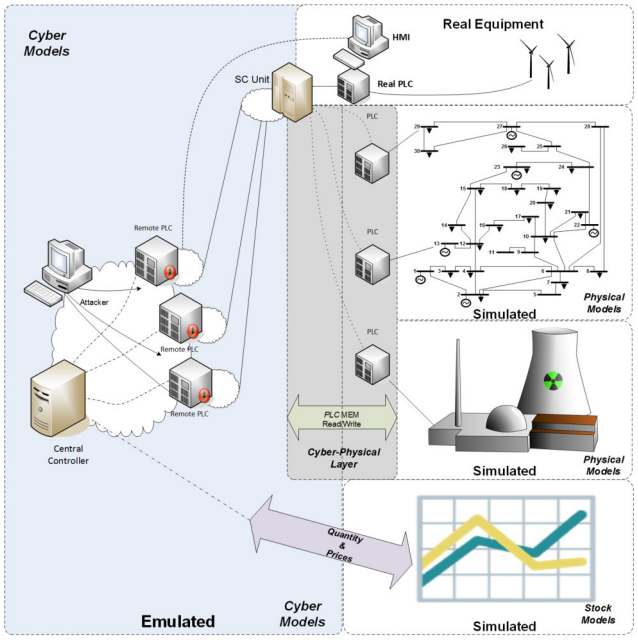


Fig. 1: Experimentation framework architectural overview

the resilience of cyber physical systems, which take into consideration both the communication infrastructure and the control of physical processes.

Nai Fovino *et al.* [18] proposed an experimental platform to study the effects of cyber attacks against NICS. In their paper the authors described several attack scenarios, including DoS attacks and worm infections that send Modbus packets to control hardware. Although the authors provided a wide range of countermeasures, they did not however identify communication parameters that affect the outcome of the attacks. They took into consideration the skills and efforts required by the attacker, as was the case of Stuxnet, where developers also had knowledge of the PLC code, OS and hardware details.

An approach where a cyber physical system is tested for cyber events, has been proposed by Hahn *et al.* [12]. They offer a high level overview of testbed functionality, including its control, communication, and physical components along with a mapping of components to research requirements. Additionally, Yardley *et al.* [13] propose that complex cyber-physical systems like the ones found in the smart grid require a combination of methodology, quantification, and testbed environments to drive tool creation to assist in the evaluation of the systems under test. They present an approach to security testing methodology and illustrate the use of testbeds in developing tools for cutting-edge systems. Both papers highlight the issue but they do not integrate additional infrastructures in order to show the interdependency issues.

Finally, a similar experiment has also been documented by Davis, *et al.* [14] that used the PowerWorld server to study the effects of communication delays between the physical process and human operators.

III. Experimentation Framework Overview

The experimentation framework developed in our previous work [16] follows a hybrid approach, where the Emulab-based testbed recreates the control and process network of NICS, including Programmable Logical Controllers (PLCs) and SCADA servers, and a software simulation reproduces the physical processes.. The main two elements of our laboratory are:

- an experimental platform for resilience, security and stability research, called Experimental Platform for Internet Contingencies (EPIC) [11], which supports the security assessment of cyber-physical systems. The EPIC test-bed can efficiently recreate realistic network topologies and conditions (e.g. delay and loss characteristics of Wide Area Network - WAN links) of the Internet infrastructure.
- a physical system simulator, called the Assessment platform for Multiple Interdependent Critical Infrastructures (AMICI)[16], which can simulate in real time critical physical infrastructures, e.g. a power grid, and can interact with the emulated network test-bed.

The architecture of EPIC suggests the use of an emulation testbed based on the Emulab software [11] in order to recreate the cyber part of NCIs, e.g., servers and corporate network, and the use of software simulation (AMICI) for the physical components, e.g., power grid and nuclear plant.

A. The controlling ICT network

The cyber layer is recreated by an emulation testbed that uses the Emulab architecture and software [17] to automatically and dynamically map physical components (e.g. servers, switches) to a virtual topology. In other words, the Emulab software configures the physical topology in a way that it emulates the virtual topology as transparently as possible. This way we gain significant advantages in terms of repeatability, scalability and controllability of our experiments.

Besides the process network, the cyber layer also includes the control logic code, that in the real world is implemented by PLCs. The control code can be run sequentially or in parallel to the physical model. In the sequential case, a *tightly coupled* code (TCC) is used, i.e. code that is running in the same memory space with the model, within the SC unit. In the parallel case, a *loosely coupled* code (LCC) is used, i.e. code that is running in another address space, possibly on another host, within the *R-PLC* unit (Remote PLC). The cyber-physical layer incorporates the PLC memory, seen as a set of registers typical to PLCs, and the communication interfaces that glue together the other two layers.

In Fig. 2, we illustrate the emulated cyber-part of our experimental setup. Each simulation of the physical processes runs at a different host. Moreover, in conjunction with Fig. 1 (i) the simulation of the main power grid is running on the lower right part of the network, (ii) the PLC is connected on the upper right, (iii) the power market on the upper left, and (iv) the nuclear plant on the left lower part. All these simulated infrastructures communicate though the ICT network .

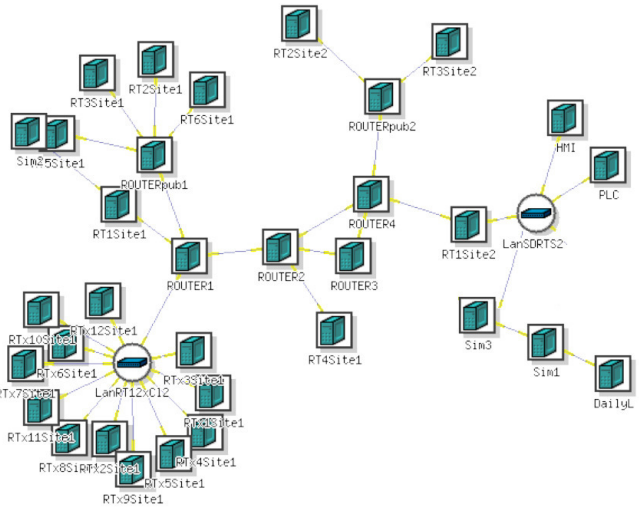


Fig. 2: The EPIC network topology for the specific implementation taken directly from Emulab's web interface.

B. Description of Nuclear Plant

In this section we present our simplified Pressurized Water Reactor (PWR). PWRs constitute the most established and diffuse types of operational Nuclear Power Plants (NPPs), so they have been included in this study to represent the nuclear typology of the generating capacity of the power grid [25].

The proposed reactor model serves as an example to illustrate the concept behind the framework of analysis of cyber-attacks on ICT communicating infrastructure of a power grid including nuclear power plants: it is admittedly simplified in its technical features and strong assumptions are made to keep the focus on the methodological framework.

In a PWR, the energy generated by the fission of atoms heats water, which is pumped under high pressure from the primary circuit to the reactor core. The heated water then flows through a heat exchanger, where it transfers its thermal energy to a secondary circuit, where steam is generated and flows to turbines, which in turn spin an electric generator [26]. The model presented here concentrates on the primary circuit. In the reactor, water, which acts as the moderator and the coolant, passes through the core with upward flow and removes the heat, which the fuel contained in the fuel bars has generated through fission (Fig. 3).

Water enters the bottom of the reactor core at about 275 °C (T_{IN}) is heated and flows upwards through the reactor core at a temperature of about 315 °C (T_{OUT}). Despite the high temperature, water remains liquid due to the high pressure in the primary coolant loop, usually around 155 bar.

Within the hypothesis of a thermal power P_{TH} uniformly distributed along the core, the dynamics of the reactor can be modeled considering the variation of the power generated by the fuel and the power absorbed by the moderator [26]:

$$M_F C_F \frac{dT_F}{dt} = P_{TH} - k\,(T_F - T_M) \tag{1}$$

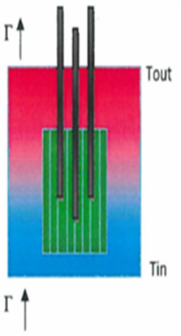


Fig. 3: Simplified model of the core of a PWR. The water flow Γ enters the bottom of the reactor at temperature T_{IN} and is heated to a temperature T_{OUT} while flowing upward through the core.

$$M_M C_M \frac{dT_M}{dt} = (T_F - T_M) - \Gamma C_M\,(T_{OUT} - T_{IN}) \tag{2}$$

$$T_{OUT} = 2T_M - T_{IN} \tag{3}$$

where

- M_F and M_M are respectively, the fuel and moderator masses,
- C_F and C_M are the fuel and moderator thermal coefficients,
- k is the global thermal coefficient, which accounts for the thermal exchange between fuel and water,
- Γ is the moderator flow and it can vary within the range of 104-105 tonn/h, and
- T_F and T_M are the fuel and the moderator temperatures. In particular, T_M is computed as the average between the inflow (T_{IN}) and the outflow (T_{OUT}) temperature of the water.

The equations (1), (2) present energetic balances on fuel and on moderator respectively. On the fuel side (first equation), the change in the produced energy is given by a source term P_{TH}, the produced thermal power, subtracted by the energy exchanged with the moderator. On the moderator side (second equation), the change in the absorbed energy is given by the difference between the energy exchanged and absorbed by the moderator. The equation (3) represents the assumed tie between the inflow and the outflow temperature of the water.

In the model, the demanded power P_{TH} and the inflow temperature T_{IN} of the moderator are the inputs. When an increase of 1kW of power and of 1 °C is given to the input variables, the step response of the system (Fig. 4) evidences that temperature rises by a factor of P/($M_F C_F$) ≈ 0.2. Since power acts mostly on the fuel temperature, the effects on the moderator temperature are negligible (Fig. 4). It is noteworthy that if the power is considered uniformly distributed inside the core, the thermal exchange between fuel and moderator is slow (Fig. 4).

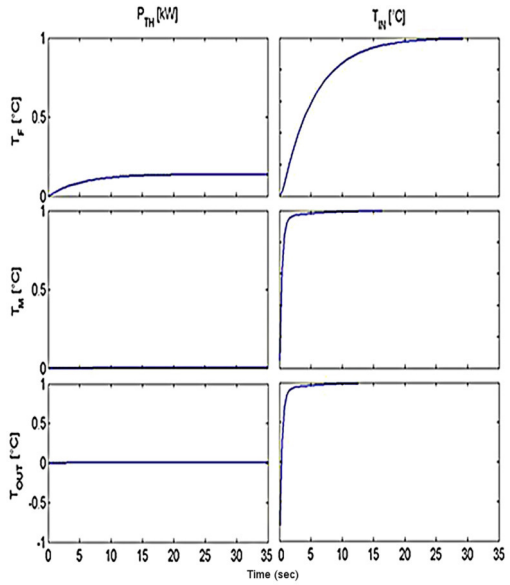


Fig. 4: Step response of the system. Thermal power P_{TH} and inflow temperature T_{IN} have been given respectively a raise of 1 kW and 1 °C.

C. Power market

Since more than one decade, a political change of mind has led to the liberalization of the power markets. Its goal: the creation of an internal European market that achieves security of supply and competitive prices and services for the customers. In this market, a growing variety of enterprises organizes the production, the trading, the marketing, the transmission and the supply of electricity, respecting appropriate regulation.

Producers compete to sell energy at the best possible price. The suppliers which deliver electricity to the final consumers buy the energy on the wholesale market from the producers or the trading companies. Power markets or spot markets offer trading platforms [21][22][23][24] to allow members to exchange information on values and prices and to submit bids for buying and selling power. Therefore a possible interruption between the communication of the power grid and the power market can lead to financial disturbances and even to market/prices manipulation.

In our testbed we have implemented a spot market, which provides automatically the prices on the requested quantity. If the requested energy quantity is not able to be provided by the lowest price energy producer, then the rest is obtained by the next one. For example, if the requested quantity is 100 KW and the lowest price producer is able to provide only 70 KW, the remaining 30 are going to be obtained by the second lowest one.

D. The power grid model

The IEEE electrical grid models [15] are extensively used by the scientific community since they are known to accurately encapsulate the basic characteristics of real infrastructures. As such, AMICI provides a broad range of grid models to experiment with, including the Western System Coordinating Council's (WSCC) 3-machine 9-bus system, and the 30-bus, 39-bus and 118-bus test cases, which represent a portion of the American Electric Power System as of early 1960. These constitute realistic models which are well-established within the power systems community and provide a wide range of power system configurations. An example graphical "bus-view" of the IEEE 30-bus power grid test system is given in Fig. 5. The IEEE 30 Bus Test Case represents a portion of the American Electric Power System (in the Midwestern US). Apart from the connecting buses, it consists of 6 generators and 20 load consuming buses. This is the main IEEE power grid we are going to use for our testbed.

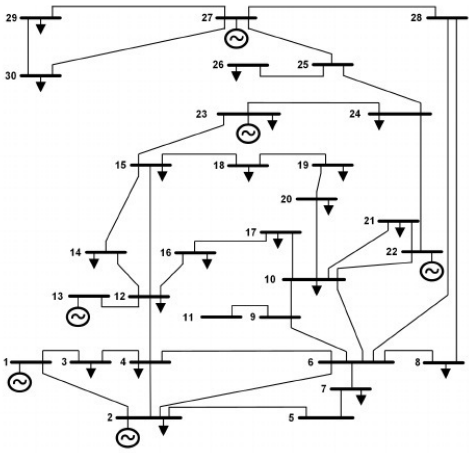


Fig. 5: The IEEE 30-bus test system.

E. Integration into AMICI

We have developed a generic approach to integrate a wide range of IEEE electrical grid models into Simulink and to prepare them for real-time simulation. This way AMICI [16] facilitates the timely integration of other models and eliminates manual, error-prone operations, which are mainly due to the large number of vectors and matrices that must be set-up.

The physical processes are implemented into AMICI by creating the Simulink models via the mathematical functions, but from a technical point of view real-time simulation of IEEE grid models in AMICI is enabled through a combination of two Matlab open-source libraries: MatPower [27] and MatDyn [28]. MatPower is an open-source Matlab package for solving power flow and optimal power flow problems. It is a simulation tool that includes several built-in IEEE test systems, which are already prepared for analysis. Since MatPower only provides static analysis of power systems, it needs to be coupled with MatDyn in order to benefit from its support for dynamic analysis, i.e., real-time simulation.

IV. Description of the Experimental Setup

A. Experimental Setup

The following experimental setup was implemented in the Joint Research Centre's (JRC) EPIC laboratory. The Emulab testbed included nodes with the following configuration: FreeBSD OS 8, Intel Xeon E5606 @2.1GHz and 2GB of RAM.

As shown in Fig. 2 the experimental setup consisted of 2 Routers (Cisco 6503), which have four Gigabit experimental interfaces and one control interface, and 22 hosts:

- one (1) host runs the power grid unit (AMICI model),
- one (1) host runs the nuclear plant (AMICI model),
- two (2) hosts for running the R-PLC units for the interconnection between power grid, nuclear plant and network,
- one (1) host for running the power market unit, and
- the (3) hosts to run the malicious software (attackers).
- the rest of the hosts to produce normal traffic.

The reason to use such a large testbed is to try replicate the attack from various hosts and verify the results. Within the Emulab testbed we emulated packet losses and background traffic (potentially DoS attack) in order to recreate a dynamic and unpredictable environment such as the Internet. For the background traffic we used UDP packets generated with both PathTest [1] and Iperf [2]. We have installed those tools in all the attacker nodes.

Additionally, we adjusted our networks in order to emulate the bandwidth limitations (10Mb/s) not only for the Internet but also for the communication to PLC. The communication between R-PLCs and the power grid model was implemented with a 100Mb/s to provide maximal performances for the interaction between R-PLC units. Finally, we should state that there is a synchronization algorithm between the models execution time and the system clocks ensuring reliable exchange of data.

B. Scenarios

The implemented scenarios are two, the first one affects the nuclear plant and the second one the real PLC, which is connected to the power grid.

1) Attack against a nuclear plant's PLC: In the implemented scenario the attacker interacts with PLCs by sending legitimate Modbus packets. This scenario assumes an attacker is able to access an internal network by bypassing the security of either the control center or substation networks. By doing so, it is possible to compromise the PLC, produce different values, and hide the fact that the plant produces approximately 30MW/h less without having any specific alarm. Since, the additional power is produced by the rest of the power grid, this means that the additional cost is around 793 euros/h, taking into account the average price for 30MW provided by the spot market (section III-C).

In Fig. 6 we see the minimal change of the temperature on the two different stages with normal and compromised PLC. It should be stated that the graph considers a timeframe of 10 hours: every hour the system registers a different level of energy demand and reacts accordingly. Since power acts mostly on the fuel temperature T_{fuel}, the effects on the moderator temperature T_m are negligible. When the system experiences a constant increase in the power demand, it reacts accordingly by varying the fuel temperature T_{fuel2} and the moderator temperature T_{m2}.

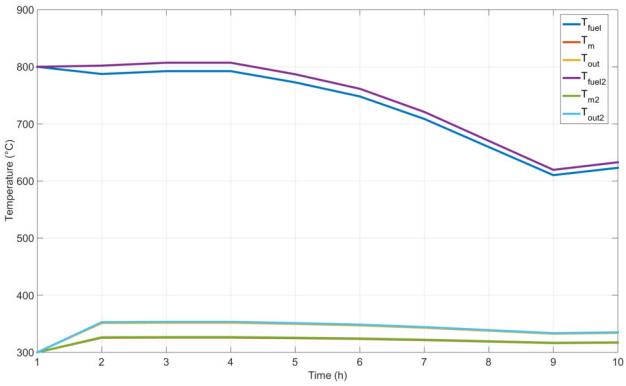


Fig. 6: Behavior of the system. The dynamics of the reactor is described by three quantities: the temperature of the fuel, T_{fuel}, the temperature of the moderator, T_m, and the temperature of the outflow water, T_{out}, and it is represented in two different load demand configurations (T_{fuel}, T_m, T_{out}) and (T_{fuel2}, T_{m2}, T_{out2}). Tm is computed as the average between the inflow (T_{in}) and the outflow temperature (T_{out}) of the moderator.

2) Attack against a generator's PLC: The DDoS attacks were implemented in different bandwidths up to 100Mbit/s, but we minimize our attacks till 20Mbit/s in order to be more realistic. We did not target the attack against a specific equipment because then the attack would be "extremely" successful. We aimed to minimize the network bandwidth between the physical equipment and the power market. Therefore there may be partial loss of communication between those entities.

In this scenario we have a dedicated router connecting the main elements. The router is not reachable through the Internet, but is used to pass communication for other services, such as web services, etc . This means that a DDoS attack cannot be aimed directly at the PLC, but by attacking a specific other service the network bandwidth is going to be limited. When the DDoS attack takes place, additional energy is needed by the consumers of the grid. So following the power market auction procedure, the producers place their bid and the power market decides on who has the best offer. During the scenario the needed load is constant 100MW split over various consumers. Moreover, the real PLC is connected to various simulated generators (each time to a different one) in order to identify any deviations based on the power grid topology.

The parameters we considered for the following experiments are packet losses and background traffic. For packet losses

[1]PathTest, Free Network Capacity Test tool, 2015
[2]Iperf: The TCP/UDP Bandwidth Measurement Tool, 2015

we used 3 rates: 0%, 10% and 20%. Finally, for the attack's background traffic we used: 5Mb/s, 10Mb/s and 20Mb/s. For each configuration setting, representing a combination of packet loss rate and background traffic we executed a separate experiment.

Within this context we measured a maximal success rate of 77% and a minimal success rate of 28%. The results show that even for 15% packet losses and 20 Mb/s the attack success rate is not 100%. More specifically, this means that from 100 attempts, an average of less than 50 will fail to produce some financial gain for the attacker or loss for the consumer. It should not come as a surprise that we measured a higher success rate for a larger loss rate. An explanation for this behavior is that the reduced number of packets does not assist the power market to verify that the offers need confirmation. These results are depicted in Tab. I. The cost is calculated by taking into account the success rate and the average price provided by the spot market for 100 MW.

TABLE I: Attack success rate and additional cost

DDoS success rate				
2.5Mb/s traffic	**5Mb/s traffic**	**10Mb/s traffic**	**Packet loss(%)**	**max Cost (euros/h)**
28%	32%	60%	0%	1586.58
29%	35%	68%	10%	1798.12
28%	37%	77%	20%	2036.11

V. Conclusions and further research

Cyber-physical systems are engineered systems that are built from, and depend upon, the seamless integration of computational algorithms and physical components. In this paper we present the implementation of a cyber-physical testbed including multiple Critical Infrastructures (CIs):

- two simulated interconnected infrastructures, power grid and nuclear plant,
- a simulated power market for providing the cost for the provided energy,
- a real PLC which is interconnected with a specific bus of the power network,
- an emulated cyber network which interconnects and controls all the aforementioned elements

To the best of our knowledge, this is the first time that those elements were presented/interconnected in a prototype and provides a step forward towards understanding the cyber security vulnerabilities of the current cyber-physical systems. In this paper we have analyzed the effects of network parameters on coordinated attacks and we show that they could be significants.

Based on this implementation, more advanced experiments will be created at the Network & Information Security Laboratory (NIS Lab) in order to show the effect of cyber-attacks against real infrastructure including the actions of real human actors (e.g. human operators) in the cyber-physical testing/simulation process. Moreover, we plan to propose and implement a set of countermeasures to tackle and mitigate the attacks, based on the exchanging signals and their statistical analysis for detecting anomalies [29][30].

References

[1] European commission, Directive on European Critical Infrastructures, COUNCIL DIRECTIVE 2008/114/EC, December 2008

[2] Wolthusen S.D., Modeling critical infrastructure requirements, Information Assurance Workshop, 2004, Proceedings from the Fifth Annual IEEE SMC, pp. 101- 108, 2004, http://dx.doi.org/10.1109/IAW.2004.1437804

[3] Yampolskiy, M., Sztipanovits, J., Yuan Xue, Koutsoukos, X.D., Horvath, P., Systematic analysis of cyber-attacks on CPS-evaluating applicability of DFD-based approach, Resilient Control Systems (ISRCS), 2012 5th International Symposium on, pp.55-62, 2012, http://dx.doi.org/10.1109/ISRCS.2012.6309293

[4] Zio, E., Sansavini, G., Modeling Interdependent Network Systems for Identifying Cascade-Safe Operating Margins, Reliability, IEEE Transactions on, vol. 60, no. 1, pp. 94-101, 2011, http://dx.doi.org/10.1109/TR.2010.2104211

[5] Zhu, B., Joseph, A., Sastry, S., A taxonomy of cyber attacks on SCADA systems. In Internet of things (iThings/CPSCom), 2011 international conference on and 4th international conference on cyber, physical and social computing (pp. 380-388). IEEE, October, 2011, http://dx.doi.org/10.1109/iThings/CPSCom.2011.34

[6] Nai Fovino, I., Carcano, A., Masera, M., Trombetta, A: An experimental investigation of malware attacks on SCADA systems. International Journal of Critical Infrastructure Protection, vol. 2, no. 4, pp. 139-145, 2009, http://dx.doi.org/10.1016/j.ijcip.2009.10.001

[7] Rysavy, Ondrej, Jaroslav Rab, and Miroslav Sveda. "Improving security in SCADA systems through firewall policy analysis." In Computer Science and Information Systems (FedCSIS), 2013 Federated Conference on, pp. 1435-1440. IEEE, 2013.

[8] Chen T, Abu-Nimeh S., Lessons from Stuxnet. Computer 2011;44(4):913, http://dx.doi.org/10.1109/MC.2011.115

[9] Fidler D., Tinker, Tailor, Soldier, Duqu: Why cyberespionage is more dangerous than you think. International Journal of Critical Infrastructure Protection 2012;5(1):289, http://dx.doi.org/10.1016/j.ijcip.2011.12.001

[10] Munro, Kate. "Deconstructing flame: the limitations of traditional defences." Computer Fraud & Security 2012.10 (2012): 8-11, http://dx.doi.org/10.1016/S1361-3723(12)70102-1

[11] Siaterlis, C., Garcia, A.P. and Genge, B., 2013. On the use of Emulab testbeds for scientifically rigorous experiments. Communications Surveys & Tutorials, IEEE, 15(2), pp.929-942, http://dx.doi.org/10.1109/SURV.2012.0601112.00185

[12] Hahn, A., Ashok, A., Sridhar, S. and Govindarasu, M. Cyber-physical security testbeds: Architecture, application, and evaluation for smart grid. Smart Grid, IEEE Transactions on, 4(2), pp.847-855, 2013, http://dx.doi.org/10.1109/TSG.2012.2226919

[13] Yardley, Tim, Robin Berthier, David Nicol, and William H. Sanders. "Smart grid protocol testing through cyber-physical testbeds." In Innovative Smart Grid Technologies (ISGT), 2013 IEEE PES, pp. 1-6. IEEE, 2013, http://dx.doi.org/10.1145/2602575

[14] Davis, C. M., J. E. Tate, H. Okhravi, C. Grier, T. J. Overbye, and D. Nicol. "SCADA cyber security testbed development." In Proceedings of the 38th North American power symposium (NAPS 2006), pp. 483-488. 2006, http://dx.doi.org/10.1109/NAPS.2006.359615

[15] University of Washington - Electrical Engineering, "Power Systems Test Case Archive," http://www.ee.washington.edu/research/pstca/, 2012, [Online; accessed January 2016].

[16] Genge, Béla, Christos Siaterlis, and Marc Hohenadel. "AMICI: An assessment platform for multi-domain security experimentation on critical infrastructures." In Critical information infrastructures security, pp. 228-239. Springer Berlin Heidelberg, 2012, http://dx.doi.org/10.1007/978-3-642-41485-5_20

[17] White, B., Lepreau, J., Stoller, L., Ricci, R., Guruprasad, S., Newbold, M., Hibler, M., Barb, C., Joglekar, A.: An integrated experimental environment for distributed systems and networks. In Proc. of the Fifth Symposium on Operating Systems Design and Implementation, pp. 255-270, 2002, http://dx.doi.org/10.1145/844128.844152

[18] Nai Fovino, I., Masera, M., Guidi, L., Carpi, G.: An Experimental Platform for Assessing SCADA Vulnerabilities and Countermeasures in Power Plants. In Proc. HSI, pp. 679-686, 2010, http://dx.doi.org/10.1109/HSI.2010.5514494

[19] Bialas, A., 2015, September. Experimentation tool for critical infrastructures risk management. In Computer Science and Information Systems (FedCSIS), 2015 Federated Conference on (pp. 1099-1106). IEEE, http://dx.doi.org/10.15439/2015F77

[20] Preisler, T., Dethlefs, T., & Renz, W. (2015, September). Simulation as a service: A design approach for large-scale energy network simulations. In Computer Science and Information Systems (FedCSIS), 2015 Federated Conference on (pp. 1765-1772). IEEE, http://dx.doi.org/10.15439/2015F116

[21] Bunn, Derek W., "Modelling prices in competitive electricity markets," 2004.

[22] Arroyo, José M., and Antonio J. Conejo. "Optimal response of a thermal unit to an electricity spot market," Power Systems, IEEE Transactions on 15.3 (2000): 1098-1104, http://dx.doi.org/10.1109/59.871739

[23] APX Power Spot Exchange, https://www.apxgroup.com/trading-clearing/spot-market/, last accessed on January 12, 2016

[24] EEX Power Spot Exchange, https://www.eex.com/en/products/power/power-spot-market, last accessed on January 12, 2016

[25] World Nuclear Association: www.world-nuclear.org/ Information-Library/ last accessed on January 12, 2015.

[26] Todreas N. E. and Kazimi M.S., Nuclear Systems Volume I: Thermal Hydraulic Fundamentals, CRC press, 2012.

[27] R.D. Zimmerman, C.E. Murillo-Sanchez, and R.J. Thomas, "MATPOWER: Steady-State Operations, Planning, and Analysis Tools for Power Systems Research and Education, IEEE Trans. on Power Systems, vol. 26, no. 1, pp. 12-19, Febr. 2011, http://dx.doi.org/10.1109/TPWRS.2010.2051168

[28] Cole S., Belmans R., "MatDyn, A New Matlab-Based Toolbox for Power System Dynamic Simulation", IEEE Trans. on Power Systems, vol. 26, no. 3, pp. 1129-1136, Aug. 2011, http://dx.doi.org/10.1109/TPWRS.2010.2071888

[29] Soupionis Y., Ntalampiras S., and Giannopoulos G., "Faults and Cyber Attacks Detection in Critical Infrastructures." In International Conference on Critical Information Infrastructures Security, pp. 283-289. Springer International Publishing, 2014.

[30] Kornecki, A. J., Subramanian, N., & Zalewski, J. (2013, September). Studying interrelationships of safety and security for software assurance in cyber-physical systems: Approach based on bayesian belief networks. In Computer Science and Information Systems (FedCSIS), 2013 Federated Conference on (pp. 1393-1399). IEEE.

9th International Symposium on Multimedia Applications and Processing

SOFTWARE Engineering Department, Faculty of Automation, Computers and Electronics, University of Craiova, Romania "Multimedia Applications Development" Research Centre

Background and Goals

Multimedia information has become ubiquitous on the web, creating new challenges for indexing, access, search and retrieval. Recent advances in pervasive computers, networks, telecommunications, and information technology, along with the proliferation of multimedia mobile devices—such as laptops, iPods, personal digital assistants (PDA), and cellular telephones—have stimulated the development of intelligent pervasive multimedia applications. These key technologies are creating a multimedia revolution that will have significant impact across a wide spectrum of consumer, business, healthcare, educational and governmental domains. Yet many challenges remain, especially when it comes to efficiently indexing, mining, querying, searching, retrieving, displaying and interacting with multimedia data.

The Multimedia—Processing and Applications 2016 (MMAP 2016) Symposium addresses several themes related to theory and practice within multimedia domain. The enormous interest in multimedia from many activity areas (medicine, entertainment, education) led researchers and industry to make a continuous effort to create new, innovative multimedia algorithms and applications.

As a result the conference goal is to bring together researchers, engineers, developers and practitioners in order to communicate their newest and original contributions. The key objective of the MMAP conference is to gather results from academia and industry partners working in all subfields of multimedia: content design, development, authoring and evaluation, systems/tools oriented research and development. We are also interested in looking at service architectures, protocols, and standards for multimedia communications—including middleware—along with the related security issues, such as secure multimedia information sharing. Finally, we encourage submissions describing work on novel applications that exploit the unique set of advantages offered by multimedia computing techniques, including home-networked entertainment and games. However, innovative contributions that don't exactly fit into these areas will also be considered because they might be of benefit to conference attendees.

Call for Papers

MMAP 2016 is a major forum for researchers and practitioners from academia, industry, and government to present, discuss, and exchange ideas that address real-world problems with real-world solutions.

The MMAP 2016 Symposium welcomes submissions of original papers concerning all aspects of multimedia domain ranging from concepts and theoretical developments to advanced technologies and innovative applications. MMAP 2016 invites original previously unpublished contributions that are not submitted concurrently to a journal or another conference. Papers acceptance and publication will be judged based on their relevance to the symposium theme, clarity of presentation, originality and accuracy of results and proposed solutions.

Topics

- Audio, Image and Video Processing
- Animation, Virtual Reality, 3D and Stereo Imaging
- Big Data Science and Multimedia Systems
- Cloud Computing and Multimedia Applications
- Machine Learning, Data Mining, Information Retrieval in Multimedia Applications
- Multimedia File Systems and Databases: Indexing, Recognition and Retrieval
- Multimedia in Internet and Web Based Systems
- E-Learning, E-Commerce and E-Society Applications
- Human Computer Interaction and Interfaces in Multimedia Applications
- Multimedia in Medical Applications
- Entertainment and games
- Security in Multimedia Applications: Authentication and Watermarking
- Distributed Multimedia Systems
- Network and Operating System Support for Multimedia
- Mobile Network Architecture
- Intelligent Multimedia Network Applications
- Future Trends in Computing System Technologies and Applications

Best Paper Award

A best paper award will be made for work of high quality presented at the MMAP Symposium. The technical committee in conjunction with the organizing/steering committee will decide on the qualifying papers. Award comprises a certificate for the authors and will be announced on time of conference.

STEERING COMMITTEE

- **Amy Neustein, Boston University,** USA, Editor of Speech Technology
- **Lakhmi C. Jain, University of South Australia and University of Canberra,** Australia
- **Ioannis Pitas, University of Thessaloniki,** Greece
- **Costin Badica, University of Craiova,** Romania
- **Borko Furht, Florida Atlantic University,** USA
- **Harald Kosch, University of Passau,** Germany
- **Vladimir Uskov, Bradley University,** USA
- **Thomas M. Deserno, Aachen University,** Germany

PUBLICITY CHAIR

- **Amelia Badica, University of Craiova,** Romania
- **Adriana Schiopoiu Burlea, University of Craiova,** Romania

ORGANIZING

- **Dumitru Dan Burdescu, University of Craiova,** Romania
- **Costin Badica, University of Craiova,** Romania
- **Marius Brezovan, University of Craiova,** Romania
- **Adriana Schiopoiu Burlea, University of Craiova,** Romania
- **Liana Stanescu, University of Craiova,** Romania
- **Cristian Marian Mihaescu, University of Craiova,** Romania

EVENT CHAIRS

- **Brezovan, Marius,** University of Craiova
- **Burdescu, Dumitru Dan,** University of Craiova, Romania

PROGRAM COMMITTEE

- **Badica, Amelia,** University of Craiova, Romania
- **Böszörmenyi, Laszlo,** Klagenfurt University, Austria
- **Botez, Ruxandra,** University of Quebec
- **Burlea Schiopoiu, Adriana,** University of Craiova
- **Camacho, David,** Universidad Autonoma de Madrid, Spain
- **Cano, Alberto,** Virginia Commonwealth University
- **Cretu, Vladimir,** Politehnica University of Timisoara, Romania
- **Debono, Carl James,** University of Malta, Malta
- **Fabijańska, Anna,** Lodz University of Technology, Poland - Institute of Applied Computer Science, Poland
- **Fomichov, Vladimir,** National Research University Higher School of Economics, Moscow, Russia., Russia
- **Giurca, Adrian,** Brandenburg University of Technology, Germany
- **Grosu, Daniel,** Wayne State University, United States
- **Groza, Voicu,** University of Ottawa, Canada
- **Kabranov, Ognian,** Cisco Systems, United States
- **Kannan, Rajkumar,** Bishop Heber College Autonomous, India
- **Korzhik, Valery,** State University of Telecommunications, Russia
- **Kotenko, Igor,** St. Petersburg Institute for Informatics and Automation of the Russian Academy of Science, Russia
- **Kriksciuniene, Dalia,** Vilnius University, Lithuania
- **Lau, Rynson,** City University of Hong Kong, Hong Kong S.A.R., China
- **Lloret, Jaime,** Polytechnic University of Valencia, Spain
- **Logofatu, Bogdan,** University of Bucharest, Romania
- **Mangioni, Giuseppe,** DIEEI - University of Catania, Italy
- **Mannens, Erik,** Ghent University
- **Mihaescu, Cristian,** University of Craiova, Reunion
- **Mocanu, Mihai,** University of Craiova, Romania
- **Morales-Luna, Guillermo,** Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, Mexico
- **Ohzeki, Kazuo,** Shibaura Institute of Technology, Japan
- **Popescu, Dan,** CSIRO, Sydney, Australia, Australia
- **Querini, Marco,** Department of Civil Engineering and Computer Science Engineering
- **RUTKAUSKIENE, Danguole,** Kaunas University of Technology
- **Salem, Abdel-Badeeh M.,** Ain Shams University, Egypt
- **Sari, Riri Fitri,** University of Indonesia, Indonesia
- **Stanescu, Liana,** University of Craiova, Romania
- **Tejera, Mario Hernández,** University of Las Palmas de Gran Canaria, Spain
- **Trausan-Matu, Stefan,** Politehnica University of Bucharest, Romania
- **Trzcielinski, Stefan,** Poznan University of Technology, Poland
- **Tsihrintzis, George,** University of Piraeus, Greece
- **Vega-Rodríguez, Miguel A.,** University of Extremadura, Spain
- **Velastin, Sergio,** Kingston University, United Kingdom
- **Virvou, Maria,** University of Piraeus, Greece
- **Watanabe, Toyohide,** University of Nagoya
- **Wotawa, Franz,** Technische Universitaet Graz, Austria
- **Zurada, Jacek,** University of Louisville, United States

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 777–785
DOI: 10.15439/2016F36
ACSIS, Vol. 8. ISSN 2300-5963

An application of the supervoxel-based Fuzzy C-Means with a GPU support to segmentation of volumetric brain images

Anna Fabijańska
Lodz University of Technology
Institute of Applied Computer Science
ul. Stefanowskiego 18/22
90-924 Lodz, Poland
Email: anna.fabijanska@p.lodz.pl

Jarosław Gocławski
Lodz University of Technology
Institute of Applied Computer Science
ul. Stefanowskiego 18/22
90-924 Lodz, Poland
Email: jaroslaw.goclawski@p.lodz.pl

***Abstract*—In this paper the problem of segmentation of volumetric medical images is considered. The fast and effective segmentation is obtained by applying the proposed approach which combines the idea of supervoxels and the Fuzzy C-Means algorithm. In particular, Fuzzy C-Means is used to cluster supervoxels produced by the fast 3D region growing. Additional acceleration of the method is achieved with the support of graphical processor (GPU). The detailed description of the proposed approach is given. The results of applying the method to volumetric CT and MRI brain images and CT images of various phantoms are presented, analysed and discussed. The issues related to accuracy of the method, memory workload and the running time are also considered.**

I. Introduction

ONE of the main challenges of recent medical image processing is the development of 3D image segmentation algorithms. These algorithms should be fast, efficient and accurate. Additionally, they should be easy to use and thus diminish the amount of user interaction required to extract the region of interest.

Although the problem of 3D image segmentation have been widely considered and numerous dedicated segmentation approaches have been proposed (e.g. [1], [2], [3]), it is still far from the satisfactory solution. This is caused mainly by the constant increase of the resolution of volumetric images acquired by computed tomography (CT) and magnetic resonance imaging (MRI) scanners. This manifests itself both, by the increase of the spatial resolution of single slices as well as the number of slices included into a scan. This in turn translates into the significant increase of the time and the memory workload required to perform segmentation of volumetric medical data.

Because of these reasons the existing approaches to image segmentation often cannot be directly used in everyday clinical routine. Therefore, recently a lot of effort have been put into the optimization and adaptation of popular segmentation approaches to fast and efficient processing of high resolution 3D medical images. This problem is also considered in this paper where Fuzzy C-Means (FCM) algorithm [4], [5] is adapted to segmentation of three dimensional images of brain. This is obtained by processing so called supervoxels (i.e. blocks of connected voxels of similar intensity) instead of single voxels. In particular, the input image is firstly divided into a number of supervoxels which are next clustered using FCM approach. This kind of processing significantly reduces the memory workload required to perform image segmentation.

It should be also mentioned, that supervoxels used in this paper extend the idea of superpixels, which has been known for few last years. However, the existing approaches to image division into blocks of pixels of similar intensity (e.g. watersheds [6], mean-shift [7], SLIC superpixels [8], Turbopixels [9] or FH superpixels [10]) are mainly dedicated to 2D images and their extension into the third dimension is not explicit or significantly increases time and memory workload and thus reduces benefit obtained due to processing blocks of pixels instead of single pixels. The method incorporated in this paper for creation of supervoxels is simple and straightforward. It is based on the fast and efficient region growing and thus can be directly adapted to three dimensional images [11].

Additionally, the graphical processor (GPU) support is also proposed to diminish the running time of both image division into supervoxels and FCM segmentation. What is more, the proposed approach reduces a user interaction to minimum, since only indication of one point is required to select the region of interest.

The following part of this paper is organised as follows. Firstly, in Section II the proposed approach is described in details. This is followed in Section III by the presentation and discussion of the results provided by the introduced method. These include both: the test performed on the CT and MRI brain scans, as well as the tests performed on various phantoms. Finally, Section IV concludes the paper.

978-83-60810-90-3/$25.00

II. The proposed approach

A. Building supervoxels in the image space

The main idea behind the introduced approach is to use Fuzzy C-Means algorithm to cluster supervoxels produced by the fast region growing [11], [12]. The supervoxels divide an image into blocks of similar intensity. Each of the supervoxels is built starting from a randomly selected seed voxel still not assigned to any region. The growing is performed with regard to the allowable difference in pixel intensities ΔI_{IMAX} and the maximum size V_{MAX} limiting the supervoxel size.

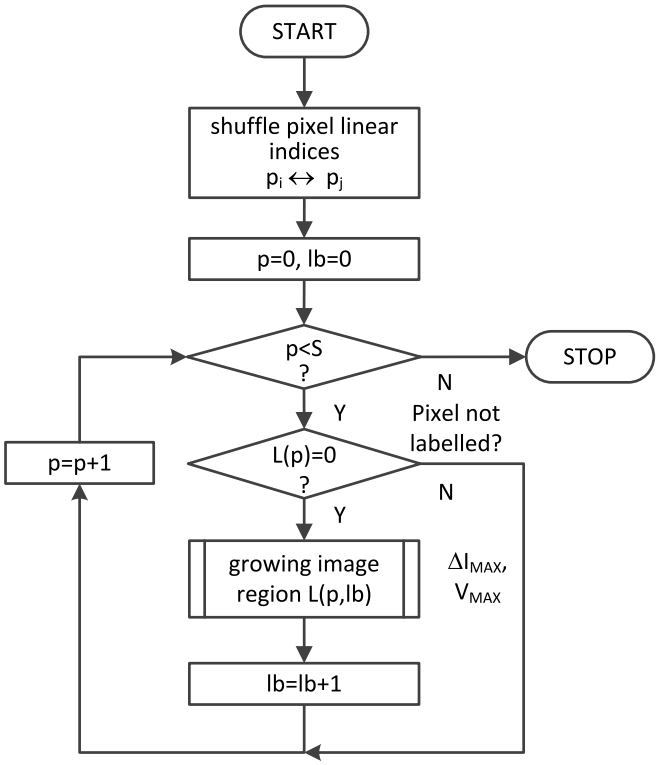


Fig. 1. Algorithm creating the label image L of three dimensional regions from the input intensity image I, L(p) - the label assigned to the voxel p ∈ [0,S), lb - the currently used label number.

The algorithm of image division into supervoxels shown in Figure 1 initially shuffles the voxel linear indices p of the input image I in the range $[0, S)$, where S is the number of voxels in the image. It speeds up the selection of unlabelled seeds. The indices randomly swapped in pairs can later be sequentially searched providing random selection of region seeds at the pixels which remain unlabelled. The procedure of growing image region around the seed p is shown in the Algorithm 1. It uses the queue Q_S of the next seed pixels and the queue of seed candidates Q_C. The queue Q_S is initially loaded with the primary seed pixel p, selected as shown in Figure 1.

For each pixel from Q_S the queue Q_C is created from its nearest 6 neighbours, assuming that their intensities fit in the range $I(p) \mp \Delta I$. Labelling the Q_C voxels increases the cumulated region volume V_R, explicitly limited to the value V_{MAX}. Exceeding volume limit V_{MAX} and empty voxel queue Q_S stops the region labelling process. Supervoxels are assigned label numbers as consecutive non negative integers. The supervoxel intensities are computed after the labelling process as the means of all original voxel intensities in the region.

Algorithm 1 The algorithm of limited region growing around the randomly selected non-labelled image pixel p

Input: I, lb, V_{MAX}, ΔI_{MAX}
Output: L

$Q_S \leftarrow p$, $Q_C \leftarrow \emptyset$
$V_R \leftarrow 0$
while $Q_S \neq \emptyset$ **do**
  $p \leftarrow Q_S$
  foreach $q \in N_B(p)$ **do**
    if $\Delta I < \Delta I_{MAX}$ **then**
      $L(q) \leftarrow lb$
      $Q_C \leftarrow q$
      $V_R = V_R + 1$
    end if
    if $V_R \geq V_{MAX}$ **then**
      return
    end if
  end foreach
  $Q_S \leftarrow Q_C$
end while

The results of image division into supervoxels are shown in Figure 2, where different colours represent different supervoxels. In particular, Figure 2a presents a sample CT brain slice, while the remaining subfigures show the corresponding slice after CT volume division into supervoxels of the increasing size V_{MAX}. All the results were obtained for the constant allowable difference in voxels intensity ΔI_{MAX} equal to 20. Due to the limited number of colours, they repeat for different supervoxels.

B. Fuzzy C-means specific solution

The segmentation of CT or MRI images with Fuzzy C-means (FCM) method [4] allows to assign any voxel with the intensity v_i, $i \in [1, N_V)]$ to a certain post-segmentation class (region) with the membership degree u_{ij}, $j \in [1, N_C)]$. The final deterministic assignment selects the class of the highest membership degree (probability) for each voxel v_i. FCM segmentation can be understood as an optimisation method minimizing the objective function G_m given in Equation (1).

$$G_m = \sum_{i=1}^{N_V} \sum_{j=1}^{N_C} u_{ij}^m \|v_i - c_j\|^2, \tag{1}$$

where $m > 1$, N_V is the image vector size, N_C is the given number of clusters (regions) of different intensities, u_{ij} is the membership degree of the voxel v_i to the region j, c_j denotes the intensity of j-th cluster centre and $\|\cdot\|$ represents the distance between the voxel v_i and the cluster centre c_j.

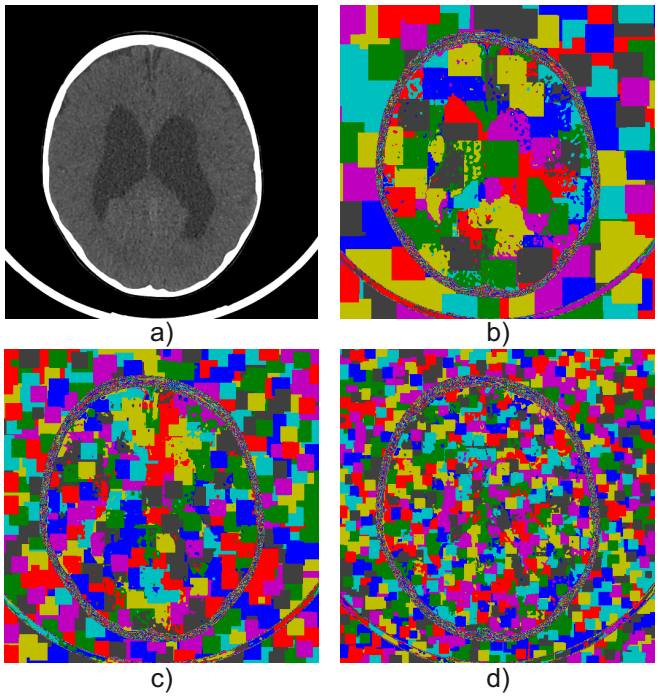


Fig. 2. The result of image division into supervoxels shown on a sample CT brain slice; a) original slice; b) ΔI_{MAX} = 20, V_{MAX} = 5000; c) ΔI_{MAX} = 20, V_{MAX} = 10000; d) ΔI_{MAX} = 20, V_{MAX} = 100000.

The execution of the FCM algorithm relays on the interactive correction of the class centres c_j based on the memberships u_{ij}. Next iteration class centres imply further modifications of u_{ij} until stopping u_{ij} changes or achieving the limit of iterations. Assuming the array of class centres $C = [c_j]$, $U = [u_{ij}]$ - the array of membership degrees and $m = 2$ the following computations are applied iteratively:

$$c_j^{(k)} = \frac{\sum_{i=1}^{N_V} u_{ij}^2 v_i}{\sum_{i=1}^{N_V} u_{ij}^2}, \tag{2}$$

$$u_{ij}^{(k)} = \left(\sum_{k=1}^{N_C} \frac{\|v_i - c_j\|^2}{\|v_i - c_k\|^2} \right)^{-1}, \tag{3}$$

The iteration stops when $\|U^{(k)} - U^{(k-1)}\|_\infty < e_{MAX}$. The number of intensity classes N_C must be defined a priori.

In the case of 3D CT or MRI images including tens of millions of voxels the FCM clustering process may be unacceptable slow. Therefore the authors proposed its acceleration in two ways:

- by using smaller number of supervoxels,
- by applying parallel computations with GPU.

The pseudocede of FCM algorithm with CUDA GPU computing [13], [14] is presented in Algorithm 2. The command *par foreach* built inside means a *for* loop executed in parallel by GPU processors. Capital letter symbols in the code refer to host data buffers, bold font of capital letters denotes arrays

Algorithm 2 FCM segmentation algorithm using GPU parallel computing. N_V –the number of supervoxels, N_B –the number of GPU memory blocks, N_t –the number of GPU threads per block, N_C –the number of classes, b –the current block number, t –the thread number in a block, sync() –the thread synchronization.

Input: $V[N_V], N_C, e_{MAX}, k_{MAX}$
Output: $C[N_C], U[N_V \times N_C]$

$\mathbf{V} \leftarrow V$
$\mathbf{U}\_ \leftarrow random([N_V \times N_C])$
$\mathbf{U}[N_V \times N_C] \leftarrow \{0\}$
$alloc\, \mathbf{P}[N_B \times N_C]$, $\mathbf{Q}[N_B \times N_C]$
$alloc\, \mathbf{C}[N_C]$
$k \leftarrow 0$

repeat
  $swap\_addresses(\mathbf{U}, \mathbf{U}\_)$
▷ *kernel function #1*
  par foreach $i \in [0, N_V)$ **do**
    $alloc\, \mathbb{S}[N_t \times N_C]$
    $\forall j \in [0, N_C),\ \mathbb{S}(t, j) \leftarrow \mathbf{U}(i, j)^2 \cdot \mathbf{V}(i)$
    $sync()$
    $\forall j \in [0, N_C),\ \mathbf{P}(b, j) \leftarrow reduction(\mathbb{S}(t, j))$
    $\forall j \in [0, N_C),\ \mathbb{S}(t, j) \leftarrow \mathbf{U}(i, j)^2$
    $sync()$
    $\forall j \in [0, N_C),\ \mathbf{Q}(b, j) \leftarrow reduction(\mathbb{S}(t, j))$
  end foreach
▷ *kernel function #2*
  par foreach $j \in [0, N_C)$ **do**
    $\mathbf{C}(j) \leftarrow \sum_{b \in [0, N_B)} (\mathbf{P}(b, j)) / \sum_{b \in [0, N_B)} (\mathbf{Q}(b, j))$
  end foreach
▷ *kernel function #3*
  par foreach $i \in [0, N_V)$ **do**
    $v \leftarrow \mathbf{V}(i)$
    foreach $j \in [0, N_C)$ **do**
      $w \leftarrow \|v - \mathbf{C}(j)\|^2$
      $s \leftarrow 0$
      $\forall k \in [0, N_C)\ s \leftarrow s + \|v - \mathbf{C}(k)\|^{-2}$
      $\mathbf{U}(i, j) \leftarrow 1/(w \cdot s)$
    end foreach
  end foreach
▷ *kernel function #4*
  $e \leftarrow \|\mathbf{U} - \mathbf{U}\_\|_\infty$
  $k \leftarrow k + 1$
until $(e < e_{MAX}) \vee (k \geq k_{MAX})$
$C \leftarrow \mathbf{C}$
$U \leftarrow \mathbf{U}$

allocated in the GPU memory, double stroked font symbols represent GPU shared memory arrays.

The algorithm input data are: the supervoxel vector image $V[N_V]$, the assumed number of intensity classes N_C, the norm e_{MAX} of maximum acceptable error and k_{MAX} – the maximum number of iterations for Equation (2) and Equation (3). The output data consists of the host vector $C[N_C]$ of output class centres and the host array of each supervoxel membership degree $U[N_V \times N_C]$. The $\mathbf{U}$ array is at first allocated in the graphic card memory and randomly initialized in the probability range of $[0, 1]$. The GPU memory also includes local temporal arrays $\mathbf{P}$ and $\mathbf{Q}$ to store component parts of the numerator and denominator in Equation (2) corresponding to N_B blocks, each of the size $N_t = 256$ GPU threads.

The kernel function #1 for each thread computes the monomials in the numerator of Equation (2) and copies them to the GPU shared memory $\mathbb{S}$ to later evaluate partial sums $\mathbf{P}$ and $\mathbf{Q}$ by the process of reduction.

The kernel function #2 adds the partial sums and computes for each thread the values c_j given in Equation (2).

The kernel function #3 completely fulfils the formula in Equation (3), because it sums the relatively small number N_C of intensity classes.

The segmentation error e in a current computing cycle is evaluated as the maximum distance norm of supervoxel membership degrees in two successive iterations. The error value is copied to the host memory to make the decision of stopping iterations.

The array of membership degrees is allocated in the GPU memory in two copies $\mathbf{U}$ and $\mathbf{U\_}$, which addresses are swapped instead of copying data between $\mathbf{U}$ and $\mathbf{U\_}$ in every iteration cycle (Algorithm 2, line 8).

The FCM output array U includes N_C column images of the degrees of membership to a particular class. The class of the highest membership in each row of U is assigned to the output vector of intensity supervoxel classes as in Equation (4).

$$V(i) = \max_j(U(i,j)),\ i \in [0, N_V),\ j \in [0, N_C). \quad (4)$$

This output vector image $V[N_V]$ is then reshaped to the original matrix form $I[Y \times X \times Z]$ after its reverse mapping from supervoxels to voxels. In the image I of N_C intensity classes (labels) only a single region is identified, which belongs to a certain class determined by the voxel marker that was selected interactively. The identification can be fulfilled by the flood fill spatial expansion covering the whole region around the marked voxel.

The complete algorithm sequence of brain image segmentation starts with GPU averaging lowpass filter to reduce data noise coming from the CT or MRI acquisition systems. The noise is gained in particular when setting low doses of radiation during brain examinations in children. The filter fulfils the formula given in Equation (5).

$$J(x,y,z) = \frac{1}{UVW} \sum_{u=-\frac{U}{2}}^{\frac{U}{2}} \sum_{v=-\frac{V}{2}}^{\frac{V}{2}} \sum_{w=-\frac{W}{2}}^{\frac{W}{2}} I(x+u, y+v, z+w), \quad (5)$$

where $[V \times U \times W]$ – the cube of image data averaging with the odd numbers U, V, W. The segmentation is finalized with the operation of morphological opening (Equation (6)).

$$J_B = (I_B \ominus S(R_X, R_Y, R_Z)) \oplus S(R_X, R_Y, R_Z), \quad (6)$$

where $S(R_X, R_Y, R_Z)$ denotes an ellipsoidal structuring element with the radii R_X, R_Y, R_Z in the particular space directions, I_B and J_B are the input and output binary images of a selected region in the original brain image. The ellipsoidal structuring element S represents an ellipsoid mask mapped into the discrete space of image voxels with the radii R_X, R_Y, R_Z respectively in X, Y, Z space directions. Using the ellipsoid instead of the ball shape enables mapping different image resolutions in space directions into the voxel space (in particular, the spacing between slices). Post processing morphological operations allow controlled smoothing of the borders in the extracted brain region.

III. Results

A. Tests on brain images

The results of applying the proposed segmentation approach to sample volumetric CT and MRI brain images are shown in Figures 3 and 4 respectively. In both figures the top panel shows the original brain slices with the region of interest indicated by the green square marker. In the middle panel the corresponding segmentation results overlaid on the input slice are presented. Finally, the segmentation results are visualised in 3D in the bottom panel. In both figures, cases are numbered from the left to the right. In the case of CT images the image division into supervoxels was performed for the maximal volume of supervoxel equal $V_{MAX} = 7500$ and maximal intensity difference $\Delta I_{MAX} = 40$. For FCM segmentation the number of classes was set to $N_C = 6$, while maximal number of iterations k_{MAX} was set to 400 (with $e_{MAX} = 0.001$). In the case of MRI images image division into supervoxels was performed for $V_{MAX} = 5000$ and $\Delta I_{MAX} = 40$. The number of classes N_C ranged from 8 (for case 1) to 6 (for the remaining cases). As previously, the maximal number of iterations k_{MAX} was set to 400 (with $e_{MAX} = 0.001$). In both cases the parameters were tuned manually, to obtain the subjectively best results. Prior to segmentation all images were subjected to preprocessing (Gaussian filtration).

The time and memory workload required to perform image segmentation in the considered CT and MRI datasets is summarised in Table I and Table II respectively. In both tables the first column indicates the case ID. This is followed by image resolution given in the second column. The third

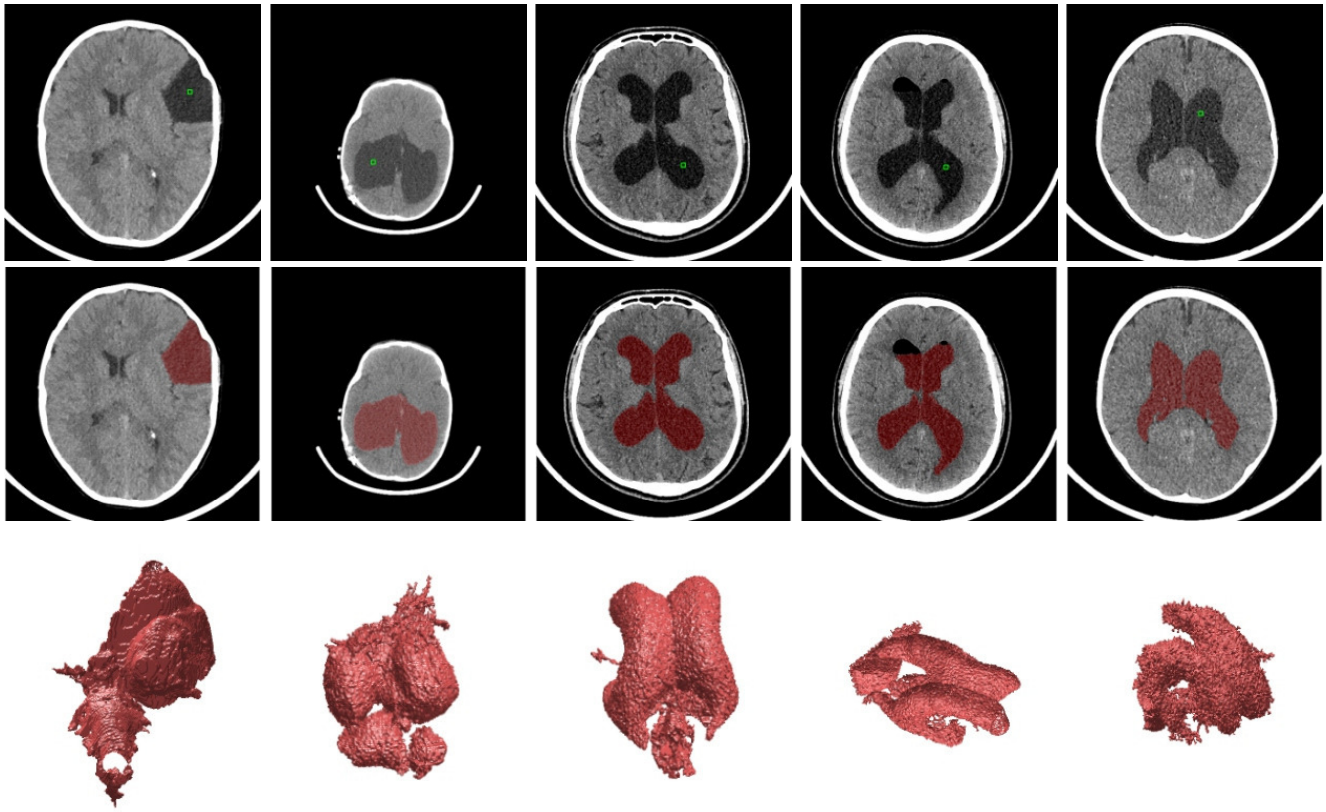


Fig. 3. The results of applying the proposed approach to sample CT brain images; top row - original images with a region of interest indicated by green square marker; middle row - the results overlaid on a sample slices; bottom row - the results shown in 3D. Cases are numbered from left to right.

column shows the level of data reduction r due to image division into supervoxels. In particular, it means that the number of supervoxels was $r\%$ lower than the number of voxels. The fourth column shows the GPU memory workload given in MB, while columns five, six and seven present time T_0 of preprocessing (filtration), time T_1 of image division into supervoxels and time T_2 of FCM execution respectively. All times are given in miliseconds. The tests were performed on a computer with Intel Core i7 3.6 GHz processor, 32 GB RAM memory and graphical card Nvidia GeForce Titan (6 GB).

TABLE I
THE TIME AND MEMORY WORKLOAD OF THE PROPOSED METHOD - THE RESULTS FOR CT DATASETS.

ID	Image size [px.]	r [%]	GPU [MB]	T0 [ms]	T1 [ms]	T2 [ms]
1	512×512×115	95.02	140.9	426.2	5410.2	560.5
2	512×512×112	96.59	80.4	435.3	9313.5	288.8
3	512×512×216	97.03	314.0	548.9	15366.1	818.9
4	512×512×220	97.24	293.9	538.8	18518.3	1087.7
5	512×512×202	97.34	344.4	544.6	17022.1	1103.9

From the Figures 3 and 4 it can be seen that the proposed approach was successful in segmenting indicated regions of interest both CT and MRI images. Based on visual assessment it can be concluded that the borders of ventricular fluid and cysts were properly determined and details of its shape were captured by the image segmentation algorithm. The accuracy of object shape determination is sufficient for further quantitative analysis.

TABLE II
THE TIME AND MEMORY WORKLOAD OF THE PROPOSED METHOD - THE RESULTS FOR MRI DATASETS.

ID	Image size [px.]	r [%]	GPU [MB]	T0 [ms]	T1 [ms]	T2 [ms]
1	512×448×25	88.15	88.6	0	1750.5	446.5
2	512×512×22	88.97	83.3	311.3	1326	406.4
3	512×512×21	86.23	75.5	323.8	1405.4	403.5
4	256×256×23	88.54	19.0	306.7	340.2	101.9
5	256×256×20	87.10	18.5	0	360.7	89.1

In the case of CT images it was possible to run the algorithm with a uniform setting of parameters, especially the number of classes N_C considered during FCM segmentation. In the case of MRI images, to obtain better quality results it was necessary

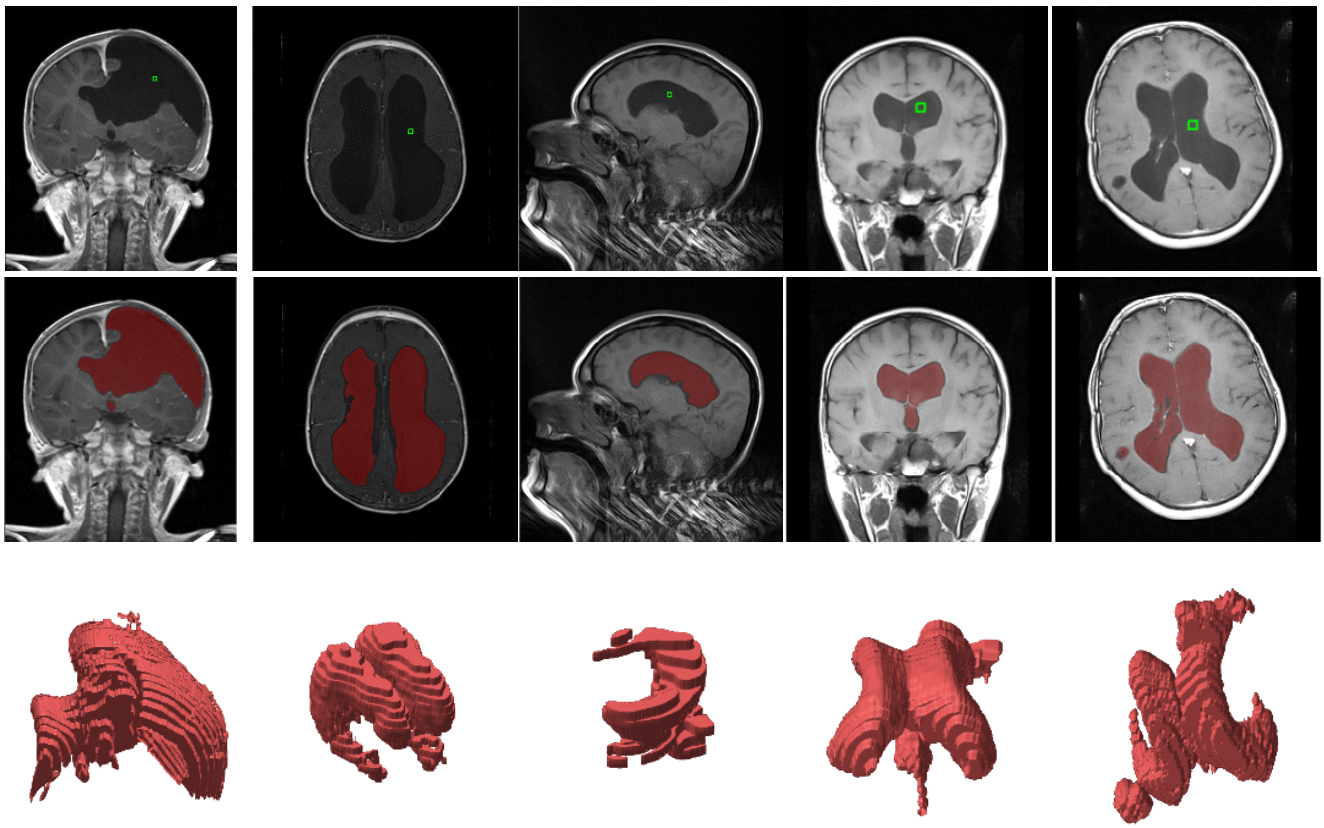


Fig. 4. The results of applying the proposed approach to sample MRI brain images; top row - original images with a region of interest indicated by green square; middle row - the results overlaid on a sample slices; bottom row - the results shown in 3D. Cases are numbered from left to right.

to tune the number of classes in the case of one considered dataset. Additionally, in cases 1 and 5 preprocessing was not applied since it deteriorated segmentation results.

From the Tables I and II it can be seen, that the proposed approach is fast and efficient. For all the considered cases the running time of the method wasn't longer than 20 seconds including preprocessing, image division into supervoxels and FCM segmentation. This is a very good result, regarding that the considered datasets consisted of up to 220 slices of resolution 512×512 pixels each. The corresponding time of FCM segmentation performed with respect to single voxels last hours. This acceleration is obtained due to significant reduction of clustered data. In the considered cases, processing of supervoxels instead of voxels diminished the number of data points clustered by FCM by on average 96.6% for CT images and 88.7% in the case of MRI images.

The memory workload of the proposed approach was also significantly reduced due to application of supervoxels. When run on the original datasets the FCM required several gigabytes of RAM memory to cluster all the voxels. As it can be seen from Tables I and II the improved method in the worst case used less than 350 MB of GPU RAM memory. In the case of volumetric images this is a very good results which makes the method available for a common use on a standard PC computer.

TABLE III
THE INFLUENCE OF SUPERVOXELS AND GPU COMPUTING ON THE FCM SEGMENTATION TIME IN CT DATASETS. T2 - THE EXECUTION TIME WITH SUPERVOXELS AND GPU, T2' - NO SUPERVOXELS, T2" - NO SUPERVOXELS OR GPU.

ID	Image size [px.]	T2 [ms]	T2' [ms]	T2" [ms]
1	512×512×115	560.5	9594	186873
2	512×512×112	288.8	7301	147686
3	512×512×216	818.9	21934	540873
4	512×512×220	1087.7	19562	595971
5	512×512×202	1103.9	24196	552163

In tables III and IV different variants of FCM segmentation times are compared for CT and MRI images respectively. The column $T2$ refers to the proposed method applying both supervoxels and parallel GPU computing. Two next columns describe the same cases omitting supervoxel creation ($T2'$) and also GPU computing ($T2''$). The exclusion of supervoxels increases CT segmentation time $17 \div 27$ times (on average

TABLE IV
THE INFLUENCE OF SUPERVOXELS AND GPU COMPUTING ON THE FCM SEGMENTATION TIME IN MRI DATASETS. T2 - THE EXECUTION TIME WITH SUPERVOXELS AND GPU, T2' - NO SUPERVOXELS, T2" - NO SUPERVOXELS OR GPU.

ID	Image size [px.]	T2 [ms]	T2' [ms]	T2" [ms]
1	512×448×25	446.5	9813	84817
2	512×512×22	406.4	11576	90933
3	512×512×21	403.5	9126	119840
4	256×256×23	101.9	2652	26286
5	256×256×20	89.1	3198	28595

$\approx$ 22 times) and additional removing of GPU computing extends it up to $330 \div 660$ times (on average ≈ 510 times). For the MRI images the FCM execution is slowed down on average 27 and 258 times in the two algorithm variants with $T2'$ and $T2''$. The acceleration with supervoxels is achieved at the expense of their preparation time. Therefore real speed-up of the algorithm in this approach is only 1.2 and 28 times for the tested CT images or respectively 6 and 56 times for MRI cases. The example CT images of larger size than the example MRI require more effort to prepare supervoxels. Hence, their use only gives less profit on time, which can be still enhanced during subsequent parallel computations.

B. Tests on phantoms

The accuracy of the introduced supervoxels based FCM segmentation approach was assessed using three different physical phantoms including: phantom of head, phantom CIRS 045 of prostate [15] and phantom CIRS 062 [16] used for calibration of computed tomography scanners. All these phantoms were scanned using a CT scanner. The resulting images were next subjected to segmentation using the proposed approach. The aim of the segmentation was to extract characteristic objects contained within phantoms. Finally, the volumes of these objects were determined based on segmentation results and compared with the volumes given in phantom specification.

The phantom of head was prepared using the 3D printing technique. The phantom consists of two main parts: the skull made from a material of the density 2.1-2.3 g/cm^3 and the brain made from a material of the density 1.02-1.04 g/cm^3. In the brain there is a hole of a shape similar to ventricular system and volume equal to 41.69 cm^3. The hole was filled with water imitating the cerberospinal fluid. The head phantom is presented in Figure 5. In particular, Figure 5a shows the general view of the phantom, while Figure 5b presents a sample CT slice. Additionally, in Figure 6 the shape of the ventricular system of the phantom is presented.

The phantom CIRS 045 is dedicated to prostate brachytherapy. Inside it contains three cysts of the increasing volumes (namely: 4 cm^3, 9 cm^3 and 20 cm^3). The general view of the phantom is shown in Figure 7a, while a sample CT scan is presented in Figure 7b.

Finally, the phantom CIRS 062 is shown in Figure 8. The phantom consists of two rings with inclusions of equal size.

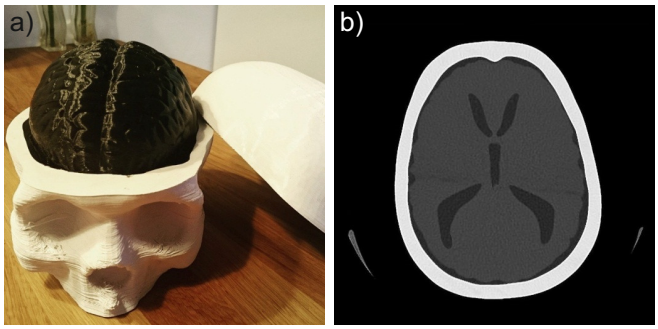


Fig. 5. The brain phantom; a) the general view; b) a CT scan.

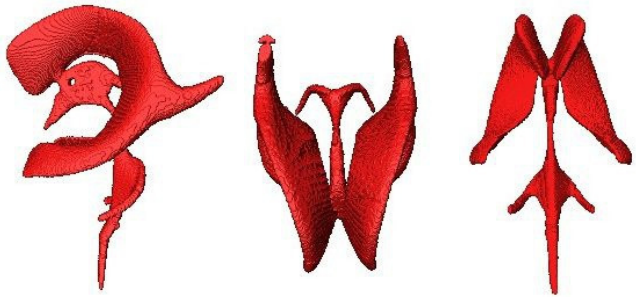


Fig. 6. The shape of the ventricular system within the brain phantom.

Each inclusion corresponds with different tissue and thus exhibit different attenuation of X-rays (see Fig. 8b).

The results of applying the introduced segmentation approach to images of phantoms are shown in Figure 9. In the case of head phantom, the proposed approach was used to extract the ventricular system (see Fig. 9a). From prostate phantom CIRS 045 the largest cyst was extracted (see Fig. 9b). Finally, from the phantom CIRS 062 a randomly selected inclusion was extracted using the proposed approach (see Fig. 9c). For each case, the segmentation result is both: shown in 3D and overlaid on a sample slice.

The results of comparison between the real and the determined volumes of the considered regions are summarised in Table V. In particular, the considered phantom is indicated in the first column. The real region volume V_0 is given in the fifth column, and followed by the determined volume V in the sixth column. The corresponding relative error of volume

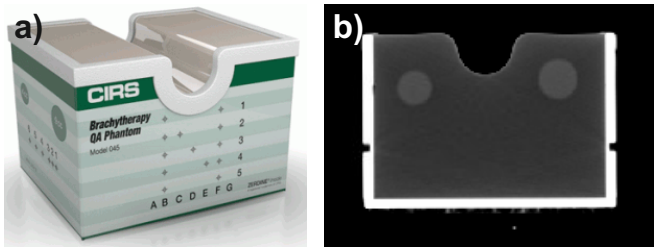


Fig. 7. Phantom CIRS 045; a) the general view (©CIRS Tissue Simulation & Phantom Technology); b) a sample slice from a CT scan.

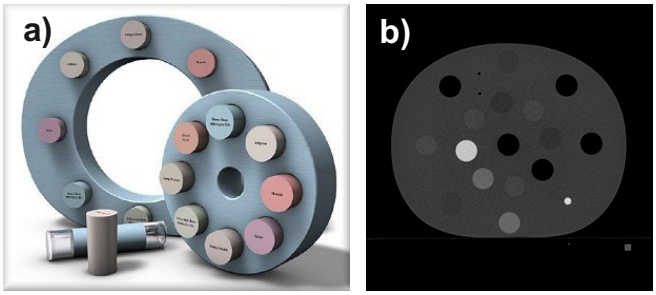


Fig. 8. Phantom CIRS 062; a) the general view (©CIRS Tissue Simulation & Phantom Technology); b) a sample slice from CT scan.

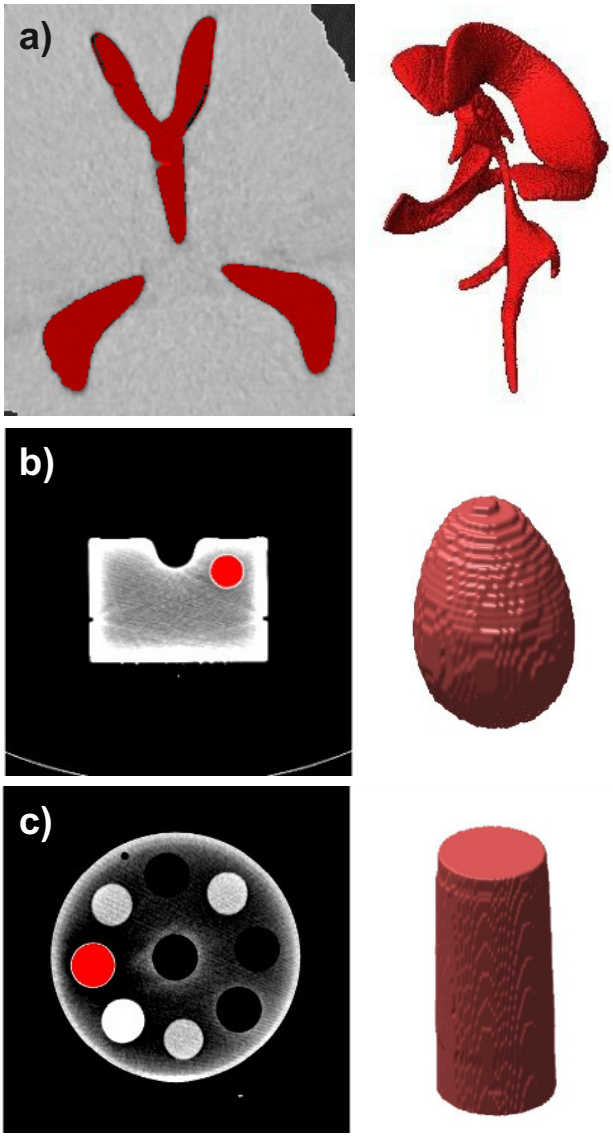


Fig. 9. The results of supervoxel based FCM segmentation applied to the considered phantoms; a) brain phantom; b) CIRS 045; c) CIRS 062.

determination is shown in the last column. Additionally, for each phantom image resolution, time of image division into supervoxels T_0 and time of FCM segmentation T are given in columns two, three and four respectively.

TABLE V
THE ASSESSMENT OF SEGMENTATION ERROR.

phantom	Image size [px.]	T0 [s]	T [s]	V0 [cm^3]	V [cm^3]	dV [%]
head	512×512×515	36.88	1.66	41.69	40.77	-2.20
CIRS 045	512×512×61	3.85	0.42	20.00	17.44	-12.80
CIRS 062	512×512×55	3.46	0.41	42.44	53.75	1.41

From Figure 9 it can be easily seen that the proposed segmentation approach successfully extracted objects of interest from the considered phantoms. The shape of the segmented objects correspond with the real one. This is especially visible in the case of the ventricular system segmented from the phantom of head (see Fig. 9a). In this case all important anatomical details are clearly visible in object after segmentation. The shape of objects extracted from the CIRS phantoms also correspond with the description given in phantoms specification.

The relative error of volume determination equals on average 5.5% (see Tab. V). It is worth highlighting, that in the case of the ventricular system extracted from the phantom of head the error is only -2.2%. This is a very good result, especially having in mind the complex shape of the ventricular object.

IV. CONCLUSIONS

In this paper the fuzzy C-means method FCM) was specifically adapted to the segmentation of volumetric brain images. In particular CT images include from tens to hundreds of millions of voxels to analyse within a reasonable time. Although the FCM method has the advantage of the same functionality in two and three dimensions, the algorithm execution time is proportional to the number of image voxels and can exceed large values for hundreds of CT slices. The proposed reduction of input data size based on supervoxels varies from 85% to 95% in the tested examples at the expense of several seconds for creating supervoxel regions. For the assumed parameters the segmentation accuracy of about 2% tested for the head phantom still remains acceptable. The above data compression and GPU parallel computing limit processing time to single seconds, practically without any extra cost for hardware equipment. Standard CUDA compatible NVIDIA graphic card of computing capabiility $3 \div 3.5\%$ will be sufficient in this case, because the GPU memory workload falls below $0.5\,GB$ for 200 CT slices. The bottle neck of such graphic computing is the maximum acceptable size of shared memory per block (up to $8\,kB$) implying the limited class number to preserve the parallel computation efficiency. Additionally in some medical cases establishing a priori the number of segmentation classes to extract desired objects is a very intuitive task. Nevertheless the proposed approach make the modified FCM method very promising for its very fast processing of bulk data and dimensionality independence.

ACKNOWLEDGEMENTS

The authors would like to thank Mikołaj Kopernik's Hospital in Lodz (Poland) for providing CT images of Electron Density Phantom CIRS 045 and Brachytherapy Phantom CIRS 062.

REFERENCES

[1] D. D. Burdescu, L. Stanescu, M. Brezovan, C. S. Spahiu, "Efficient Volumetric Segmentation Method", *Proceedings of the 2014 Federated Conference on Computer Science and Information Systems*, pp. 659–668, 2014, http://dx.doi.org/10.15439/978-83-60810-58-3

[2] K. Xiao, A. E. Hassanien, N. I. Ghali, "Medical Image Segmentation Using Information Extracted from Deformation", *Proceedings of the 2011 Federated Conference on Computer Science and Information Systems*, pp. 157–163, 2014.

[3] M. R. Ogiela, T. Hachaj, "Automatic segmentation of the carotid artery bifurcation region with a region-growing approach", *Journal of Electronic Imaging*, vol. 22(3), 033029, 2013, http://dx.doi.org/10.1117/1.JEI.22.3.033029

[4] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, New York, 1981.

[5] R. Nock, and F. Nielsen, "On weighting clustering", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28(8), pp. 1–13, 2006, http://dx.doi.org/10.1109/TPAMI.2006.168

[6] C. Couprie, L. Grady, L. Najman, and H. Talbot, "Power watershed: A unifying graph-based optimization framework", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33(7), pp. 1384–1399, 2011, http://dx.doi.org/10.1109/TPAMI.2010.200

[7] W. Tao, H. Jin, and Y. Zhang, "Image segmentation based on mean shift and normalized cuts", *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 37(5), pp. 1382–1389, 2007, http://dx.doi.org/10.1109/TSMCB.2007.902249

[8] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk,"SLIC superpixels compared to state of the art superpixel methods", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34(11), pp. 2274–2282, 2012, http://dx.doi.org/10.1109/TPAMI.2012.120

[9] A. Levinshtein, A. Stere, K. Kutulakos, D. Fleet, S. Dickinson, and K. Siddiqi,"Turbopixels: Fast superpixels using geometric flows" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31(12), pp. 2290–2297, 2009, http://dx.doi.org/10.1109/TPAMI.2009.96

[10] P. F. Felzenszwalb, and D. P. Huttenlocher, "Efficient graph-based image segmentation" *International Journal of Computer Vision*, vol. 59(2), pp. 167–181, 2004, http://dx.doi.org/10.1023/B:VISI.0000022288.19776.77

[11] A. Fabijańska, and J. Gocławski, "The segmentation of 3D images using the random walking technique on a randomly created image adjacency graph" *IEEE Transactions on Image Processing*, vol. 24(2), pp. 524–537, 2015, http://dx.doi.org/10.1109/TIP.2014.2383323

[12] W. Pratt, *Digital Image Processing*, 4th ed. Los Altos, California: John Wiley & Sons Inc., 2007.

[13] J. Sanders and E. Kandrot, *Cuda by Example. An Introduction to General Purpose GPU programming*, NVIDIA Coropration, 2011.

[14] N. Wilt, *The CUDA Handbook. A Comprehensive guide to GPU Programming*, Addison-Wesley, Upper Saddle River, NJ, 2013.

[15] Brachytherapy QA Phantom CIRS 045, http://www.cirsinc.com/products/all/71/brachytherapy-qa-phantom

[16] Electron Density Phantom CIRS 062, http://www.cirsinc.com/products/all/24/electron-density-phantom

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 787–790
DOI: 10.15439/2016F472
ACSIS, Vol. 8. ISSN 2300-5963

A compact deep convolutional neural network architecture for video based age and gender estimation

Bartłomiej Hebda
AGH University of Science and Technology
Krakow, Poland
E-mail: hebda.bartlomiej@gmail.com

Tomasz Kryjak, *Member IEEE*
AGH University of Science and Technology
Krakow, Poland
E-mail: tomasz.kryjak@agh.edu.pl

***Abstract*—In this paper research on a compact deep convolutional neural network (DCNN) architecture for age and gender estimation from facial images has been presented. The proposed solution was tested on the FERET and the Adience Benchmark databases. In the first case a 98.6% accuracy for gender and 86.4% for age estimation was obtained. For the Adience database, which contains images recorded in unconstrained conditions and is much more demanding, a 62.0% for gender and 42.0% for age accuracy was obtained. When compared to the reference results on a much larger network, the performance should be considered as satisfactory. The research shows that a compact DCNN with small input images can provide quite good classification results.**

I. Introduction

A VISION system which allows the estimation of age and gender of a person using a face image can have a number of important practical applications in biometric and statistics systems, as well as advanced human-computer interfaces (HCI) and so-called smart advertising (with personalized content). It typically consists of a face detection [1] and the actual estimation module. The designed vision system must be resistant to changes in face appearance (facial expressions, hairstyle, presence of beard or moustache, glasses and to some extent also make-up), different lightening conditions, inaccurate face localization in the image, face orientation (frontal, from profile and also rotated), size of the input image and it's quality (presence of noise, blur caused by movement of the person, underexposure, overexposure, shadows, etc.).

Deep convolutional neural networks (DCNN) are one of the most interesting tool available for the image processing and machine learning community in recent years. It is worth noting that the approach in not new – the first concepts were already proposed in the 70's of the last century [2]. However, due to limited performance of the available computing platforms, these solution were not used. The breakthrough came with the emergence of programmable graphic processing units (GPU). They proved to be an almost perfect platform for neural networks implementation, mainly because of the massive parallelization possibility and floating point support. At the same time learning methods for such complex structures were developed and refined. In addition, the dynamic growth of social networks services like Facebook, Flickr or Instagram allowed to obtain quite easy access to huge image databases. Currently DCNNs are used for almost all machine learning tasks – from speech recognition through image recognition to information about social networks users categorization. Only in computer vision the following applications should be mentioned: face detection, pedestrian detection, road sign detection and recognition and object tracking. What is more, the DCNN based approaches usually significantly outperform the "classic" (i.e. feature extraction and classification – e.g. HOG + SVM) ones.

In this paper, a compact DCNN for age and gender estimation from facial images is presented. The obtained results indicate, that it is possible to use a quite small, energy and resource efficient architecture, without much loss on classification performance. Moreover, all experiments were performed on a typical PC computer, without a powerful GPU accelerator.

The reminder of this paper is organized as follows. In Section II a brief review of age and gender estimation algorithms is presented. The proposed solution is described in Section III. Then, in Section IV the evaluation results are presented and discussed. The paper ends with a summary and indication of future research directions.

II. Age and gender estimation systems

Age and gender estimation is a quite popular topic in the computer vision community. An in-depth review is far beyond the scope of this article and therefore only selected works directly related to DCNNs are presented.

A. DCNN based solutions

One of the earlier works on gender recognition using CNNs was presented in paper [3]. The solution consisted of a face detection and gender recognition module – both using neural networks. The architecture involved three layers (two hidden and output). Input images had a 32×32 pixels resolution. The reported accuracy on the FERET dataset was 97.2%.

Most of the work related to the DCNNs appeared in 2015 and later. In the article [4] two approaches were compared: "classic" and DCNN based. In the first case the following

978-83-60810-90-3/$25.00

features were considered: HOG (Histogram of Oriented Gradients), LBP (Local Binary Patterns) and SURF (Speeded Up Robust Features). As regression the CCA (Canonical Correlation Analysis) was applied. In the second, many different variants of network architectures were examined (the Caffe library was used). The best result were obtained for two convolutional and one fully connected layer. Input images had 50×50 pixels size. The authors noted a significant disproportion between the time required for learning and actual operation in both cases. Finally, for the MORPH database, the "classic" solution obtained 4.25 and DCNN 3.88 mean absolute error (MAE) value.

In the work [5] a DCNN for age and gender estimation was proposed. The network had three convolutinal and two fully connected layers. Input images of size 256×256 were cropped to 227×227. The authors did not use a pre-trained network model – in contrast to many other approaches. On the Adience dataset this solution achieved 86.8% $\pm$ 1.4 accuracy for gender and 50.5% $\pm$ 5.1 for age estimation. In the latter case 8 age categories were used. If an "off by one" error is allowed, the performance increases to 84.7 $\pm$ 2.2. In this study the Caffe library and GPU with 1,536 CUDA cores and 4 GB of RAM were used.

In 2015, at the IEEE International Conference on Computer Vision Workshop (ICCVW) a competition on age estimation (ChaLearn) was conducted. The Looking at People 2015 (LAP) dataset with 4,961 images was used [6]. The DEX system from ETH Zurich, Switzerland obtained the best results [7]. In the first step the face was detected. Then, it was rescaled to 256×256 pixel size. The authors used a pre-trained DCNN on the ImageNet dataset. In addition, it was fine-trained on the IMDB-WIKI database (260,282 images) and finally on the LAP set. In total 20 separate networks were used. This allowed to achieve 3.21 MAE value. The Caffe framework and NVIDIA Tesla K40C GPU were used. Learning a single network lasted 5 days, fine-learning 3 hours and processing a single image about 200 ms.

This short (compared to the available material) overview can be concluded with the following statement: "age and gender estimation with DCNNs is very accurate (sometimes even better than human), however the used network architectures are very complex". Therefore, their training and usage requires a lot of computing resources – most authors used specialized and expensive GPU accelerators like NVIDIA K40.

B. Our previous solution

During our earlier work [8], a "classical" age, gender and race estimation system was created. In the pre-processing stage the face was detected (Viola-Jones method), eyes and mouth were detected (also by the Viola-Jones approach) and image alignment was performed – an affine transform was used to assure that eyes and mouth are at predefined locations for every considered image. The local binary patterns (LBP) were used as features. After their computation, the image was divided into non-overlapping blocks in which histograms were computed – their concatenation formed the feature vector. The support vector machine (SVM) with radial basis functions (RBF) was used as classifier. In case of non-binary problems (race, age) multiple SVMs were used – one for each class.

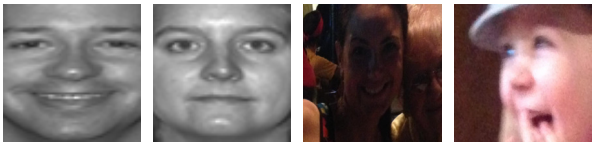


Figure 1: Samples from the FERET (a) and Adience (b) databases

In the experiments on the FERET database quite good performance (94.0% for gender, 86,7% for race and 69.5 % for age in 10 years interval) was obtained. Unfortunately, real-life tests in unconstrained conditions on images acquired with a standard USB webcam revealed the weakness of this solution. Actually, even for gender determination it was quit difficult to get a proper result. This was a direct motivation to improve the vision system and use the DCNN approach.

III. The proposed solution

In this work the impact of simplifying the network structure, especially the size of the input image, on the overall accuracy of age and gender recognition was evaluated. Also the computing resources and single image processing time were considered. The ultimate goal of the ongoing research is to implement DCNN solutions in embedded devices and support real-time video stream processing.

A. The used datasets

In the experiments two databases were used: FERET [9] and Adience [10]. The first was chosen to compare the obtained results with our previous work [8]. It includes 2,413 images of 856 individuals. Examples are shown in Figure 1a. The biggest drawback is that these pictures were taken under controlled conditions. This results in poor generalization of the trained classifier, especially for cases registered under real-life conditions. In addition modern methods obtain almost 100% accuracy for this set, which indicates that it is not longer an appropriate challenge.

The Adience set was used because it is considered as challenging and the images were registered in unconstrained environment. It contains 26,580 annotated images of 2,284 people. Samples are shown in Figure 1b.

B. The used hardware and software

All experiments were performed on mid-range laptop – Intel Core i5-2410M (2.3 GHz up to 2.9 GHz), 4 GB DDR3 RAM, NVIDIA GeForce GT 540M with 2 GB DDR3 RAM, 96 CUDA cores and 672 MHz core frequency. There are several computing libraries for DCNNs – eg. Caffe, Torch and Google Tensor Flow. After a preliminary analysis, Torch (`http://torch.ch/`) was selected, mainly due to easier installation and configuration (in comparison to Caffe), as

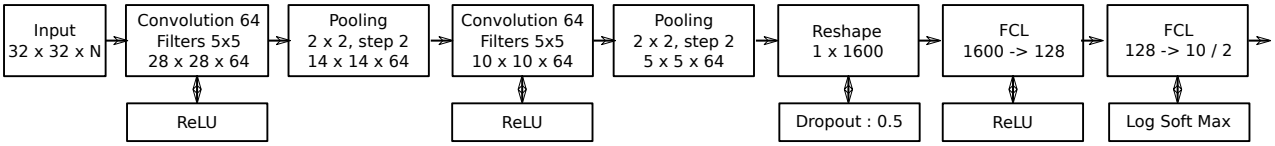


Figure 2: The used DCNN architecture

Table I: Results for the FERET (F) and Adience (A) dataset

	Conv. 1	Conv. 2	# Training	# Test	# Epoch	Acc. gender	Acc. age	Computing time
F1	5×5	5×5	100	100	100	81.3%	41.2%	48s
F2	5×5	5×5	250	100	200	96.0%	70.1%	181s
F3	5×5	5×5	500	200	300	88.5%	74.5%	534s
F4	5×5	5×5	1000	500	500	93.7%	58.9%	1795s
F5	5×5	5×5	2000	500	500	**98.6%**	85.4%	3324s
F6	3×3	5×5	2000	500	200	97.6%	**86.4%**	1572s
F7	7×7	5×5	2000	500	200	97.6%	85.6%	1683s
F8	3×3	3×3	2000	500	200	98.0%	**86.4%**	1354s
F9	7×7	7×7	2000	500	200	97.6%	85.8%	1709s
A1	5×5	5×5	2000	500	300	50.2%	42%	2317 s
A2	5×5	5×5	12000	500	200	51.4%	32.1%	3h
A3	5×5	5×5	12000	500	200	59.5%	33.1%	3h
A4	5×5	5×5	12000	500	200	**62.0%**	**42.0%**	3h

well as a more convenient to manage network model and the whole project (LUA language). Additionally, in our opinion the available documentation is very extensive and there are many resources on the Internet (ready scripts and educational materials). It is also possible to import models from the Caffe library.

C. Image preprocessing

In the first stage, the images from the FERET and Adience databases were rescaled to 32×32 pixels. Unfortunately, on the used hardware, experiments for higher resolution lasted too long (several hours) and for larger models were even impossible. Then, the original RGB colour space was transformed to YUV, which allowed to separate luminance and chrominance information and slightly improve the performance.

In the next step the input data was normalized using the well known scheme with mean and standard deviation. First, this two values were computed for each colour component separately. Then, from all pixel in the input image the mean value was subtracted and the result was divided by the standard deviation.

D. Network architecture

As stated earlier, the main assumption of the presented work was to design a simple and computationally efficient network architecture. Therefore, it was decided to use 32×32 pixels input images. It should be emphasized that usually larger images are considered (cf. Section II-A).

The used network architecture was essentially based on the one described in [5]. However, many simplifications were introduced. The scheme is presented in Figure 2. It consists of the following components: two convolutional layers (Convolution 64), three ReLU transfer layers (in-place operation – $f(x) = max(0, x)$), two subsampling layers with a *max* operator (Pooling), data reshape, two fully connected layers (FCL), in-place dropout with 0.5 probability and LogSoftMax operation ($f_i(x) = log(\frac{1}{\sum_j e^{x_j}} \cdot e^{x_i})$).

E. Network training

In the experiments the following labels were used. For gender: 1 – man, 2 – woman. For age: 0 – 9, as the range $[0, 100]$ was divided into 10 intervals (we used the value 10 due to compatibility with our previous research). Some solutions allow to determine the age with one year accuracy – for example [7]. However, this increases the complexity of the network structure and for a lot of applications this precision is not required.

After creation of the network model, the weights were initialized (in this work random values were used). Then, during an iterative learning process (using stochastic gradient descent (SDG) method) the weights were modified to reduce the classification error.

IV. Evaluation

For the FERET database 9 experiments were carried out. Their main aim was to examine how the size of used convolutional filters and sizes of the training and test sets, as well as the number of iterations (epochs) affected the classification performance. The obtained results are summarized in Table I (upper part). As the training times for both networks (age, gender) were similar, only the mean value is presented.

Like expected, increasing the number of training samples had a positive effect on the recognition efficiency – using 2,000 images allowed to obtain a 98.6% gender recognition accuracy. It is worth noting, that this is 4% better than the "classic" solution descried in Section II-B and work [8].

Age estimation is a more complex issue. In addition, it was not easy to provide enough training and test samples for each of the 10 classes. Again, the best results were obtained for

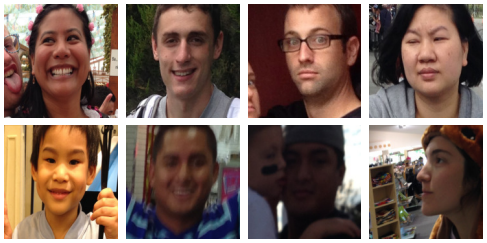


Figure 3: Sample positive (top row) and negative (bottom row) gender classification results from the Adience dataset

the largest network – 86.4%. It is over 15% better than the "classical" solution and slightly better than reported in [3].

Using different filter sizes (3×3, 5×5 and 7×7) had only a minor impact on the final classification performance (not more than 1% difference). A small 3×3 filter allowed to achieve best age estimation – 86.4%.

For the Adience dataset 4 test were conducted. Their results are summarized in Table I (bottom part). Images of size 32×32 were used. In the first test 50% for gender and 42% for age accuracy was obtained. This is much worse than on the FERET base, but the reference DCNN from [5] on this demanding set allowed for approx. 87% for gender and 50.7% for age accuracy. Because the size of the input image could not be increased, two optimizations proposed in the work [5] were evaluated on the compact DCNN. In the third test, the input image was restricted only to the centre part of the original image. In the fourth test, 5 images were passed to network – one centre and 4 from corners. This allowed to compensate the incorrect alignment of faces in the input images. Especially the second modification had a positive impact on the final solution – 62% for gender and 42% for age accuracy. Some sample gender classification results are presented in Figure 3. It should be noted that the unsuccessful cases (bottom row) are quite difficult.

V. Conclusion

In this paper the use of a compact DCNN architecture for age and gender estimation was proposed. The input image size was defined as 32×32 pixels. This allowed to obtain 98.6% accuracy for gender and 85.34% for age in 10 years intervals on the FERET database. When compared to our previous solution based on LBP and SVM, respectively a 4% and 15% improvement was obtained.

For a much more demanding Adience database 62% gender and 42% age accuracy was measured. These are results respectively 25% and 8% worse than for the large DCNN with input image size 227×227. It is worth to emphasize that the measured classification time was 16 ms (vs 200 ms) and this despite of the used computing platform (mid-range notebook).

The obtained results are interesting due to at least three reasons. Firstly, the access to powerful GPU platforms or even clusters in not common. Especially nowadays, where most personal computers are notebooks which are being slowly replaced by advanced smartphones. Secondly, the "didactic" aspect should not be missed. It seems that deep learning and convolutional neural networks should be introduced for graduate students of electrical engineering and computer science faculties. For an effective teaching process, students should be able to carry out simple experiments on their own hardware. In addition, the network training should not take too long (e.g. not several hours) and the classification result should be "decent" (reasonable when compared to large DCNNs). Thirdly, the use of a compact DCNN in an embedded vision system allows to reduce costs (less computing resources required), minimize the energy consumption and perform real-time video stream processing.

In the near future our research will concentrate on analysing the use of reprogrammable FPGA devices, heterogeneous Zynq SoC (which both are proven platforms for embedded real-time vision systems), as well as the recently proposed solutions like NVIDIA Jetson FX-1 for implementing compact, energy efficient but also accurate DCNN based classifiers and vision systems.

Acknowledgment

The work was supported by AGH University of Science and Technology project number 15.11.120.879.

References

[1] M. Szkudlarek and M. Pietruszka, "Fast gpu and cpu computing for head position estimation," in *Proceedings of the 2015 Federated Conference on Computer Science and Information Systems*, ser. Annals of Computer Science and Information Systems, M. Ganzha, L. Maciaszek, and M. Paprzycki, Eds., vol. 5. IEEE, 2015. doi: 10.15439/2015F410 pp. 231–240.

[2] A. G. Ivakhnenko, "Polynomial theory of complex systems," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-1, no. 4, pp. 364–378, Oct 1971. doi: 10.1109/TSMC.1971.4308320

[3] F. H. C. Tivive and A. Bouzerdoum, "A gender recognition system using shunting inhibitory convolutional neural networks," in *The 2006 IEEE International Joint Conference on Neural Network Proceedings*, 2006. doi: 10.1109/IJCNN.2006.247311. ISSN 2161-4393 pp. 5336–5341.

[4] I. Huerta, C. Fernández, C. Segura, J. Hernando, and A. Prati, "A deep analysis on age estimation," *Pattern Recognition Letters*, vol. 68, Part 2, pp. 239 – 249, 2015. doi: http://dx.doi.org/10.1016/j.patrec.2015.06.006 Special Issue on "Soft Biometrics".

[5] G. Levi and T. Hassncer, "Age and gender classification using convolutional neural networks," in *2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, June 2015. doi: 10.1109/CVPRW.2015.7301352. ISSN 2160-7508 pp. 34–42.

[6] S. Escalera, J. Fabian, P. Pardo, X. Baro, J. Gonzalez, H. J. Escalante, D. Misevic, U. Steiner, and I. Guyon, "Chalearn looking at people 2015: Apparent age and cultural event recognition datasets and results," in *2015 IEEE International Conference on Computer Vision Workshop (ICCVW)*, Dec 2015. doi: 10.1109/ICCVW.2015.40 pp. 243–251.

[7] R. Rothe, R. Timofte, and L. V. Gool, "Dex: Deep expectation of apparent age from a single image," in *2015 IEEE International Conference on Computer Vision Workshop (ICCVW)*, Dec 2015. doi: 10.1109/ICCVW.2015.41 pp. 252–257.

[8] B. Hebda and T. Kryjak, "Age, race and gender estimation based on facial images," in *Zeszyty Studenckiego Towarzystwa Naukowego*, 2015, pp. 137—141.

[9] P. Phillips, H. Wechsler, J. Huang, and P. J. Rauss, "The {FERET} database and evaluation procedure for face-recognition algorithms," *Image and Vision Computing*, vol. 16, no. 5, pp. 295 – 306, 1998. doi: http://dx.doi.org/10.1016/S0262-8856(97)00070-X

[10] E. Eidinger, R. Enbar, and T. Hassner, "Age and gender estimation of unfiltered faces," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 12, pp. 2170–2179, Dec 2014. doi: 10.1109/TIFS.2014.2359646

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 791–796
DOI: 10.15439/2016F494
ACSIS, Vol. 8. ISSN 2300-5963

Quality Metric for Shadow Rendering

Krzysztof Kluczek
Gdańsk University of Technology
Department of Intelligent Interactive Systems,
Narutowicza 11/12, 80-233 Gdańsk
Email: krzysiek@devkk.net

***Abstract*—Shadow rendering is one of the most important aspects of rendering 3D environments, yet, the problem is far from trivial. A number of shadow rendering algorithms exist, with various degrees of rendering quality, fidelity and performance. Additionally, many of such algorithms offer high degrees of flexibility when it comes to fine tuning. This paper proposes a new method of measurement of quality of shadows produced by rendering algorithms, which method can be used for automation of algorithm choice and fine-tuning of such algorithms to specific data sets and use cases.**

I. Introduction

ONE of the most important aspects of rendering virtual objects and environments is rendering of shadows [1]. While a great number of shadow rendering algorithms was developed, all of them require some degree of compromise between image quality, fidelity and rendering performance. With ray-tracing allowing very high shadow fidelity at the cost of rendering performance, and shadow map techniques offering [2] real-time performance even on low-end devices at the cost of image quality, the choice of shadow rendering algorithms is often one of key decision taken during development of rendering applications. Many of such algorithms offer further parametrization, which allows fine-tuning to given purpose. To simplify the task of such choices and enable some degree of automation in this matter an automated shadow rendering quality measurement algorithm is required. This paper presents a proposal of such an algorithm.

II. Related Work

While a lot of research has already been done in the area of image quality assessment, it is surprising that we couldn't find any proposal of a complete metric fit to measure and compare quality of rendered shadows. The proposed NoRM No-Reference Image Quality Metric [3] was capable of detecting certain type of shadow rendering artifacts (shadow map aliasing), but because the method was heavily based on machine learning, it was impossible to be easily reproduced in our environment. Another study [4] evaluates existing full-reference image quality metrics in the context of detection and measurement of two certain types of shadow artifacts (acne and peter panning) showing that SSIM [5] and MSSIM [6] metrics outperform other evaluated methods. Still, in our research of existing works we were unable to find a quality metric that would focus on all aspects of evaluating quality of shadow rendering algorithms.

III. Proposed Model

The proposed quality metric is capable of quality assessment of shadow rendered for a single light source (for any type of such light source) as seen from a single view point. The model is based on weighted average of several submetrics, each focusing on a key aspect of rendered shadow quality:

- shadow fidelity
- aesthetics
- detail
- rendering performance

The weighted average of the above aspects allows tuning the metric itself to given purpose and expectations. For example, while rendering performance will be much more important in the real-time applications than absolute shadow fidelity, the exact opposite will be true for offline rendering in cinematography.

Because the proposed metric was developed mostly for use in comparing shadow rendering algorithms, the metric computation itself does not need to be real-time. Therefore, we allowed ourselves to use full-reference, white-box methods for simplicity. The reference images were generated using ray-tracing algorithm with accurate soft-shadow simulation and high sample count (1024 samples per pixel). White-box approach allows easy separation of the shadow information from the rest of the scene by use of shadow masks. The shadow mask is an additional output from a renderer, which represents only the geometry term (the visibility factor) for a given light. The example of shadow mask image generated for our test scene can be seen on Figure 1.

IV. Shadow Fidelity Measurement

The purpose of the shadow fidelity metric component is measurement of how closely the shadow under evaluation matches the expected physical shadow. This component includes assessment of correct shadow placement and topology, as well as correct representation of fully shadowed and fully unshadowed area, but does not include noise errors, which will be included in the shadow aesthetics component of the metric. Such separation was implemented in order to enable fine-tuning the final metric by specifying the weights for each submetric separately to adjust the importance factors to specific needs.

The shadow fidelity submetric is, just like the main metric, a weighted average of several components:

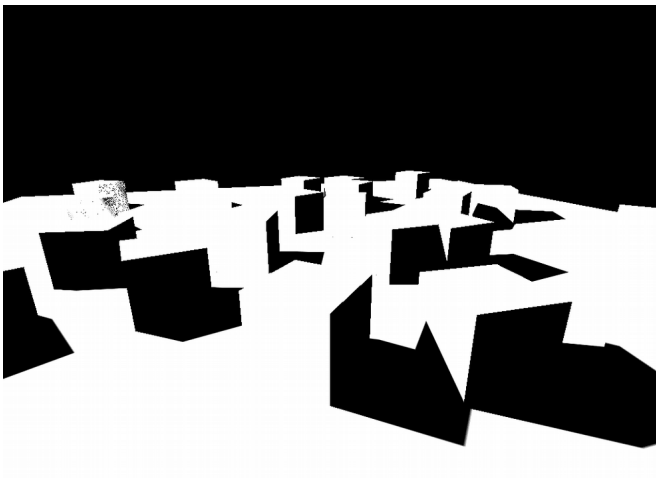


Fig 1.An example of shadow mask image for our test scene

- 10% shadow contour match
- 50% shadow contour match
- 90% shadow contour match
- dark area coverage
- lit area coverage

The first three components assure that the soft shadow boundary in test image is located where it ought to be, while the last two components assure that there are no extra features present in regions, which are empty in the reference image.

A. 10%, 50% and 90% Shadow Contour Matching

The shadow contour line is defined as isoline, for which the shadow factor (the geometry factor from the lighting equation) is equal to specified percentage value. Three contour lines are used in our method: the 10% line, 50% line and 90% line. These factors were chosen experimentally to provide 10% margin from fully lit and fully shadowed areas. Such margin is needed, as the submetric should minimize impact of noise errors, which are to be detected with different submetric. Additional line, the 50% contour line, was introduced to test the fidelity of shadow cross-section – the distribution of shadow mask values across the penumbra region.

The method for extracting shadow contour lines is identical for all contour line values. First, the shadow mask image is blurred to reduce impact of noise on the shadow line. In our tests we used 11x11 blur kernel for images with 1024x768 pixel resolution. Then, the isoline is extracted by comparing each pixel of the blurred mask with its right, lower and lower-right neighbors. The pixel is marked if and only if a region consisting of the pixel and the neighbors contains shadow mask value transition across the contour value. As a corner case, if the region has all pixel values equal to the contour line threshold, the pixel is still marked in order to guarantee continuity of the contour line. The above operation (blurring and contour line extraction) is carried on both the test and the reference images. In the final step, for each contour pixel in reference image a contour pixel in test image is found, and quadratic mean of distances to such pixels is computed. The final score of the given contour line match submetric is computed as:

$$CM = 1 - \frac{D_{RMS}}{D_{MAX}} \tag{1}$$

where:

CM – is the contour line submetric score,

D_{RMS} – is the quadratic mean of contour distances,

D_{MAX} – is the worst-case maximum contour distance

Ideally, the mean would equal zero, which means perfect match, yielding a perfect score of 1. A worst-case scenario, yielding a score of 0, is assumed to be a scenario in which the quadratic mean of contour distance is equal to D_{MAX} – an arbitrarily chosen maximum contour distance. In our implementation, that distance was set as a 20% of image size (a geometric mean of image width and height in pixels). The choice of that parameter can be arbitrary, because the metric is intended to be a comparative one, and any fine-tuning is meant to be done by choosing the weights of the submetrics.

B. Dark and Light Area Coverage

The dark and light area coverage submetrics are computed in a manner similar to each other. First, the shadow mask image is blurred to reduce the impact of noise and shadow acne on the submetric score. In our implementation, we use the same blurred image as in shadow contour matching submetrics. Then, depending on whether we compute the dark or the light area coverage submetric, we mark all pixels which contain blurred mask values below 10%, or above 90%, for dark and light areas correspondingly. This process is carried on both the test image as well as on reference image. In the final step, both images are compared. Only the region of pixels marked on the reference image is considered. Within that region all unmarked pixels on test image are counted, giving final score of the submetric as:

$$AM = 1 - \frac{N_{unmarked}}{N_{region}} \tag{2}$$

where:

AM – is the area coverage submetric score,

$N_{unmarked}$ – is the count of unmarked pixels in the test image (within considered region),

N_{region} – is the count of all pixels of the considered region

In the event of a perfect match, all pixels marked on the reference image will also be marked on the test image, resulting in ideal score of 1. In worst case, none of the considered pixels will be marked on the test image, giving the score of 0.

V. Shadow Aesthetics Measurement

With the aesthetic submetric score we aim to detect two types of shadow rendering artifacts: the shadow acne and shadow noise in the shadowed region. While the aesthetic quality of shadow can be hard to measure, we decided to focus on shadow noise in this metric, as image noise is one of the most frequent and typical image rendering errors. The shadow aesthetics sumbetric is, again just like the main met-

ric, a weighted average of several (in this case: two) components:

- shadow ACNE value
- shadow noise value

With both components scaled to yield a value of 1 in case of perfect shadow image and a worst-case value of 0.

A. Shadow Acne Detection

The shadow acne rendering artifacts are typical to shadow map rendering algorithms and are characterized by appearance of high frequency noise within the fully lit regions of the shadow mask. Therefore, to detect such errors we need to measure the noise level within the lit portion of the shadow mask. To compute that level, we first process the shadow mask image using the Roberts cross operator to isolate high-frequency image noise [7]. Then we compute the quadratic mean of computed noise values within the lit region of the shadow mask. The lit region is considered to be the light region computed during measurement of the light area coverage in one of the earlier submetrics. Both the test shadow mask and the reference shadow mask images are processed in the above manner and the final acne detection score is computed as:

$$AV = min\left(\frac{1-T_{RMS}}{1-R_{RMS}}, 1\right) \quad (3)$$

where:

AV – is the score of acne detection,

T_{RMS} – is the quadratic mean of noise detected within the lit region on the test image,

T_{RMS} – is the quadratic mean of noise detected within the lit region on the reference image

The above equation normalizes the acne score of the test image to the acne score of the reference image because some features of the reference image (such as penumbra edges) can be falsely detected as acne, therefore making ideal score of 1 impossible to reach. The final value is limited to 1 to prevent the score from going above that value if test image lacks features, present on the reference image, that would otherwise be falsely detected as acne.

B. Shadow Noise Detection

The shadow noise is detected in a manner identical to the acne detection method outlined above, with the exception that now only the dark region is considered. The dark region is the region detected as a dark portion of the shadow mask during dark region coverage measurement. The rest of the noise detection algorithm remains the same as in acne detection algorithm.

VI. Shadow Detail Measurement

The insufficient shadow detail comes in most cases from the low shadow map resolution, as compared to its projection on the screen. To measure the quality of shadow detail as a shadow resolution compared to image resolution, we have to utilize white-box testing method and produce yet another image output from our shadow rendering algorithms: the shadow map texels projection on the scene. Such texel projection image is prepared first by assigning each shadow map texel an unique color index, then projecting such shadow map directly on the scene. To limit further computations to region of interest where shadow map detail can possibly affect rendering, in such projection we ignore any geometry that is not facing the light for which shadow quality is measured, as well as any regions that don't contain any geometry (e.g. sky). The example of texel projection image generated for low resolution shadow map (for the purpose of readability) can be seen on Figure 2.

To compute the shadow detail submetric score the following method is used. For each pixel of interest within the shadow map texel projection image we count the number of pixels, that share the same texel index. Then this count is averaged using quadratic mean. Then the final score of shadow detail submetric is computed as:

$$SD = \frac{1}{S_{ratio}} \quad (4)$$

where:

SD – is shadow detail submetric score,

S_{ratio} – is share ratio computed using the quadratic mean

Because we consider only pixels that have any shadow texels projected on them, the share ratio can never drop below 1. Therefore, the perfect score, when each pixel has its own shadow map texel assigned, is 1. Note that the shadow map resolution exceeding the pixel resolution across any portion of the scene does not further increase the score.

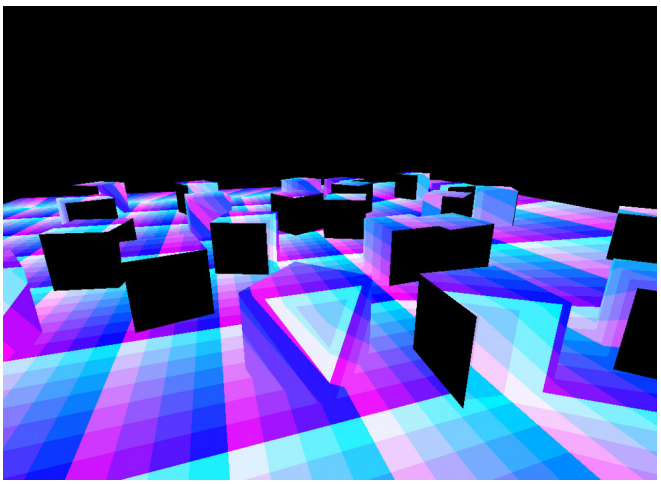


Fig 2.An example of a shadow texel projection image

In the event the given shadow rendering technique does not rely on intermediate projected texture maps and has infinite detail resolution (like raytracing or shadow volume techniques), the shadow detail metric is skipped entirely assuming perfect score of 1.

VII. Rendering Performance Score

The rendering performance submetric score is computed basing on shadow rendering time and is computed as:

$$RP = \frac{1}{1 + \frac{T_{shadowed} - T_{unshadowed}}{T_{ref}}} \quad (5)$$

where:

RP – is the rendering performance score,

$T_{shadowed}$ – is the rendering time of the scene with the shadow rendering algorithm enabled,

$T_{unshadowed}$ – is the rendering time of the scene with the shadow rendering algorithm disabled,

T_{ref} – is the reference time value, scaling the score distribution

We suggest using value of T_{ref} equal to the inverse of the minimum final target frame rate (maximum time of rendering allowed for a complete single frame). With rendering performance score specified this way, the score assumes value of 1 in ideal case of shadow rendering not taking any time. When shadow rendering time rises to maximum time allowed for rendering the whole frame (making the shadow algorithm impossible to use), the performance score drops to value of 0.5. Further increase in shadow rendering time causes the score to drop further towards 0 as the rendering time rises towards infinity.

During our tests we did not use performance measurement submetric, as our implementations of shadowing algorithms were not fully optimized. Therefore, any performance comparisons would not produce any meaningful results.

VIII. Example Results

To test our quality metric on real data sets we have implemented four different shadowing techniques, in addition to raytracing, which was used as reference. All of the implemented test techniques were based on the shadow mapping technique for point light shadows. The most widely used for such cases Cube Shadow Map technique (CSM) was implemented, as well as slightly modified Tetrahedral Shadow Map technique introduced by [8]. Additionally, both Cube Shadow Map and Tetrahedral Shadow Map techniques were implemented with Variance Shadow Map variants (CVSM and TVSM, respectively). The four techniques were tested with various shadow map texture resolutions using a simple synthetic test scene with a single point light. The test cases are named after the technique being tested and shadow map resolution being used (e.g. CSM 1024). For cube shadow maps, the resolution is specified as edge length of the shadow map cube, in texels. For tetrahedral shadow maps the resolution is specified as edge length of a cube map, which would have a number of map texels closest to the number of texels in the tetrahedral map, thus making the resolution specification comparable to the cube shadow maps.

The reference shadow mask image for our data set can be seen on Figure 3.The shadow mask images used for each test variant are presented on Figure 4. The quality measurement results are presented in Table I, separately for each submetric component. As expected, the lower resolution shadow map variants achieve lower scores, with score changes most noticeable for really low resolution shadow maps. The tetrahedral shadow mapping techniques score on average worse than cube shadow map techniques, as the map shape distortions are more prominent in the tetrahedral maps, resulting in uneven distribution of shadow map texels over the lit scene. The Variance Shadow Mapping variants perform comparably to the non-VSM variants, with noticeable improvement within the lit region at the cost of degradation of shadow region coverage, resulting from softening the shadow edge across the region transitions. It can be seen, that this also slightly influences acne and noise measurements, as this influence was limited, but not completely removed from the metric.

IX. Conclusion

In this article we presented a quality metric proposal for quality measurements of shadow rendering algorithms. We presented several components of such metric, which were designed to measure quality of various characteristics of the shadow mask images. While there is still room for improvements when it comes to the methods and algorithms used, the quality measurement metric presented can already be useful in comparison of shadow rendering algorithms, detecting common rendering artifacts and yielding stable and comparable results.

References

[1] P. Boulenguez, B. Airieau, M.-C. Larabi, D. Meneveaux, “Towards a perceptual quality metric for computer-generated images,” in *Proc. SPIE 8293, Image Quality and System Performance IX*, Burlingame, CA 2012, doi:10.1117/12.908067.

[2] L. Williams, "Casting curved shadows on curved surfaces", *Proc. of the 5th annual conference on Computer graphics and interactive techniques* - SIGGRAPH '78, 1978.

[3] R. Herzog, M. Čadík, T. O. Aydčin, K. I. Kim. K. Myszkowski, H. P. Seidel, H. (2012). NoRM: No-Reference Image Quality Metric for Realistic Image Synthesis. *Computer Graphics Forum*, 31(2pt3), pp. 545-554.

[4] R. Piórkowski, "Automatic Detection of Shadow Acne and Peter Panning Artefacts in Computer Games." in *Central European Seminar on Computer Graphics for students 2015,* Smolenice, Slovakia, pp 117-123.

[5] Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," in *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600-612, April 2004.

[6] Z. Wang, E. P. Simoncelli, A. C. Bovik, "Multiscale structural similarity for image quality assessment," *Signals, Systems and Computers, 2004. Conference Record of the Thirty-Seventh Asilomar Conference on*, 2003, pp. 1398-1402 Vol.2. doi: 10.1109/ACSSC.2003.1292216

[7] L. S. Davis, "A survey of edge detection techniques", *Computer Graphics and Image Processing*, vol 4, no. 3, pp 248-260, 1975

[8] H.-C. Liao, “Shadow mapping for omni-directional light using tetrahedron mapping”, in *GPU Pro: Advanced Rendering Techniques*, Boca Raton, CRC Press, 2010, pp. 455–475.

TABLE I.
QUALITY SUBMETRICS VALUES FOR DIFFERENT SHADOW RENDERING TECHNIQUES FOR THE TEST SCENE

Shadow rendering method	10% contour score	50% contour score	90% contour score	Dark area coverage	Lit area coverage	ACNE score	Shadow noise	Detail
Ray-tracing	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
CSM 128	0.909	0.909	0.907	0.920	0.764	0.995	0.994	0.000
CSM 256	0.935	0.931	0.925	0.952	0.776	0.996	0.992	0.002
CSM 512	0.945	0.940	0.937	0.973	0.803	0.995	0.991	0.006
CSM 1024	0.952	0.948	0.944	0.985	0.858	0.990	0.994	0.021
CVSM 128	0.848	0.890	0.901	0.851	0.790	0.986	0.992	0.000
CVSM 256	0.911	0.913	0.925	0.913	0.802	0.985	0.993	0.002
CVSM 512	0.942	0.943	0.940	0.953	0.830	0.988	0.994	0.006
CVSM 1024	0.948	0.948	0.945	0.975	0.865	0.994	0.995	0.021
TSM 128	0.871	0.866	0.864	0.953	0.596	0.989	0.987	0.000
TSM 256	0.882	0.879	0.862	0.969	0.630	0.991	0.988	0.000
TSM 512	0.925	0.908	0.894	0.975	0.730	0.995	0.986	0.001
TSM 1024	0.950	0.937	0.932	0.985	0.829	0.993	0.988	0.005
TVSM 128	0.858	0.894	0.901	0.842	0.713	0.985	0.993	0.000
TVSM 256	0.902	0.921	0.915	0.887	0.761	0.982	0.992	0.000
TVSM 512	0.926	0.936	0.935	0.934	0.837	0.987	0.993	0.001
TVSM 1024	0.949	0.948	0.945	0.965	0.862	0.991	0.994	0.005

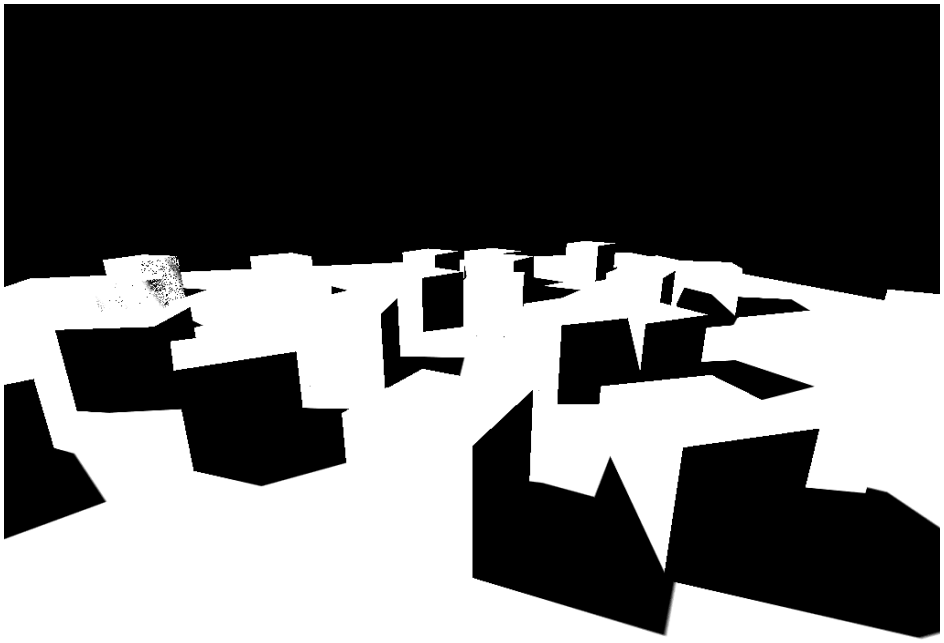


Fig 3. The reference shadow mask image generated using a commercial ray-tracer

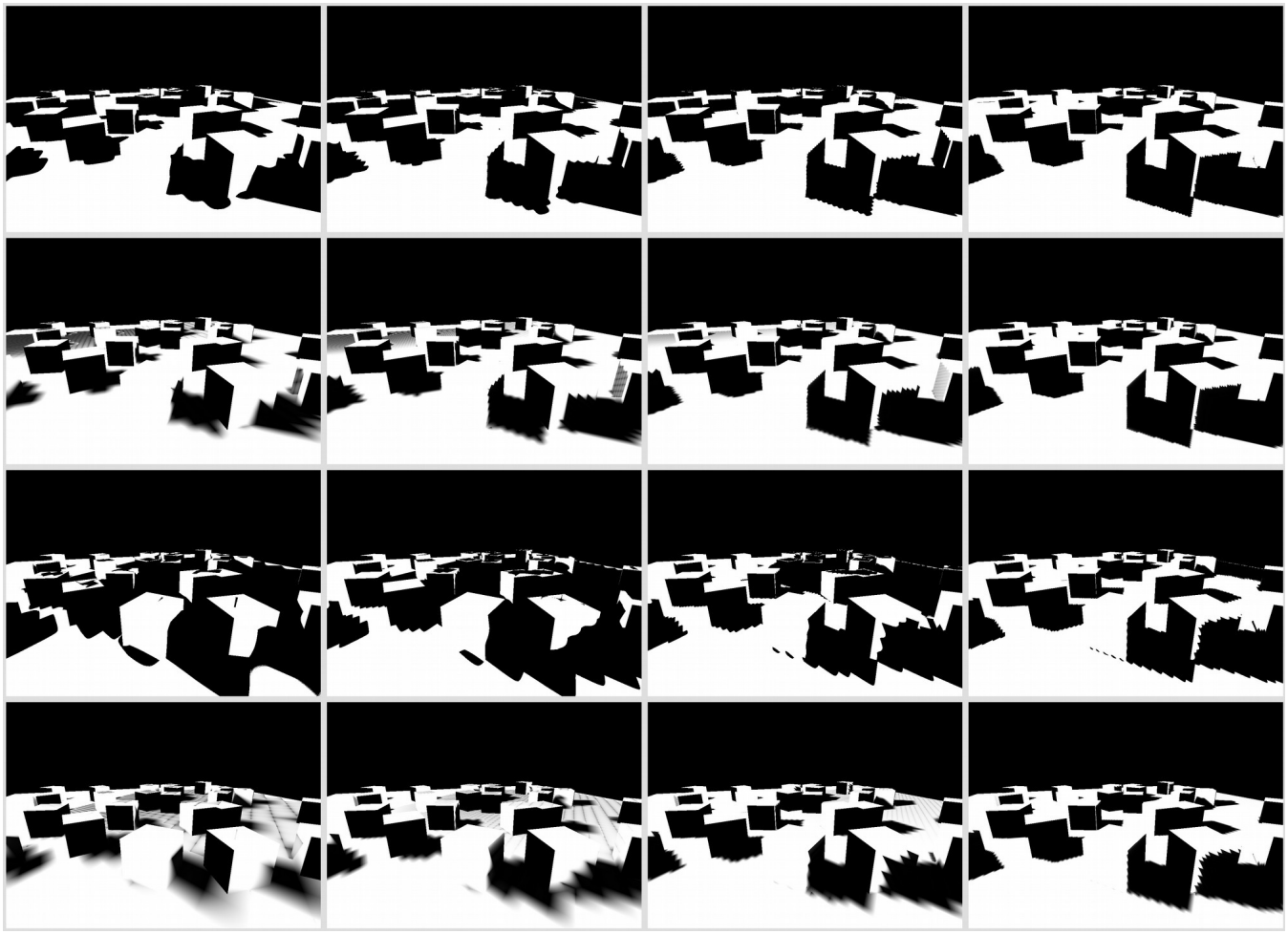


Fig 4. Shadow mask images used in our tests; techniques in rows: CSM, CVSM, TSM and TVSM; resolutions in columns: 128, 256, 512, 1024

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 797–801
DOI: 10.15439/2016F518
ACSIS, Vol. 8. ISSN 2300-5963

Using formant frequencies to word detection in recorded speech

Łukasz Laszko
Cybernetics Faculty,
Military University of Technology,
ul. Gen. S. Kaliskiego 2,
00-908 Warsaw, Poland
email: lukasz.laszko@wat.edu.pl

Abstract—**The paper considers increasing the precision of detection of words in unsupervised keyword spotting method. The method is based on examining signal similarity of two analyzed media description: registered voice and a word (textual query) synthesized by using Text-to-Speech tools. The descriptions of media were given by a sequence of Mel-Frequency Cepstral Coefficients or Human-Factor Cepstral Coefficients. Dynamic Time Warping algorithm has been applied to provide time alignment of the given media descriptions. The detection involved classification method based on cost function, calculated upon signal similarity and alignment path. Potential false matches were eliminated in the algorithm by applying two-staged verification, using the Longest Common Subsequence algorithm and analyzing formant frequencies of eleven English monophthons. The use of formant frequencies at the stage of verification increased overall detection precision by about 10% as compared to original algorithm.**

***Index Terms*—keyword spotting, formant frequency analysis, pattern matching, audio information retrieval**

I. Introduction

INCREASING use of digital sound processing methods to simple daily tasks is currently very popular due to widespread availability of mobile devices having implemented this type of methods. Regarding this trend [1] considers an approach that could be used to detect words in recorded speech of unknown language without training, by using publicly available, free of charge online translation services with Text-To-Speech support e.g.: Google Translate, Bing Translator, Yandex Translate[1].

The problem of word detection consists in searching for given words in a speech medium, which is either solid container or a stream. The detection is usually given by the two coupled values [2]: time code of the beginning of the word and a quality ratio. This problem in contemporary literature is usually called "keyword spotting[2]" (shorten as KWS) [3], [4].

This work was supported by Cybernetics Faculty (Military University of Technology, Warsaw, Poland) under the grant no. RMN/765/2015

[1] Names of the products have been presented in this paper only in relation to the contemporary, publicly available technology, not for marketing purposes.

[2] Also „spoken term detection" (shorten STD).

Classical solutions to this problem address supervised approach where models such as hidden Markov model (HMM) or support-vector machine (SVM) are trained like in a typical automatic speech recognition (ASR) system, using Large Vocabulary Continuous Speech Recognition (LVCSR) methods [5]. In consequence the speech signal is divided into segments of equal-size, from which speech features are extracted. Next, an appropriate algorithm is employed to determine the type of signal in each segment. As a result, recognized words, together with the corresponding indexes are stored in a database. Then, a text query is performed within the indexed data [6].

Based on the fact that for some applications it is not possible to have model trained, either due to lack of relevant training data or due to time-specific limitations, different unsupervised approaches to the problem have been developed [2], [7].

Under the concept of unsupervised matching lay suitable speech features and a classification strategy. The approach presented in [1] employs cepstrum-based features: Mel-Frequency Cepstral Coefficients (MFCC) and Human-Factor Cepstral Coefficients (HFCC). As for classification strategy that approach points Dynamic Time Warping (DTW) algorithm.

However results of applying the method shows relatively high overall rate of false positives: 13,51% for MFCC and 14,86% for HFCC.

In this paper, the author propose an approach to minimize this insufficiency, by adding additional verification stage, based on the analysis of formant frequencies. The motivation for this came from [12]. Using formant frequencies analysis, as shown below, has a positive influence on the results, but limits the versatility of the KWS method to specific language only. Different improvement techniques used for KWS could involve combining of multiple features, like described by Mitra et al. in [5].

II. Problem statement

A. Method background

This paper considers the same use case as in [1]. In this approach the KWS method supports human operator in searching for specific words in a given speech medium. For

this scenario, the sound queries are synthesized directly from text. The method gives coarse detection, prior to involving precise detecting methods or just hearing by the operator.

B. Speech features model

The approach assumes at this point the choice of appropriate speech signal features. In the research two types of feature vectors have been used: Mel-Frequency Cepstral Coefficients (MFCC) and Human-Factor Cepstral Coefficients (HFCC). MFCCs have been computed according to the following algorithm:

1) given signal S has been windowed by Hamming window resulting in N segments, $s_1 \ldots s_N$;

2) each segment has been processed by short-time Fourier transform (STFT) with length of 51 ms and step size of 10 ms;

3) then the triangular filter bank has been developed with 40 equally spaced mel-scale center frequencies f_i, $i = 1,\ldots,40$ and with uniform bands controlled by the neighbor center frequencies $f_{i\pm1}$;

4) next, the actual filtering has been done, by multiplication of each STFT segment (representing magnitude spectrum) with magnitude spectrum of bands for MFCC;

5) finally, the result has been decorrelated using Discrete Cosinus Transform (DCT), keeping only 15 the most decorrelated vectors (MFCC coefficients).

The same model has been applied to Human-Factor Cepstral Coefficients, with a change in point 3). In HFCC, center frequencies are still equally spaced in mel frequency scale, but unlike MFCC filter bandwidth is treated as a parameter, which determines filter bands' cut-off frequencies, using measure called Equivalent Rectangular Bandwidth (ERB) [8]:

$$ERB(f) = 6.23f^2 + 93.39f + 28.52 \text{ Hz} \quad (1)$$

where f states for filter center frequency, expressed in kHz.

C. Textual query

In the presented method Text-to-Speech (TTS) system is exploited to generate synthetic voice from a text query. Next, the query (pattern) is transformed to chosen speech feature space. Then a chunk of speech signal from a given source is read. This chunk is transform to the same speech feature space. Then a classification strategy is applied. In case of the pattern matched, time code of the corresponding sound segment is registered.

Using textual query and TTS make it easy to extend the approach to reflect language variations assumed in the scenario, to search for the same word translated to several languages [1].

D. Similarity and time alignment

DTW is used in the method to compare two feature vectors of different length (analyzed voice and the reference pattern) and to find an optimal alignment path P of both. P is usually calculated upon the local distance matrix (similarity matrix) from the minimal indexes (usually lower left corner) to maximal indexes (usually upper right corner) of the matrix. Optimal means here the lowest cost path P for passing from one point of matrix to another, within given constraints. For details of applying DTW to exemplary speech features vectors, see [1].

Building similarity matrix $D_{A,R}$ where A stands for analyzed voice feature vector and R stands for reference pattern feature vector, is the first step considered in speech classification. Feature vector consists of either MFCC or HFCC coefficients computed for segments $s_1 \ldots s_N$. Individual element $d(a,r)$ of similarity matrix, where a,r stands for specific element of vector A and vector R respectively, is given by inner product:

$$d(a,r) = \frac{\langle A_a, R_r \rangle}{\|A_a\| \|R_r\|} \quad (2)$$

Next, the two-staged cost path algorithm is executed. The first stage stands for the calculation of an accumulator $C_{A,R}$ (where C is of size D). The resulting structure contains at each of its element $c(a,r)$ the value of accumulated lowest transition cost to this element from its neighbors, including the cost of lowest transition to the neighbors from theirs consequent neighbors until the starting element $c(1,1)$. The computation retains directional constraints, according to the recursion:

$$c(a+1,r+1) = d(a+1,r+1) + \min \begin{cases} c(a-1,r) \\ c(a,r) \\ c(a,r-1) \end{cases} \quad (3)$$

where: $a,r \geq 1$ and $c(1,1) = d(1,1)$.

The second stage stands for an optimal aligning of analyzed voice and the reference pattern. In this stage the path P is created. Its creation is based on the accumulator traceback, starting from its last point $c(N_A, N_R)$ and ending in point $c(1,1)$ recursively by searching across all allowable predecessors to each point. Because each point holds the value of the lowest transition cost to itself, the actual calculation of the path is based on choosing the next point upon the minimal value.

E. Classification and verification

After applying DTW, to proper classification an additional matching procedure is proposed in the method, see Fig. 1.

This procedure assigns weight values v based on referring points of matrix $D_{A,R}$ and a path threshold T_P to the path P, satisfying inequality (4).

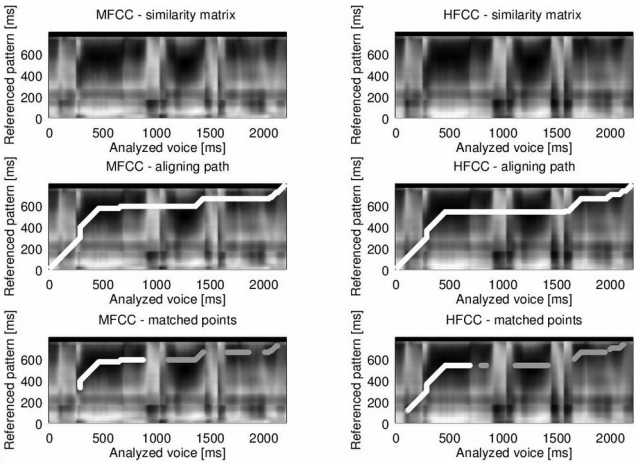


Fig. 1. Pattern matching procedure. Upper images present computed similarity between the pattern (the word "school" - synthesized woman's voice) and analyzed voice (recorded man's voice "school is closed today"). Images in the middle present global alignment path. Bottom images present resultant match: white strip is for the best match, gray strips are for remaining matches.

$$T_P \le v := 1 - d \le 1 \tag{4}$$

where T_P controls the number of points suspected to indicate detected words.

Assuming, that several possible word could occurre, which is indicated by subsequences: $p_{k_1}^{(l)} \dots p_{k_{N_p}}^{(l)}$, $l = 1,2\dots$, the verification step is executed.

Originally this step consisted of applying Longest Common Subsequence (LCS) with maximization criterion of cumulative weights of subsequences, i.e. the longest subsequence with the highest sum of weights wins.

The number of possible word reoccurrence was controlled by sequence threshold value, which restricted the minimal cumulative cost for a subsequence.

As a result of matching procedure, assuming only one occurrence of the searched word, the best match is projected to the analyzed voice time domain (e.g. white strips in the bottom images of Fig. 1).

F. Format frequencies analysis

Conclusions from experiments described in [1] indicate that relatively high level of false positive results could be minimized after applying less speaker-dependent speech features. By following that suggestion and motivated by [12], in this paper the author propose to extend verification step with formant frequencies analysis. Formant frequencies are defined either as an acoustic resonance of the human vocal tract or more technically as local maxima of the envelope of the signal spectrum. As stated in [12] they are important in determining phonetic content of speech sounds, but they are not quite good speech features. This is because of, on one hand: their little speaker dependency (assuming the same speakers gender) and existence of cataloged form for specific language (usually including frequency range for a specific phonetic unit). On the other hand: their strict connection with high signal energy of phonetic units, like vowels, and unreliable measure of purely defined signals (silence, weak fricatives, etc).

Nevertheless in this paper it was hypothesized that knowledge of at least a part of speech segment will positively influence the quality of detection.

III. KWS SUPPORTED BY FORMANT FREQUENCIES ANALYSIS

A. Formants estimation

In the described research formant frequencies have been estimated only for English vowels, as for all other phones such frequencies either do not exist or are difficult to be identified. The following phonetic convention of the vowels has been adopted (see table 1) in the presented research.

TABLE 1. PHONETIC CONVENTION OF ENGLISH VOWELS SELECTED FOR THE RESEARCH[3]

Classi-ficatio n	***IPA[4] notation***	***Own[5] notation***	***Example***	***Choice***
short vowels	/ʌ/	a	c<u>u</u>p	Yes
	/æ/	ae	c<u>a</u>t	Yes
	/e/	e	b<u>e</u>d	Yes
	/ə/	e_	<u>a</u>bout	No
	/ɪ/	y	h<u>i</u>t	Yes
	/i/	i	happ<u>y</u>	No
	/ɒ/	o	h<u>o</u>t	Yes
	/ʊ/	u	g<u>oo</u>d	Yes
long vowels	/ɑ:/	aa	<u>ar</u>m	Yes
	/ɜ:/	ee	b<u>ir</u>d	Yes
	/i:/	ii	s<u>ee</u>	Yes
	/ɔ:/	oo	c<u>a</u>ll	Yes
	/u:/	uu	f<u>oo</u>d	Yes

In English language, there are 5 vowels: <a, e, i, o, u>, but their pronunciation depends on a variety of factors, resulting in several distinguishable monophthongs[6]. Usually for phonetic analysis from 8 to 13 monophthongs are chosen [9]. According to [10] for further analysis 11 monophthongs have been chosen, the choice is marked in table 1, in the far right column[7].

TABLE 2. AVERAGE VALUES OF F_1, F_2 AND F_3 IN HZ [10].

	Male			Female		
	F_1	F_2	F_3	F_1	F_2	F_3
/i:/	280	2249	2765	303	2654	3203
/ɪ/	367	1757	2556	384	2174	2962
/e/	494	1650	2547	719	2063	2997
/æ/	690	1550	2463	1018	1799	2869
/ʌ/	644	1259	2551	914	1459	2831
/ɑ:/	646	1155	2490	910	1316	2841
/ɒ/	558	1047	2481	751	1215	2790
/ɔ:/	415	828	2619	389	888	2790
/ʊ/	379	1173	2445	410	1340	2697
/u:/	316	1191	2408	328	1437	2674
/ɜ:/	478	1436	2488	606	1695	2839

[3] Own work based on [9] as well as other materials available at official IPA website: https://www.internationalphoneticassociation.org/

[4] International Phonetic Alphabet

[5] Own notation was used because of programming and results presentation reasons.

[6] Single and the smallest phonetic unit; pure vowel sound.

[7] The choice was caused by the most recent research found in this area, which published a comprehensive list of results.

B. Algorithm

In general overview the KWS algorithm discussed in this paper is presented in Fig. 2. Description of presented processing blocks can be found in [1], but the verification stage is described in details below. The input to the verification stage is given from the LCS algorithm, resulting in a few best matches (e.g. in the bottom images of Fig. 1 there is one best match colored white, the other stripes, colored gray, present remaining matches).

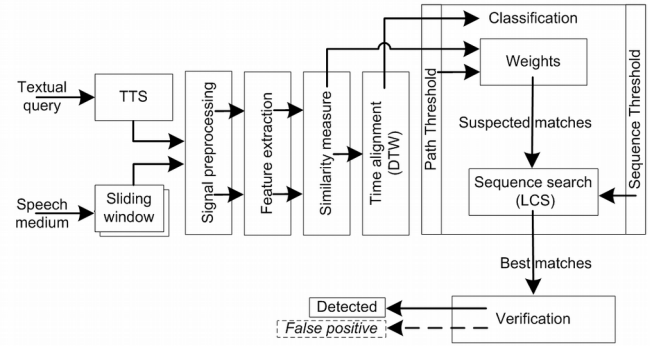


Fig. 2. Unsupervised key detection algorithm.

Let's assume there is only one best match and the verification stage is to decide if this match is a true detection or a false positive detection. Then the resultant match is subjected to verification, which is based on the analysis of formant frequencies: the reference pattern vector of formant frequencies and the analyzed voice vector of formant frequencies. It is worth noting that these two signals are after the whole procedure time-aligned, thus could be extracted and analyzed separately. Then for these excerpts the algorithm described in [11] is carried out to estimate formant frequencies.

As for the reference pattern excerpt the resultant formant sequences (exactly two formant frequencies F_1, F_2 have been chosen) are averaged and compared with the values described in table 2 to estimate appropriate monophthong.

The estimation of monophthong is performed by comparison of averaged estimated frequency F_1 from the excerpt, to all F_1 values from the table 2. The difference between the values (its absolute value), measured in the sense of Euclidean, is treated as the quality of the detection (monophthong cost). The smaller the difference the better the detection. The same estimation is done for F_2. As a result of this procedure monophthongs are detected (if any) for the reference pattern. From the collection of all detected monophthongs only the one with the smallest cost is chosen. The example of detection of /u:/ monophthong for the reference pattern is presented in the Fig. 4. on the vertical axis[8].

This estimation is carried out in an analogous manner to the analyzed voice excerpt. Finally, if the estimated monophthong for reference pattern matches analyzed voice monophthong, then according to the scenario assumed in the research, the excerpt with analyzed voice is played back to the human operator.

[8] It should be noted that the time position indicated by white lines has only illustrative value and not necessarily reflect the detection of a monophthong. This is due to successive averages made in the algorithm.

C. Experiments

A series of preliminary experiments have been conducted with regard to the presented algorithm. The target was to determine the influence of the proposed verification stage based on formant frequencies analysis to the quality of detection.

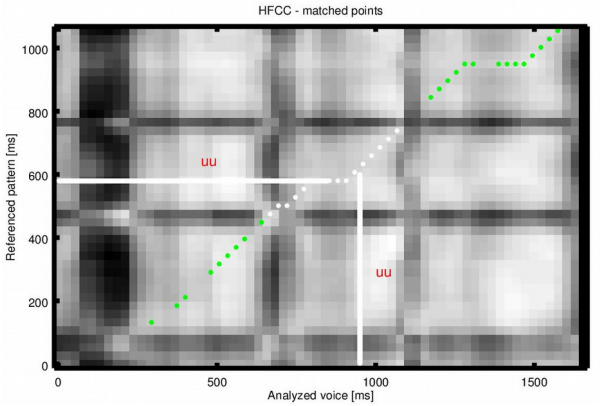


Fig. 3. An example of detecting monophthong /u:/ in the word "school". The dots represent optimal alignment path (as a result of DTW). White dots represent best matched sequence (as a result of LCS). While the white lines with the description "uu" represent the (averaged) position of the detected monophthong in the matched sequence.

The experiments have been conducted on the same research material as the original method [1], but only male voices have been chosen. Therefore the material consisted of five short (from 1 – 5 seconds) sentences in English language: spoken by one man (natural speech) and synthesized by six TTS systems with fourteen men voices. This material has been stored on a hard drive in the WAV containers.

The queries have been produced online by one chosen TTS system, different from these used in prepared research material. The TTS was accessed through the World Wide Web via HTTP protocol.

During examination all used sounds were resampled to 8000 Hz. The model (Fig. 2) was configured according to the same guidelines as in the original examinations.

Experiments have been conducted according to the following strategy: selected word to find (textual query) has been sent to the TTS system to obtain speech signal, the signal then has been read by program and compared with the entire research material according to the algorithm (Fig. 2).

D. Results

A series of preliminary results were obtained. They were compared to the results of the original method [1]. This allowed to determine the influence of the new verification approach on the quality of word detection. These results are shown in Table 3.

As it was hypothesized the percentage of false positives decreased. The decrease is about 40%. Surprisingly this had also positive effect to the percentage of detected words and had no negative effect on misses ("No detection").

The results showed that the unsupervised detection of word in a given set is possible with relatively high detection rate. Moreover, new verification method has given satisfactory results, approaching the method to industry

standards. Comparing the actual results with the original results of the method (MFCC: 82.43%, HFCC: 85.14%), it can be concluded, that MFCC recorded an increase of 8.32 percentage points, and HFCC recorded an increase of 8.66 percentage points. This gives an overall increase in precision of about 10% for both MFCC and HFCC features.

However taking into account small research material involved in the examination, these results do not allow for general statements.

TABLE 3. **Overall results by speech features**

	Detected words	**No detection**	**False positive (new method)**	**False positive (original method)**	**Overall increase in the elimination of false positives**
MFCC	90,75%	4,05%	5,19%	13,51%	38,41%
HFCC	93,8%	0,00%	6,2%	14,86%	41,72%

IV. CONCLUSIONS AND LESSON LEARNED

Results of the work shows that the inclusion in the KWS formant frequencies analysis, increases its quality. This partly proves the hypothesis that knowledge of a part of speech segment has a positive influence on the quality of detection. The main advantage of this approach is the elimination of false positives. However the presented approach limits the application of the method only to one language, due to the requirement of having a catalogue of formant frequencies. The requirement of possessing 3 formant frequency models (for male, female and children) for each language is also a strong one.

During the examination the author also encountered the problem of covering (overlapping) formant frequencies that lie in close proximity to each other (in the spectrum), also well known from the literature (see [12]). For example, for F_1 and F_2 in the spectrum only one peak is perceptible. In the described research it was noticed that generally F_2 overlaps F_1, therefore F_2 becomes F_1 and F_3 becomes F_2 in consequence.

One solution to this problem was to properly assign these frequencies as F_2 and F_3, leaving F_1 undetected (or arbitrary setting its value to 0).

Leaving the problem unresolved significantly deteriorates estimation of vowels, to the extent, that the result becomes random.

The problem in the presented approach (especially while applying DTW and LCS) that cannot be fully circumvent is the determination of the threshold values. According to literature search, the most popular technique for solving this problem in KWS, is to parallel true positive rate with false positive rate for several chosen threshold values, to create Receiver Operating Characteristic (ROC) and to find optimal threshold value by using graphical method.

REFERENCES

[1] Ł. Laszko, "Word detection in recorded speech using textual queries", Proceedings of the 2015 Federated Conference on Computer Science and Information Systems, 2015, pp. 849-853, DOI 10.15439/2015F341

[2] D. von Zeddelmann, F. Kurth, and M. Müller, "Perceptual audio features for unsupervised key-phrase detection," Proc. ICASSP2010, 2010, pp. 257-260, DOI:10.1109/ICASSP.2010.5495974.

[3] S. Tabibian, A. Akbar, B. Nasersharif, "A fast search technique for discriminative keyword spotting," Artificial Intelligence and Signal Processing (AISP), 2012 16th CSI International Symposium on, pp.140-144, 2-3 May 2012, DOI:10.1109/AISP.2012.6313733.

[4] M. Sigmund, "Search for Keywords and Vocal Elements in Audio Recordings", Elektronika ir elektrotechnika, ISSN 1392-1215, vol. 19, no. 9, pp. 71-74, 2013

[5] V. Mitra, J. van Hout, et. al., "Feature Fusion for High-Accuracy Keyword Spotting", Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on, pp. 7143-7147 2014

[6] J. Tejedor, D. T. Toledano, et al., "Access Spoken term detection ALBAYZIN 2014 evaluation: overview, systems, results, and discussion", EURASIP Journal on Audio, Speech, and Music Processing (2015) 2015:21, DOI 10.1186/s13636-015-0063-8

[7] A. S. Park and James R. Glass, (Cited in [2]) "Unsupervised pattern discovery in speech," IEEE Trans. on Audio, Speech and Language Processing, vol. 16, no. 1, pp. 186–197, 2008.

[8] M. D. Skowronski and J. G. Harris, "Exploiting independent filter bandwidth of human factor cepstral coefficients in automatic speech recognition," The Journal of the Acoustical Society of America (JASA), vol. 116, no. 3, pp. 1774–1780, 2004.

[9] G. Hunter, H. Kebede, Formant frequencies of British English vowels produced by native speakers of Farsi. Societe Francaise d'Acoustique, Acoustics 2012, Apr 2012, Nantes, France

[10] D. Deterding (1997). The Formants of Monophthong Vowels in Standard Southern British English Pronunciation, Journal of the International Phonetic Association, 27, pp. 47-55

[11] R. Snell, F. Milinazzo, Formant location from LPC analysis data, IEEE Transactions on Speech and Audio Processing. Vol. 1, Number 2, 1993, pp. 129-134.

[12] J. Holmes, W. Holmes, P. Garner, Using formant frequencies in speech recognition, Eurospeech, Vol. 97, pp. 2083-2087, http://www.idiap.ch/~pgarner/pubs/holmes1997.pdf

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 803–806
DOI: 10.15439/2016F6
ACSIS, Vol. 8. ISSN 2300-5963

An Improvement of Just Noticeable Color Difference Estimation

Kuo-Cheng Liu
Information Educating Center
Taiwan Hospitality & Tourism University, Taiwan
Email: { kcliu@mail.tht.edu.tw }

***Abstract*—In this paper, the estimation of just noticeable color difference in color images is improved by using a new spatial masking. The internal generative mechanism of human of brain theory implies that the human visual system (HVS) is sensitive to the orderly stimulus possessing structural regularity which is easily to be predicted and is insensitive to the disorderly stimulus containing structural irregularity. It is obviously that the spatial masking in color images may be overestimated in the region with orderly structures and underestimated in the region with disorderly structures. By using a simple prediction model imitating the brain works of the HVS, the structural irregularity is computed to build a new masking function that can be used to improve the estimation of just noticeable color difference for color images. The masking function is further extended to build a color visual model of estimating the visibility thresholds of color images for performance comparison. Simulation results demonstrate that the proposed method is able to obtain better performance of estimating just noticeable color difference.**

I. Introduction

THE well known properties that the human visual system (HVS) has a limited sensitivity in perceiving visual information are always utilized to represent images more efficiently. Through assessing the human visual sensitivity inherent in images, the performance of many techniques has been improved in the perceptual research community and can be found in [1]-[5].

By using different levels of just noticeable color difference (JNCD), a simplified visual model introduced in [4] for estimating the perceptual redundancy for each color pixel as the visibility threshold of color difference was proposed to modify the coding efficiency of two existing image coders. In [6], Zhu *et al*. proposed a perceptual based no-reference objective image quality metric by integrating perceptually weighted local noise into a probability summation model. Unlike existing objective metrics, the proposed no-reference metric is able to predict the relative amount of noise perceived in images with different content, without a reference. Hsieh *et al*. [7] presented a copyright identification scheme for color images that takes advantage of the complementary nature of watermarking and fingerprinting was proposed to embed the watermark into the less sensitive R and B channels of the host image in the RGB color space. To gain high performance in color image processing, the properties of the human visual perception of color stimuli must be well utilized.

In this paper, the concept indicating that the brain works actively predict the input scenes and avoid the residual uncertainty/disorder is utilized to improve spatial masking for color images. Based on a simple prediction model, the luminance dominated structural irregularity is taken into account to explore a more strict spatial masking function in color images, while only background non-uniformity and texture content are used in the prior works. To avoid underestimating or overestimating the spatial masking effect in the region with structural uncertainty, the nonlinear additivity model is adopted to build a new masking function. By using this function, the visibility thresholds of color images are estimated for performance comparison under a fair viewing test.

II. Structure-based Adjustment for Spatial Masking

In [4], the spatial masking effect considering the local color image context is exploited to calculate variable just noticeable color difference (JNCD) or variable JNCD (VJNCD) of each color pixel in the uniform CIELAB color space. That is,

$$VJNCD_p = JNCD_{Lab} \cdot s_c(a_p, b_p) \cdot s_L\big(E(L_p), \Delta L_p\big) \qquad (1)$$

where $s_c(a_p, b_p)$ is a weighting function used by the CIE94 color difference equation for adjusting the dimension of the ellipsoid along the chroma a, b axis, $s_L(E(L_p), \Delta L_p)$ is texture masking adjustment primarily induced by average

background luminance $E(L_p)$ and luminance gradient ΔL_p for pixel p of the color image, and $JNCD_{Lab}$ is the basic visibility threshold of color difference in the CIELAB space. The basic threshold has been found around 2.3[10]. By using the colors on the surface of the VJNCD sphere, the perceptual redundancy inherent in each color pixel in color images can be estimated. According to the principle recently introduced in [8], the input scene information received by human eyes is not fully processed by the HVS and some information with structural irregularities is avoided and hard to be predicted. Herein, the structure-based adjustment, $s_U(LR_p)$, caused by considering the amount of luminance residual, LR_p, is proposed and incorporated to improve the estimation of variable just noticeable color difference. In this paper, both the structure-based adjustment and the texture masking adjustment are adequately combined to design a new spatial masking function $\Lambda(p)$ for modifying Eq. (1). For simplicity, only the luminance dominated structural irregularity inherent in the color image is investigated, while considering the fact that the human eye is more sensitive to luminance than to chrominance.

Based on the concept of internal generative mechanism of human of brain theory, the structural irregularity of an image is from the uncertain information which is hard to be predicted for the HVS. We reasonably regard the uncertain information of the image as the residual part between the image and its prediction part [9]. An computational prediction autoregressive (AR) model for the luminance component of the color image is therefore exploited and given by

$$L'_p = \sum_{p_i \in \Re} k_i L_{p_i} + \nu_p \tag{2}$$

where L'_p is the predicted luminance value of pixel p, p_i the i-th neighboring pixel in the surrounding region $\Re = \{p_1, p_2, \ldots, p_N\}$ and $\{k_i\}$ the model coefficients which are determined by minimizing the variance of the white noise $\{\nu_p\}$. The residual part is then computed as the uncertain information to construct the relation between structure-based adjustment and structural irregularity.

III. Verification of the Improved Variable Just Noticeable Color Difference

The performance of estimating the improved variable just noticeable color difference is verified by incorporating Eq. (1) into Chou's model [4] to compare the accuracy of estimating the visibility thresholds of color images. For a particular color pixel, the perceptual redundancy is quantitatively measured by analyzing the loci of colors which are perceptually indistinguishable from this color. In [9], the loci form a sphere centralized at this color's coordinate with the radius of VJNCD in the uniform CIELAB color space and used to compute the visibility thresholds of color pixels in color images. The improved VJNCD of the color pixel p within a complex image is then redefined as

$$VJNCD'_p = JNCD_{Lab} \cdot s_c(a_p, b_p) \cdot \Lambda(p) \tag{3}$$

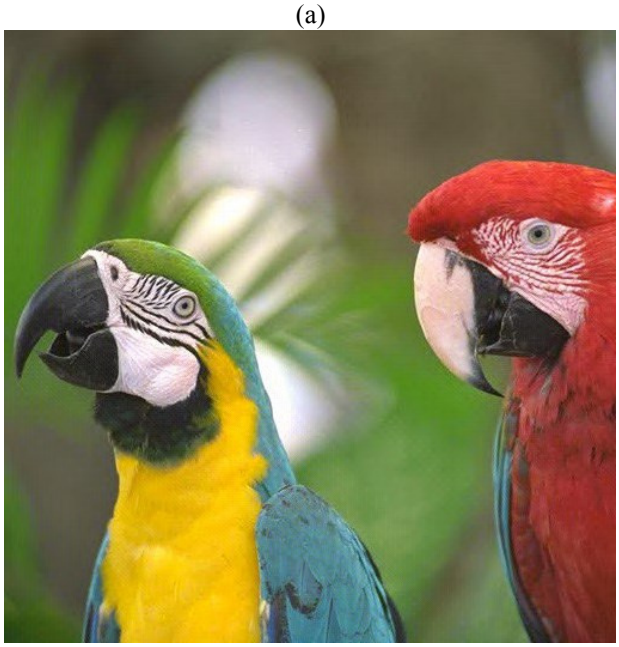
The procedure for estimating the perceptual redundancy of a pixel in an arbitrary color space is firstly to transform the color pixel to the CIELAB space. By using Eq. (1) that utilizes the improved spatial masking function in this paper, the corresponding improved VJNCD threshold is obtained. Under the perceptually conservative restriction controlled by the luminance, some critical colors on the surface of the improved *VJNCD* sphere are selected to transform back to the target color space. Finally, an approximate rectangular subspace is obtained to quantify the perceptual redundancy and the visibility thresholds of the color pixel for each color channel are calculated.

The verification of the improved variable just noticeable color difference for color images is inspected by comparing the accuracy of estimating the visibility thresholds of color images in each color component. Herein, a subjective test is conducted to inspect if the estimated visibility thresholds is consistent with the HVS. Suppose a test image is represented in the YCbCr color space, the visibility threshold for each color pixel in each color component of the color image is randomly added to or subtracted from the corresponding color pixel. The accuracy of estimating the visibility thresholds is better if the PSNR of the contaminated image is lower, while the visual quality of the contaminated image has nearly the same as the original image under the specified viewing condition.

IV. Simulation Results

In the viewing test, the original color image and its noise contaminated image are randomly placed side by side on the monitor. The test is carried out in the dark room when the subject observes the image pair on the monitor at a viewing distance of six times the image's height. In the simulation, a variety of standard test color images are used. The color pixels are represented in YCbCr format. 16 subjects who have normal eyesight or had been corrected to be normal take part in test.

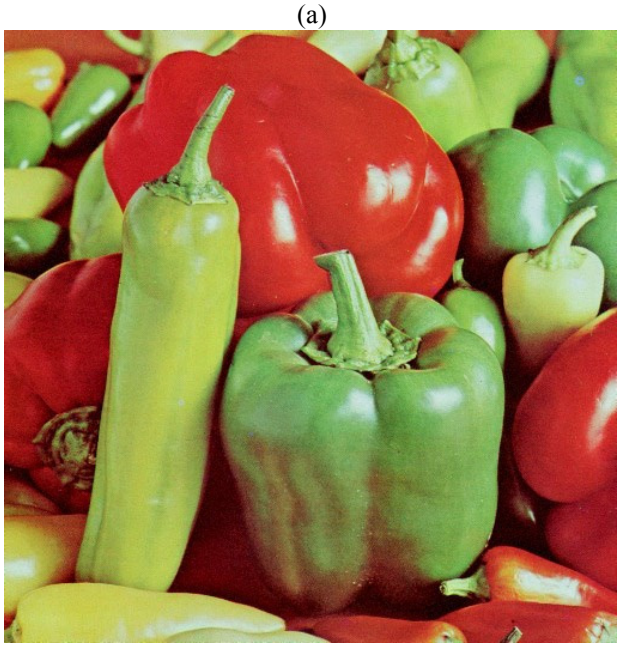
(a)



(b)

Fig. 1. (a) Original color image of "Zuerst" and (b) its noise contaminated image (PSNR=34.95B)

(a)



(b)

Fig. 2. (a) Original color image of "Pepper" and (b) its noise contaminated image (PSNR=33.20dB)

In Fig. 1, the original color image of "Zuerst" (Fig. 1a) and its noise contaminated version (Fig. 1b)of PSNR=34.95dB by the associated visibility thresholds in three color components are shown, while the two images are perceptually indistinguishable from each other under the viewing condition mentioned above. To achieve a fair comparison for verifying the improvement of the existing variable just noticeable color difference, the simulation results are compared with the color visual model proposed in [4]. The PSNR of the noise contaminated color "Zuerst" image by the method presented in [4] is 36.87dB at nearly the same visual quality. Fig. 2 shows the experimental results for "Pepper" color image. It is shown that the improved spatial masking estimation based on the uncertain information indeed achieve larger noise concealment in the regions with structural irregularity, while the visual quality of the noise contaminated image has nearly the same as the original image under the specified viewing condition. Under the same perceptually indistinguishable visual quality for 12 standard test color images, the average PSNR of the noise contaminated color images by the proposed method is 1.4dB lower than that by the Chou's method. The

proposed spatial masking adjustment successfully shows better performance.

V. Conclusions

In this paper, the estimation of just noticeable color difference in color images is improved by using a new spatial masking. By using the brain works of the human visual perception which is sensitive to the orderly stimulus that is easily predicted and is insensitive to the disorderly stimulus, the estimation is incorporated with a simple prediction model to effectively obtain a new spatial masking function. We use the function to improve the just noticeable color difference for computing the visibility thresholds of color images for performance comparison. With the new spatial masking function, the proposed spatial masking adjustment successfully shows better performance than the existing masking method.

Acknowledgment

The work was supported by the Ministry of Science and Technology, Taiwan (R.O.C.), under contract MOST 104-2221-E-278-001, and Image & Video Processing Laboratory of Department of Computer Science and Information Engineering, National Dong-Hwa University, Taiwan.

References

[1] H. Tian, Y. Fang, Y. Zhao, W. Lin, R. Ni, and Z. Zhu, “Salient Region Detection by Fusing Bottom-Up and Top-Down Features Extracted From a Single Image”, IEEE Trans. Image Process., vol. 23, no. 10, pp. 4389-4398, 2014.

[2] H. R. Sheikh and A. C. Bovik, “Image information and visual quality,” IEEE Trans. Image Process., vol. 15, no. 2, pp. 430–444, Feb. 2006.

[3] I. Höntsch and L. J. Karam, “Locally adaptive perceptual image coding,” IEEE Trans. Image Processing, vol. 9, no. 9, pp. 1472-1483, Sep. 2000.

[4] C. H. Chou and K. C. Liu, "Colour image compression based on the measure of just noticeable colour difference," IET Image Processing, vol. 2, no. 6, pp.304-322, Dec. 2008.

[5] P. B. Nguyen, A. Beghdadi, M. Luong, "Perceptual watermarking using a new JND model," Signal Processing: Image Communication, vol. 28, no. 10, pp. 1506-1525, Nov. 2013.

[6] T. Zhu and L. Karam, "A no-reference objective image quality metric based on perceptually weighted local noise,” EURASIP Journal on Image and Video Processing, vol. 1, no. 5, pp. 1 - 10, 2014.

[7] S.-L. Hsieh, C.-C. Chen, and W.-S. Shen, "Combining Digital Watermarking and Fingerprinting Techniques to Identify Copyrights for Color Images," Scientific World Journal, Jan. 2014 http://dx.doi.org/10.1155/2014/454867.

[8] K. Friston, “The free-energy principle: A unified brain theory?” Nat. Rev. Neurosci., vol. 11, no. 2, pp. 127–138, Feb. 2010.

[9] J. Wu, G. Shi, W. Lin, A. Liu, and F. Qi, “Just Noticeable Difference Estimation For Images with Free-Energy Principle”, IEEE Trans. Multimedia, vol. 15, no. 7, pp. 1705-1710, Nov. 2013.

[10] M. Mahy, L. Van Eyckden, and A. Oosterlinck, “Evaluation of uniform color spaces developed after the adoption of CIELAB and CIELUV,” Color Res. Appl., vol. 19, no. 2, pp. 105-121, 1994.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 807–816
DOI: 10.15439/2016F233
ACSIS, Vol. 8. ISSN 2300-5963

Content Based Image Retrieval using Query by Approximate Shape

Stanisław Deniziak
Kielce University of Technology
al. Tysiaclecia Panstwa Polskiego 7, 25-314 Kielce, Poland
Email: s.deniziak@tu.kielce.pl

Tomasz Michno
Kielce University of Technology
al. Tysiaclecia Panstwa Polskiego 7, 25-314 Kielce, Poland
Email: t.michno@tu.kielce.pl

***Abstract*—In this paper we present a new method for content-based image retrieval. The method is based on querying database by approximate shape representing given object. In this way all images containing the object may be found. Shapes are specified as a set of geometric primitives and attributes. Relations between primitives are represented by a graph. Our graph matching algorithm is used for computing the level of similarity between shapes. The method may be also used for searching transformed as well as partially covered objects. Experimental results showed the efficiency of our approach.**

I. Introduction

SEARCHING for specific images or any graphical objects is one of the most challenging problem in image analysis, called image retrieval. The large image data sets are stored in databases, file systems, Internet resources, and other repositories. Moreover, the size of each individual image is increasing due to the development of new high-resolution sensors. Methods which are used in order to retrieve images may be grouped into the following three categories: Keyword Based Image Retrieval (KBIR), Semantic Based Image Retrieval (SBIR) and Content Based Image Retrieval (CBIR) [1].

The KBIR methods are based on metadata which describes images stored in the database. During the query, the set of keywords is compared with the metadata- textual description for each image, which most often contains also a set of keywords. KBIR methods provide good results and are fast, but their quality depends on the image annotations. In many cases image descriptions are not available, annotations are incomplete or not ambiguous. For example the person who is preparing keywords for images may not have enough knowledge or identify some objects wrongly. There are also some algorithms which tries to prepare keywords automatically without human interaction, but they suffer for the same problems. Moreover the number and precision of descriptions is also very important. In order to overcome problems with proper keywords, the CBIR algorithms were proposed [2]. In this group, a query has a form of a sample image, which is used to retrieve similar images. In SBIR methods queries concern semantic meaning of the image. They are based on textual annotations or automatic image recognition.

In this paper, a new Content Based Image Retrieval method is proposed. The idea is based on image decomposition into primitives with attributes. The detected primitives are transformed into a graph which is used for object matching. The main advantage of our method is that it is not based on strict comparison of objects, but it compares approximate shapes, therefore it may be applied for transformed (e.g. scaled, rotated, tilted) as well as partially covered objects.

The paper is organized as follows: next section describes the review of existing content-based image retrieval methods and the motivation to this work. Our algorithm is presented in the section 3. Finally, experimental results, conclusion and further directions of our research are given.

II. Content-Based Image Retrieval

Two types of CBIR algorithms may be distinguished: low-level approach and high-level one. The low-level algorithms, during image processing, use the whole frame, e.g. computing normalized color histogram [3], a difference moment and entropy [5], spatial domain [4] or MPEG-7 shape and texture descriptors [6]. The results of low-level algorithms are most often precise when searching for the whole image (e.g. a painting), but when querying for a specific object they are insufficient. In this case, the high-level algorithms provides much better results. Most of them are based on regions which are groups of similar pixels. After region extraction, they are transformed into graphs and compared with graphs from the database. During the region extraction different techniques are used, e.g. color-based or fuzzy patterns recognition [7]. This group of algorithms provides very precise results but are problematic for users when they does not have full knowledge about searched objects or does not have sample query image.

Our work focuses on CBIR method based on query by shape. Such methods are useful when a user looks for images containing given class of objects, and the query is specified as an approximate sketch drawn manually or using graphical editor. There are some methods which deal with this problem. In [8] the human drawn image sketches are compared, but their algorithm relied on low-level image sketch and was not oriented on objects. There are a lot of methods for shape extraction from images for image retrieving. They are based of FFT [10] or on extracting some features of the shape [9]. Others are using stroke points with gradient fields which are combined with Poisson HOG [15] or edge based shape vectors [16]. There were also some attempts to use both raster and vectorized image, e.g. using color moment and topogeo

978-83-60810-90-3/$25.00

descriptors [17]. It was shown that shape-based image retrieval methods are very efficient. But existing methods are based on strict shape matching, therefore transformed or partially covered shapes may not be recognized. In [11] a method where query image may be a hand-drawn sketch was proposed. The method compares the query with the whole image using wavelet decompositions.

In our previous work [12], [13], [14] we proposed some high-level methods which are based on predefined shapes, easy to drawn by a human but also to extract from the query image. Our preliminary works showed that our graph-based representation of shapes is very suitable for object matching, even when object are deformed. Hence, we decided to continue our research. The goal of this research was to propose a new CBIR algorithm which will be suitable not only for queries specified by image objects but also for human drawn queries. The queries decompose object shape into smaller parts which are then compared with shapes extracted automatically from images stored in the database. Moreover, because objects are represented by approximate shapes, the matching is defined by the level of similarity.

Our previous researches were focused only on two primitive types - lines and ellipses. During the tests, we found that this set provide good results for human-based, angular objects, but for the other types, the results may be insufficient. In order to improve the precision of results for such objects, the primitive set had to be extended with rounded ones. Moreover, during experiments we also noticed that in some parts of objects there is very important to store the relations between primitives of the same type. For example, some of the lines in the bicycle object are connected with each others and when combined, they create a chassis. In our previous researches that situation was covered by connections between primitives in the graph, but we decided to strengthen the validity of that fact.

III. Query by Approximate Shape

The method is based on two ideas: an object representation and a matching algorithm. Objects are specified using shapes decomposed into primitives which are extracted from the image or are drawn by an user. Shapes are represented with graphs, where nodes correspond to primitives and edges are between nodes associated with adjoining primitives. After building a graph, the matching algorithm, which compares the queried object with an object from the database, is applied. The system overview is shown in the Fig. 1.

A. Object representation

The main idea of our approach is based on approximation of any shape with limited set of simple primitives. Each primitive shape may have attribute defining size, color, orientation etc. The two most basic types of primitives are line segment and arc. Usually lines and arcs compose more complex shapes, therefore we also consider polylines and poly arcs as primitives. When polyline (or poly-arc) creates closed loop, it creates a plane. Since the plane has additional attributes like texture or color of the surface, we also consider polygons and arc-sided polygons as primitives. Hence, 6 different primitives are distinguished (Fig. 2). All shapes are approximated using the above primitives with attributes. It is very important that approximation should be deterministic, i.e. similar objects should be approximated with similar shapes. Query images as well as images from the database are approximated in the same way.

The predefined set of primitives allows approximation of different shapes, even composed of curved segments. For each primitive we also define the following basic attributes, which describes orientation of the primitive:

- for line segment: the size and the angle defining the slope,
- for polyline and polygon: the number of line segments, attributes of the following line segments,
- for arc: the size and the angle,
- for poly-arc and arc-side polygon: the number of segments, attributes of the following arcs.

In order to allow comparisons of shapes with different sizes, all sizes are normalized and have values between 0 and 1.

The approximation of any object is done using the following procedure: first edge detection is performed. Then all segments are approximated with lines and arches, next all polygons and arc-sided polygons are extracted, finally connected lines and arches are converted into polylines and poly-arcs. Since after line segments detection in real life images very often lines are divided into smaller parts, a line merging should be applied. If the distance between endings of two line segments is below the line merging threshold (Fig. 3 a)), their angles should be tested. If for both segments angles are the same, the merging could be performed. Moreover there may be also a situation when an arc was detected as a set of lines (Fig. 3 c)). In order to detect and convert set of segments into an arc, firstly it should contain at least 3 connected segments with endings in very close distance. Next, the angles between lines should be measured. If their values are the same—the set of segments could be converted to an arc (Fig. 3 d)).

The predefined set of primitives covers most geometric shapes, the Fig. 4 shows the example car (b), bicycle (c) and flower (d) objects. There may be noticed that all circles are detected as arches (with 360° angle). Next, all primitives based on arches and line segments have to be detected. Firstly all polygons and polygon arches are extracted, next polylines and poly-arches. When a primitive is detected, all its line segments or arches must not be used as part of other primitives.

After detection of primitives, the graph representation is being built. During this process, the following relations between primitives are stored:

- which primitives are close to each other or which primitives are connected,
- positions of the above primitives (using 8 directions: N, E, W, S, NW, NE, SW, SE - see Fig. 5).

When some primitives are connected to each other, they create a more complex shape (e.g. adjoined triangles, quadrangles and arches). Because information about which primitives belong to which complex shape and how complex shapes are

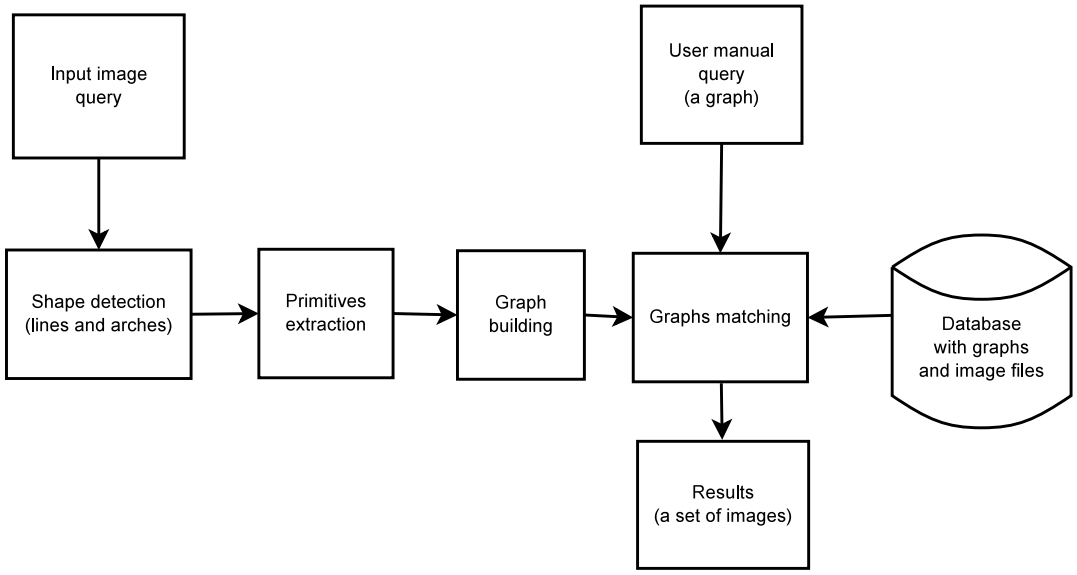


Fig. 1. The system overview.

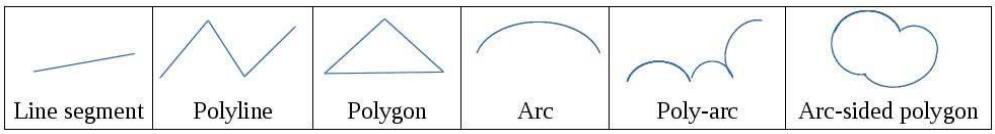


Fig. 2. Predefined primitives

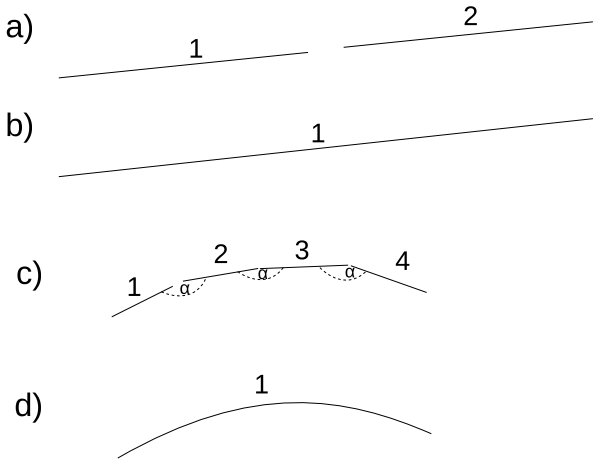


Fig. 3. Lines preprocessing. a) a splitted line into two segments with the same angles and very close distance between endings, b) line after merging, c) lines which may construct an arc d) an arc after lines conversion.

situated to each other may be useful during object matching, it is also stored. Moreover, in real life images, sometimes primitives which should be connected are not connected but are placed in a very close distance, for example because of inaccuracies of line segment detection algorithm. To overcome this problem, the maximal distance between primitives is used when creating connections.

The Fig. 4 shows the conversion of a sample car, bicycle and flower objects into graphs. Firstly, the algorithm detected all base primitives - arches and lines. For the car and bicycle objects, wheels and gear were detected as arches with $360°$, while their chassis as a set of lines. Some parts of car's body were recognized both as arches and lines. For the flower object, stalk and all petals were detected as arches. The second algorithm step, construct more complex primitives on the basis of the previous detection. Firstly, polygon arches and polylines are extracted as a result of the process of line and arches endings examination. For example, if the distance between two lines endings is lower than maximal distance threshold, they are combined as a polyline. The maximal distance threshold was introduced as a result of our previous tests with real life images. Very often we noticed the situation when the lines had to be connected, but as a result of some failures of line detection algorithm there was some space between the endings. In this step, all petals were combined as a one polygon arc, and car's windows and bicycle chassis as polylines. After polylines detection, the algorithm checks if they could construct a polygon (as a result, one of the windows and part of chassis were transformed). The last algorithm step constructs complex shapes and graphs. All detected primitives which are close to each others are combined using maximal distance threshold. All their position relations are stored, as was stated previously (e.g. the stalk arc is slightly on the left and below the petals polygon arc - the SW direction). For

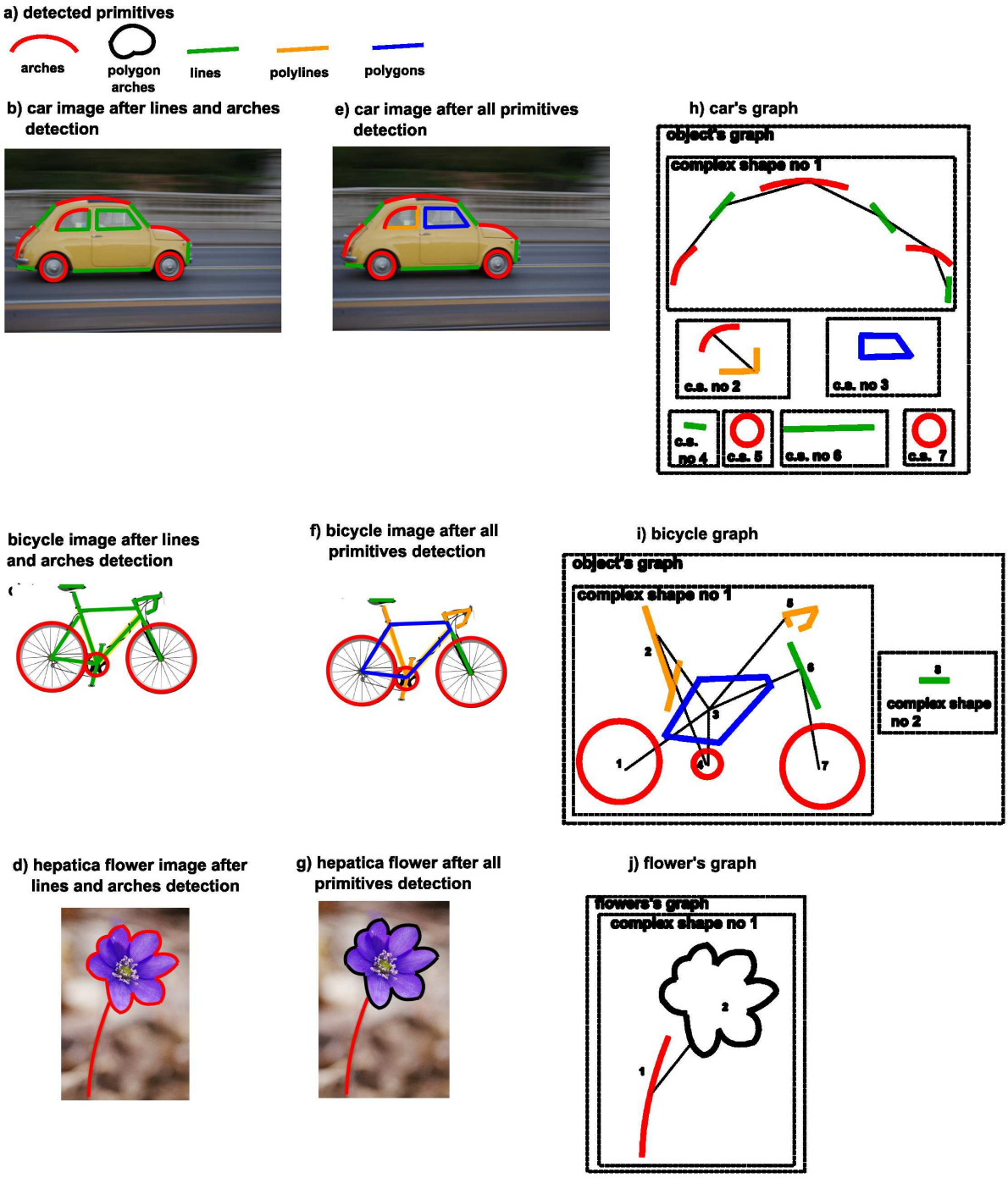


Fig. 4. The examples of primitives detection and graphs used for comparisons.

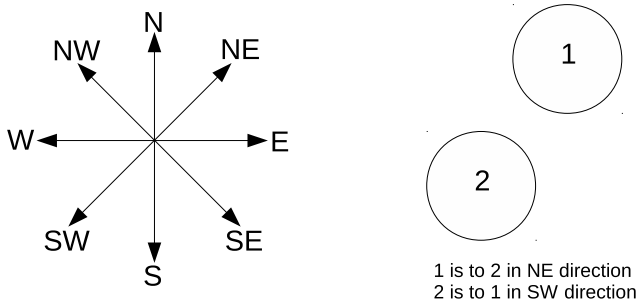


Fig. 5. The directions used in the algorithm and example of usage.

the car, 7 complex shapes were created, because there were unconnected groups of primitives. But in case of the flower, all primitives were connected and created only one complex shape.

B. Matching algorithm

The matching algorithm is partially based on our previous research [12]. All comparisons between objects are performed hierarchically in order to improve the performance, but also to reject objects which are not similar in the first steps. The algorithm is as follows:

Assumptions:

- O - searched object
- M - object from the database with which O is compared
- T_{CS} - threshold which defines the minimal number of complex shapes from O matched with M - values: $< 0, 1 >$, 0 - none, 1 - all
- T_P - threshold which defines the maximal difference between parameters, values: $< 0, 1 >$, 0 means the values are the same
- T_N - threshold which defines the minimal number of connections to other primitives with the same type and mutual position between primitives - values: $< 0, 1 >$, 0 - none, 1 - all
- T_{sim} - threshold which defines the minimal similarity of two objects graphs, values: $< 0, 1 >$, 0 - objects are completely different, 1 - the same objects
- thresholds: T_{CS}, T_P and T_N are used only to improve the performance and to reject not similar complex shapes or primitives in the earliest algorithm stages.

1) Try to match all complex shapes of O to all complex shapes of M, comparing the number and types of primitives and mutual positions.
   a) N_{CSP} = the number of primitives in O's complex shape with the same type as in M's complex shape
   N_{CSO} = the number of primitives in the O's complex shape
   N_{CSM} = the number of primitives in the M's complex shape
   $sim_{CS} = \frac{N_{CSP}}{max(N_{CSM}, N_{CSM})}$
   b) If $sim_{CS} < T_{CS}$ then go to a) and compare O's complex shape with the next M's complex shape.
   c) Add the pair of complex shapes to the list, storing theirs sim_{CS}.
2) For each pair of complex shapes CS_O (O's complex shape) and CS_M (M's complex shape) from the list:
   a) Try to match primitives - compare each primitive P_{CSO} of CS_O with each primitive P_{CSM} of CS_M:
      i) If P_{CSO} and P_{CSM} types are different, check another CS_M's primitive and go to a)
      ii) Compare primitives attributes and compute sim_P coefficient:
      $$c = \sum_{i=1}^{n} |(i^{th}\ attribute\ of\ P_{CSO}) - (i^{th}\ attribute\ of\ P_{CSM})|$$
      $$sim_P = \frac{c}{n}$$
      iii) If $sim_P > T_P$ then check another CS_M's primitive and go to a)
      iv) Compare connections of P_{CSO} and P_{CSM}:
      N_{PCSO} = the number of P_{CSO}'s connection to other primitives with the same type and mutual position as in P_{CSM}
      $$sim_N = \frac{N_{PCSO}}{the\ number\ of\ connections\ in\ P_{CSM}}$$
      v) If $sim_N < T_N$ then check another CS_M's primitive and go to a)
      vi) If $(1 - sim_P) * sim_N$ is greater than previously stored values, store them as a new P_{CSO}'s matching:
      $$P_{sim_{CS}} = (1 - sim_P) * sim_N$$
   b) Modify sim_{cs} value for complex shapes pair CS_O and CS_M:
   $$sim_{CS} = \frac{sum\ of\ P_{sim_{CS}}\ for\ each\ primitive}{number\ of\ primitives\ in\ CS_M}$$
3) For each CS_O choose the CS_M matching using the highest sim_{CS} values. If there are more than one maximum value, use the mutual positions of complex shapes.
4) Compute the graphs similarity coefficient:
   $$sim = \frac{sum\ of\ each\ complex\ shape\ sim_{CS}\ of\ O}{the\ number\ of\ complex\ shapes\ in\ M}$$
5) If $sim < T_{sim}$ then objects are not similar.

IV. Experimental results

The algorithm was firstly evaluated using cars, bicycles and flowers objects. The Fig. 6 and Fig. 7 shows sample images, their graphs and normalized attributes. In the tables, the 'Complex shape' column contains the sequence number (which may be used e.g. as a reference to a specific complex shape). The 'Primitive' column was divided into two sub-columns: 'No' and its 'type'. The 'No' contains the sequence number of a primitive which is unique in the whole image. The

'Attributes' column consists of the 'count' which informs how many sub-primitives constructs the primitive and the 'Values' which is a list of attribute values for each sub-primitive. As mentioned earlier, the attribute is computed as a normalized sub-primitive angle which stores values between 0 and 1. For lines the direction angle is used, for arches the central angle is used.

When comparing the same classes of objects, the sim values should be high. In our tests, for comparison of car 1 with car 2 from Fig. 6 the sim value was equal to 0.75 which means that objects are similar in 75%. Because in both graphs there are two complex shapes containing only one circle (arc with 360°= 1 after normalization), after algorithm execution, there were two candidates of matching for complex shapes no 5 and 7 (in car's 1 graph). In order to choose the best matching, the algorithm checked the mutual positions with other complex shapes. Comparisons with different classes should result with lower sim values. For example when car 1 was compared with a bicycle from Fig. 7 a) and b), the sim value was equal 0.14. For completely different classes, like hepatica flower (Fig. 7 c) and d)), the sim reached 0 value. The hierarchical comparisons enabled faster rejections of not similar complex shapes, which resulted in much lower number of detailed comparisons.

In order to evaluate the algorithm with greater number of images, the prototype application was developed using C++ and OpenCV image processing library. For line detection the Line Segment Detector (LSD) algorithm was used, which is known for providing precise line segment detection results [19]. All detected segments were tested if they can construct one bigger segment and an arc. Moreover, Circular Hough Transform was used to detect circles and improve number of correctly extracted primitives. The images were preprocessed using i.e. morphological operations. After primitives detection, all of them were grouped into complex shapes using minimal distance criterion. During our future research, the primitives extraction algorithm will be refined. As a database, 105 real life images of cars, motorbikes, bicycles and scooters were used. Some images contained background with other objects. In order to evaluate the performance of our algorithm the *precision* and *recall* coefficients were used, defined as following:

$$precision = \frac{number\ of\ relevant\ results\ images}{total\ number\ of\ results\ images} \quad (1)$$

$$recall = \frac{number\ of\ relevant\ results\ images}{total\ number\ of\ relevant\ images\ in\ the\ database} \quad (2)$$

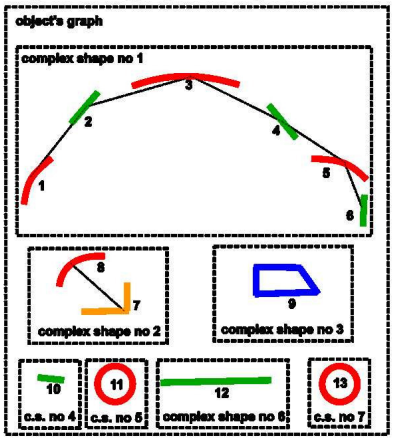
Moreover, the algorithm was also compared with simple region-based method. The precision and recall results for example car, bicycle, scooter and motorbike query images (Fig. 8) are presented in the Table I. It can be seen that our algorithm provided much higher $precision$ values in comparison to the region-based method. For the region-based algorithm, bicycle objects are problematic due to the small uniform color areas which leads to smaller precision values. Contrary, car objects very often contain many big uniform areas and the results are much more precise. Our algorithm is not sensitive to such problems and for both object classes provided high number of correct results. The Query by Shape algorithm provided the best results for the bicycle sketch image, because it contained only the most important primitives for each bicycle. The worst results were obtained for the scooter object because for some bicycle and motorbike objects it obtained enough sim value to qualify them as a correct results.

One of our aim is to provide the algorithm which is able to perform both types of queries: manual (a sketch manually drawn by an user which is then converted into a graph) and automated (graphs automatically detected from input query images) using the same database without any modifications. Because of that fact, we performed also experiments using manually drawn objects and the same database (with the same automatically generated graphs) as in previous tests. For example the bicycle object (Fig. 9) provided results with 0.77 precision.

Since nowadays more and more people are using web searching engines to find specific images, we performed also another tests in order to compare our algorithm with Google Images Search engine. Due to the fact that our Query by Shape system was run on an average performance laptop and Google Images is using huge data centers to process queries, the results cannot be compared directly, but they show differences between approaches. As the query images, we used the same files as for our algorithm (Fig. 8). The results for almost all images were very precise, because the Google Images Search algorithm tries to firstly find the same images and then if none are available it tries to find similar ones. The Google algorithm during query processing uses not only the image data, but also some textual annotations which were chosen to best describe the image. The most precise results were obtained for Mercedes Benz (Fig. 12) and motorbike queries, which was caused by very precise recognized textual descriptions - "mercedes benz s class 1998" and "yamaha 125". However, Google Image Search did not provide the best results for all queries, e.g. the sketch of a bicycle and the scooter. The bicycle sketch was recognized as a "cyclist", but also as a simple, black and white sketch which resulted in only schematic images in the result set (e.g. our Query by Shape provided all types of bicycle images). For the scooter image, the results were much worse. The keywords assigned to the query were "dk raven" and as a result none scooter images were returned but only one type of a specific kind of a bicycle. The tests showed that Google Images Search engine provides very good results for images which are known for it, e.g. previously indexed with proper keywords. For objects which are new, the results may be moderate or even completely incorrect like for the scooter image. Moreover, the Google Image Search engine tries to find the exact object images (for example the same model, color and year of a car). Our algorithm, Query by Shape, tries to find images of objects with the same class (e.g. a car or a bicycle), allowing different

a) car 1 image after all primitives detection

b) car's 1 graph after primitives

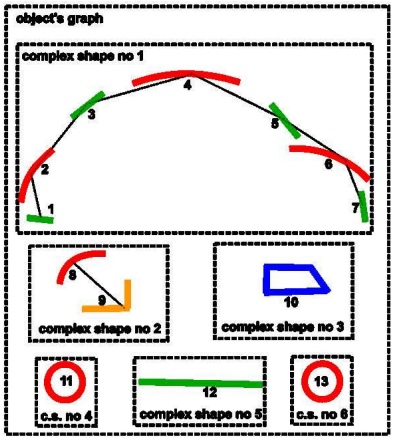


c) car's 1 attributes

Complex shape	Primitive		Attributes	
	No	Type	Count	Values
1	1	arc	-	0,16
	2	line	-	0,14
	3	arc	-	0,22
	4	line	-	0,86
	5	arc	-	0,17
	6	line	-	0,24
2	7	polyline	2	0; 0,25
	8	arc	1	0,31
3	9	polygon	4	0; 0,01; 0,24; 0,85
4	10	line	-	0,8
5	11	Arc	-	1
6	12	Line	-	0,01
7	13	Arc	-	1

d) car 2 image after all primitives detection

e) car's 2 graph after primitives

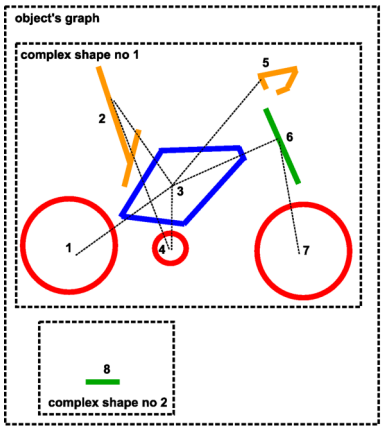


f) car's 2 attributes

Complex shape	Primitive		Attributes	
	No	Type	Count	Values
1	1	Line	-	0,8
	2	arc	-	0,17
	3	line	-	0,11
	4	arc	-	0,22
	5	line	-	0,86
	6	arc	-	0,16
	7	line	-	0,28
2	8	polyline	2	0; 0,25
	9	arc	1	0,31
3	10	polygon	4	0; 0,01; 0,24; 0,84
4	11	Arc	-	1
5	12	Line	-	0
6	13	Arc	-	1

Fig. 6. The example attributes of two compared cars.

a) bicycle graph after primitives extraction

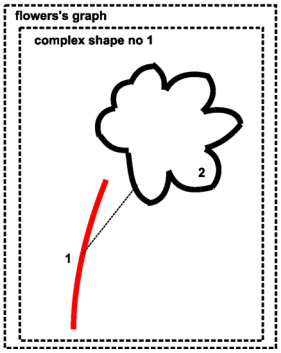


b) bicycle's attributes

Bicycle:

Complex shape	Primitive		Attributes	
	No	Type	Count	Values
1	1	Arc	-	1
	2	polyline	2	0,21; 0,8
	3	polygon	5	0; 0,16; 0,83; 0,87; 0,98
	4	Arc	-	1
	5	polyline	4	0,03; 0,06; 0,68; 0,83;
	6	line	-	0,81
	7	Arc	-	1
2	8	line	-	0

c) flower's graph after primitives



d) flower's attributes

Hepatica flower:

Complex shape	Primitive		Attributes	
	No	Type	Count	Values
1	1	Arc	-	0.08
	2	Polygon arc	2	0.35; 0.36; 0.21; 0.23; 0.17; 0.13; 0.13; 0.24; 0.19; 0.17; 0.16; 0.15

Fig. 7. The example attributes of bicycle and flower objects.

TABLE I
THE PRECISION AND RECALL RESULTS FOR CHOSEN TEST OBJECTS

object	Query by Shape		Region-based	
	precision	recall	precision	recall
car (Fiat 500)	0.89	0.33	0.53	0.75
car (Mercedes Benz)	0.79	0.73	0.51	0.5
bicycle	0.93	0.37	0.23	0.42
bicycle (a sketch)	1.0	0.60	0.28	0.47
motorbike	0.86	0.40	0.75	0.4
scooter	0.67	1.0	0.21	1.0

colors, orientations and other differences in attributes.

V. CONCLUSION AND FUTURE WORKS

In this paper the new CBIR algorithm, using query by approximate shape, was presented. The idea of the method is based on decomposition of shapes into smaller segments - primitives, which are described by theirs attributes. Based on detected primitives, a graph representation of the shape is built, then it is compared with graphs stored in the database. The algorithm is suitable for queries using input image as well as for human-drawn queries. In comparison with our previous research a complete set of primitives was defined, a new graph constructing procedure was used, the matching

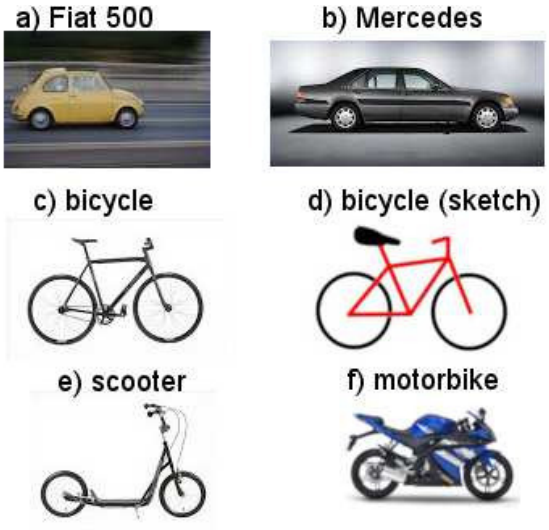


Fig. 8. Example image objects used for tests.

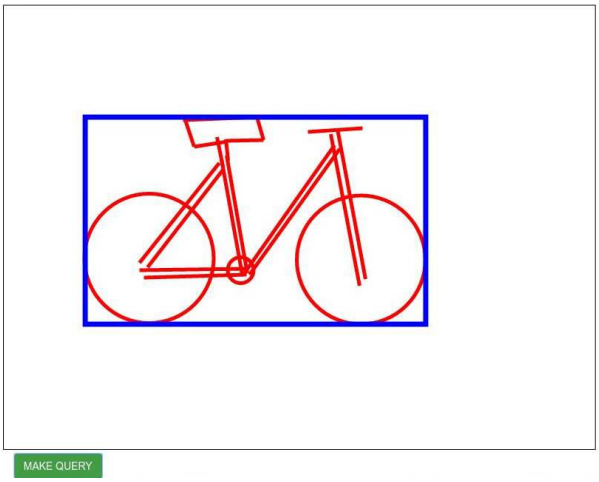


Fig. 9. A bicycle drawn as a manual query. The blue rectangle indicates the detected object during graph constructing stage, the red lines shows lines and arches drawn by an user.

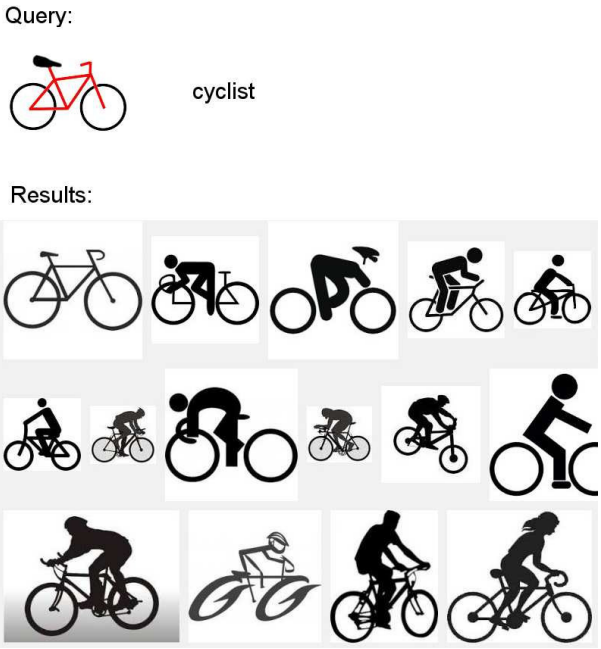


Fig. 10. Bicycle sketch search example using google images.

algorithm was improved and adopted to complex primitives. The experimental results, especially in conjunction with our previous works, are very promising. The main advantage of our approach is that it may be applied to transformed or partially covered objects.

In the future research we will evaluate our approach using greater number of object classes from available image databases, and we also compare the efficiency of our method with more existing state of the art CBIR approaches. Moreover, some modification should be added in order to add ability to achieve better matching e.g. to detect mirrored objects. More advanced graph matching should be also performed, e.g.

Fig. 11. Scooter search example using google images.

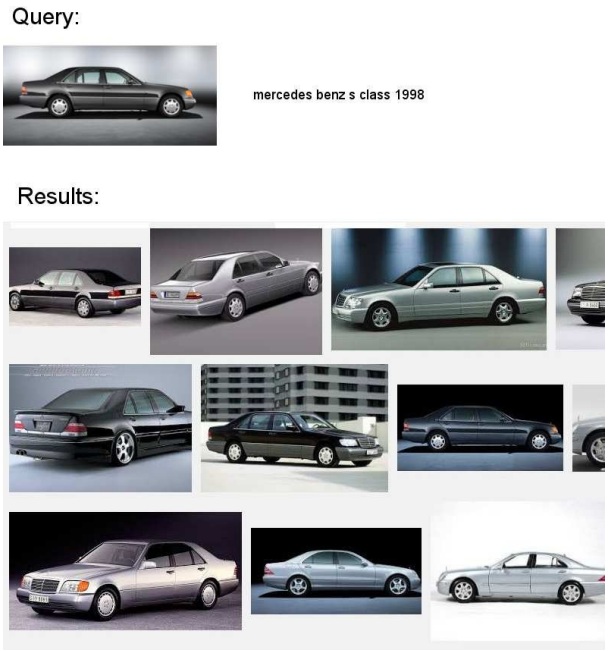


Fig. 12. Mercedes search example using google images.

solving an optimization problem with constraints [18]. Another direction of future research is an efficient storing of objects graphs in the database. Some initial works were performed in [14], but more advanced research should be done.

ACKNOWLEDGMENT

The research used equipment funded by the European Union in the Innovative Economy Programme, MOLAB - Kielce University of Technology.

REFERENCES

[1] H. H. Wang, D. Mohamad, and N. A. Ismail, "Approaches, challenges and future direction of image retrieval" Journal of Computing, vol.2, No.6, 2010, pp. 193-199

[2] R. Datta, D. Joshi, J. Li, J.Z. Wang "Image Retrieval: Ideas, Influences, and Trends of the New Age." ACM Computing Surveys, 40, 2, 2008, 5:1–5:60, doi: 10.1145/1348246.1348248

[3] M. Mocofan, I. Ermalai, M. Bucos, M. Onita, and B. Dragulescu, "Supervised tree content based search algorithm for multimedia image databases", 6th IEEE International Symposium on Applied Computational Intelligence and Informatics, 2011, pp. 469–472, doi: 10.1109/SACI.2011.5873049

[4] T. K. Shih, "Distributed multimedia databases" T. K. Shih, Ed. Hershey, PA, USA: IGI Global, ch. Distributed Multimedia Databases, 2002, pp. 2–12

[5] H.-P. Kriegel, P. Kroger, P. Kunath, and A. Pryakhin, "Effective similarity search in multimedia databases using multiple representations" in 12th International Multi-Media Modelling Conference Proceedings, 2006, pp. 389–392, doi: 10.1109/MMMC.2006.1651355

[6] C. Lalos, A. Doulamis, K. Konstanteli, P. Dellias, and T. Varvarigou, "An innovative content-based indexing technique with linear response suitable for pervasive environments" in International Workshop on Content-Based Multimedia Indexing, 2008, pp. 462–469, doi: 10.1109/CBMI.2008.4564983

[7] M. Bielecka and M. Skomorowski, "Fuzzy-aided parsing for pattern recognition" in Computer Recognition Systems 2, ser. Advances in Soft Computing, M. Kurzynski, E. Puchala, M. Wozniak, and A. Zolnierek, Eds. Springer Berlin Heidelberg, vol. 45, 2007, pp. 313–318, doi: 10.1007/978-3-540-75175-5_39

[8] T. Kato, T. Kurita, N. Otsu, and K. Hirata, "A sketch retrieval method for full color image database-query by visual example" in 11th IAPR International Conference on Pattern Recognition, Vol.I. Conference A: Computer Vision and Applications, 1992, pp. 530–533, doi: 10.1109/ICPR.1992.201616

[9] J. F. Nunes, P. M. Moreira and J. M. R. S. Tavares, "Shape based image retrieval and classification", 5th Iberian Conference on Information Systems and Technologies (CISTI), 2010

[10] D. Zhang, G. Lu "Shape-based image retrieval using generic Fourier descriptor" Signal Processing: Image Communication. 17, 10, 2002, pp. 825–848

[11] C. E. Jacobs, A. Finkelstein, D.H. Salesin "Fast Multiresolution Image Querying" Proc. of the 22nd Annual Conference on Computer Graphics and Interactive Techniques, 1995, pp. 277–286

[12] S. Deniziak, T. Michno "Query by Shape for Image Retrieval from Multimedia Databases." Communications in Computer and Information Science, Springer, 521, 2015, pp. 377–386, doi: 10.1007/978-3-319-18422-7_33

[13] S. Deniziak, T. Michno "Query-by-Shape Interface for Content Based Image Retrieval" 8th IEEE International Conference on Human System Interactions (HSI), 2015, pp. 108–114, doi: 10.1109/HSI.2015.7170652

[14] S. Deniziak, T. Michno, A. Krechowicz "The Scalable Distributed Two-layer Content Based Image Retrieval Data Store" 8th International Symposium on Multimedia Applications and Processing, Federated Conference on Computer Science and Information Systems (FedCSIS), 2015, pp. 827–832, doi: 10.15439/2015F272

[15] Chao Ma, Xiaokang Yang, Chongyang Zhang, Xiang Ruan, and Ming-Hsuan Yang, "Sketch Retrieval via Local Dense Stroke Features" Image and Vision Computing (IVC), 2016, doi: 10.1016/j.imavis.2015.11.007

[16] R. Krishnamoorthy, , S. Sathiya Devi, "Image retrieval using edge based shape similarity with multiresolution enhanced orthogonal polynomials model" Digital Signal Processing, Volume 23, Issue 2, March 2013, pp. 555–568 , doi: 10.1016/j.dsp.2012.09.018

[17] A. S. Mouratoa, R. Jesus, "Clip art retrieval using a sketch. Tablet application." Conference on Electronics, Telecommunications and Computers – CETC 2013, doi: 10.1016/j.protcy.2014.10.246

[18] P. Sitek, J. Wikarek, "A Hybrid Programming Framework for Modeling and Solving Constraint Satisfaction and Optimization Problems." Scientific Programming, vol. 2016, Article ID 5102616, 13 pages, 2016. doi:10.1155/2016/5102616

[19] R. Grompone von Gioi, J. Jakubowicz, J.-M. Morel, G. Randall, "LSD: a Line Segment Detector", Image Processing On Line, 2 (2012), pp. 35–55, doi: 10.5201/ipol.2012.gjmr-lsd

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 817–822
DOI: 10.15439/2016F136
ACSIS, Vol. 8. ISSN 2300-5963

Studying the influence of object size on the range of distance measurement in the new Depth From Defocus method

Krzysztof Murawski
Military University of Technology
ul. Kaliskiego 2,
01-489 Warsaw, Poland,
IEEE Member # 92707852
Email:
Krzysztof.Murawski@wat.edu.pl

Artur Arciuch
Military University of Technology
ul. Kaliskiego 2,
01-489 Warsaw, Poland,
Email:
Artur.Arciuch@wat.edu.pl

Tadeusz Pustelny
Department of Optoelectronics
Silesian University of Technology
ul. Krzywoustego 2,
44-100 Gliwice, Poland
Email:
Tadeusz.Pustelny@polsl.pl

***Abstract*—The article presents new results achieved during researching the distance measuring method that is a part of Depth From Defocus techniques. The method has been developed to determine the shape of the flaccid diaphragm used in the Ventricular Assist Device (VAD). The shape is determined on the basis of distance measured between the CCD sensor plate of the camera and objects (markers) located on the flaccid diaphragm. Experiments were carried out using a stationary camera and circular markers with a diameter from 3 mm to 9 mm.**

The goal of this paper is to present the influence of the object (marker) size on the distance range measured between the camera and diaphragm used in the external pneumatic prosthetic heart.

I. Introduction

THE ARTICLE presents the impact of the size of the observed object (marker) on the result of the distance measurement method presented in [1], [2]. This method is part of the group of techniques defined by the formulated Depth From Defocus (DFD). It was developed specifically for the video sensor determining the momentary volume of ejected blood (*SV*) from the blood chamber of the pneumatic artificial heart. In [1], [2], the image is produced by a stationary camera equipped with a lens with a fixed focus. The camera is placed above the diaphragm so as to allow the observation of its entire surface. The proposed solution is characterized by the fact that during distance measurement the position of the camera and all lens and camera settings (focus, aperture, focal length) remain unchanged. Only the position of the observed markers located on the flaccid diaphragm is subject to change, Fig. 1a, which is located between the blood chamber and the air chamber, Fig. 1b. Method [1], [2] simultaneously determines the position of all of the markers in 3D space based on an analysis of only one image frame. In this respect, it has no equivalent in the literature. The vision systems, mentioned in the literature, used for measuring the distance consist of a light source and a camera, which usually form a stereoscopic system [3]. The distance in such systems is calculated by knowing the optical parameters of the cameras and their mutual position. Measuring systems equipped with one camera are also not rare [4]. The camera then performs two to eight photos of the object [4], and the distance is determined on the basis of the inverse perspective transformation [5]. There is also a measuring system variant with a camera equipped with autofocus. Such a system is calibrated with the help of a standard with known parameters. Measuring the distance to the object then depends on taking a picture at a specific focus setting and calculating the distance using the lens equation [6]. Other methods used to measure the distance is photogrammetry [7], [8] and the fringe projection technique [9] –[11]. Areas of application are, however, limited by the scope of the distances, the speed of the autofocus settings, image processing time, the resolution of the camera image sensor, the number of processed frames per second, as well as the dimensions and weight of the sensor.

In the application, considered in the article, real-time operation is particularly important with the simultaneous determination of the location of 49 markers located on the flaccid diaphragm in 3D space, Fig. 1a, which shape can change with the frequency of up to 3 Hz.

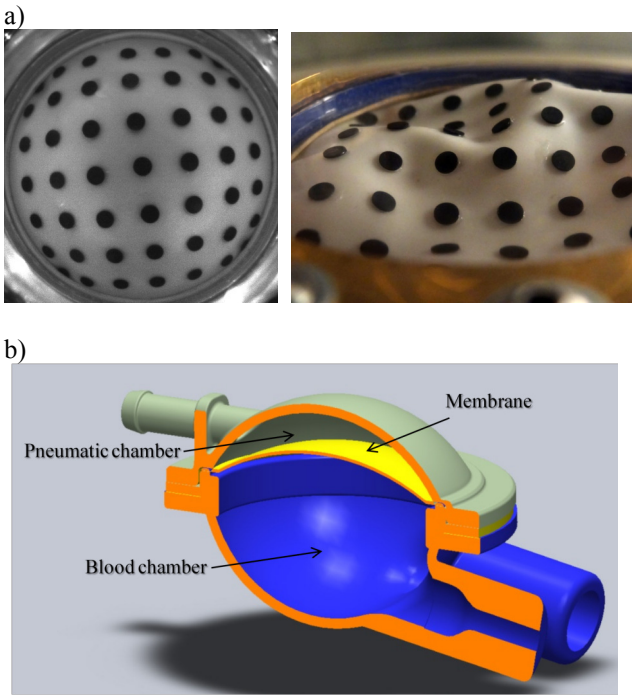


Fig. 1 View of flaccid diaphragm (a), location of the observed diaphragm in the artificial heart prosthetic (b)

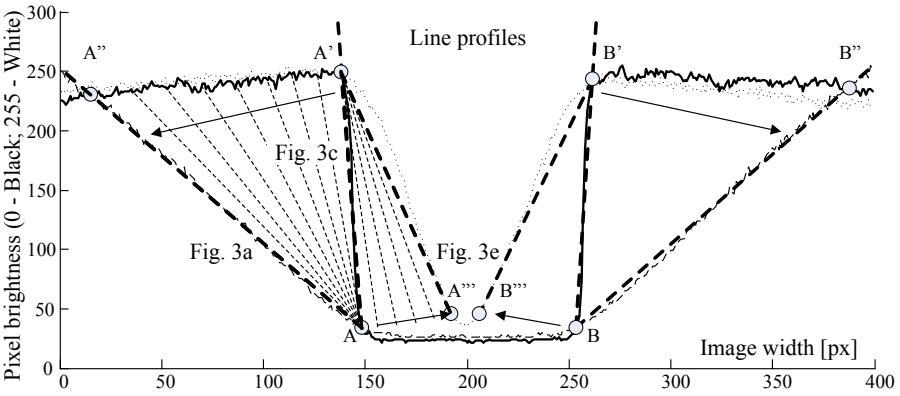


Fig. 2 Profile images of horizontal lines designated by the center of gravity of the object (the view on the standardization of brightness) [1]

II. New Depth From Defocus Method [1], [2]

The essence of the presented measurement solution is based on the analysis of the image produced by the marker moving away from its location for which its focus was set. Figure 2 shows the profiles of horizontal lines of images indicated by the center of gravity of the marker as well as their behavior during changes in the distance of the marker from the camera. For sharp views of marker lengths AA’ and BB’ in Fig. 2 are almost vertical, which indicates a sharp cut-off of the view of the marker from the background. When the marker approaches the camera image blurs. As a result, points A’ and B’ move respectively in direction A” and B”. At the same time the distance between points A and B remains practically unchanged. When the marker moves away from the camera points A” and B” return to their original position (A” $\rightarrow$ A’, B” $\rightarrow$ B’), and the marker image becomes sharp. The marker further distancing from the camera results in a decrease in the view of the marker and the image blurs. In the case under consideration a displacement of points A $\rightarrow$ A”’ i B $\rightarrow$ B”’is observed. The position of points A’ and B’ remains unchanged. The sequence of changes in the position of points A, A’, B, B’ determined when approaching and moving away from the marker with the camera at a step ΔL = 0.01m is presented in Fig. 3. Changes in positions of points A, A’, B, B’ shown in Fig. 2 and Fig. 3 are the basis for determining the distance marker. For this purpose, the image from the camera is subjected to defuzzification. Defuzzification was performed using image binarization with a threshold T_H equal to 70. The selection of the binarization threshold consists of determining such a T_H value, so that a uniform distribution of intersecting points of horizontal image line profiles is obtained (indicated by the marker center of gravity) with a line showing the T_H test value, Fig. 3. The location of the determined points univocally associated with distance d of the marker to the plane of camera image, Fig. 3. The points distribution is described by the equation during the calibration process. The result is a relation describing the distance of the marker to the image plane of the camera.

III. Motivation

The motivation to work on video sensor hardware and software (soft-sensors) to measure the *SV* pneumatic pulsating heart supporting pump were the test results obtained in the framework of the "Polish Artificial Heart"

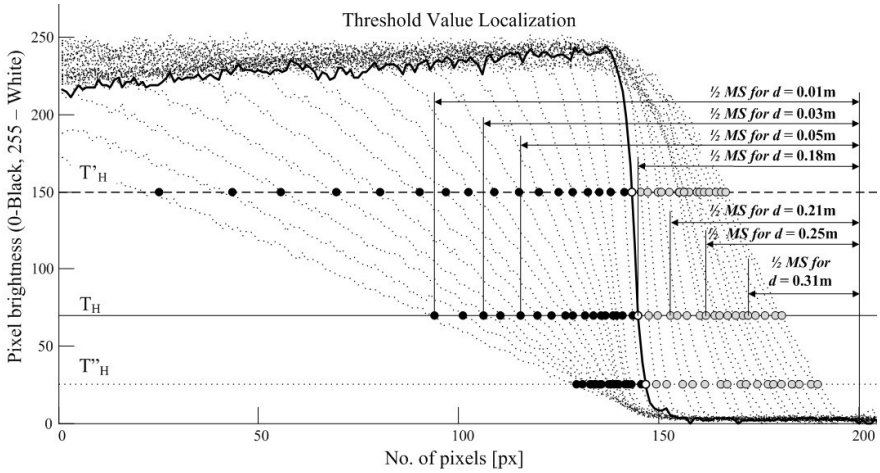


Fig. 3 The distribution of points relevant due to determining the distance to the marker for the established threshold T_H at L_{MIN} = 0.07m, L_{MAX} = 0.42m and ΔL = 0.01m [1].

project [12]. Both [13], [14] have demonstrated that the instantaneous stroke volume can be determined by an acoustic technique using the Helmholtz resonator theory. Knowing the limitations of the developed method the project also examined the possibility of using other techniques and measuring devices [15], including the video camera [16]. [16] shows that the measurement of momentary stroke blood volume of the ReligaHeart® EXT heart prosthesis using a video camera and a marker is ambiguous and burdened with error. The experiment described in [16] was verified by conducting a test using a model of the heart supporting pneumatic pump. In the test, just like in [16], occurrences of stroke volume measurements were noticed for which different shapes of the flaccid diaphragm were characterized with identical positions and marker areas, Fig. 4. This behavior of the marker prevents unambiguous determination of blood stroke volume only on the basis of analyzing changes in its position and the size of the surface area.

Using flaccid diaphragms in the heart prosthesis, even though raises many problems, is necessary and justified from a medical point of view. Such a diaphragm limits the formation of coagulation and eliminates the problem of sedimentation of blood (dividing into fractions). This does not change the fact that the problem of determining the momentary stroke volume of blood from the blood chamber of the pneumatic pump heart assist device (artificial heart) remains unresolved, and the safety of its use is based solely on its visual inspection: *"One of the main advantages of the extracorporeal, polyurethane blood pump is its transparency that allows running continuous visual inspection of the pump status and its quality of work."* [17] As an alternative to the present status it is proposed to use virtual reality technology in the sensor system of the chamber. In the method presented in [18], the camera captures a two-dimensional image of the diaphragm equipped with passive markers, Fig. 5. Markers are used to determine characteristic points of the flaccid diaphragm in 3D space. Knowing the position of these characteristic points in [18] the method of reconstructing the view of the diaphragm was stated.

This method initialized on an IBM PC type computer allows calculating the coordinates of nodal points of the diaphragm and to generate the visualization in 3D space with a frequency of 7 Hz. Exemplary results of the reconstruction

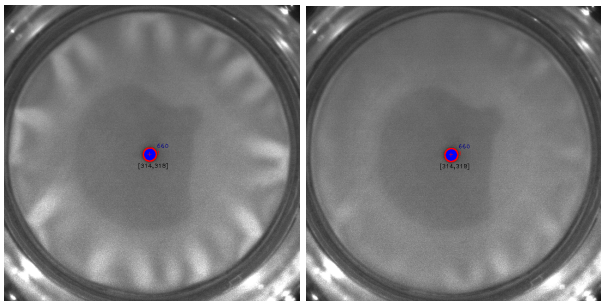


Fig. 4 Unambiguous measurement of SV resulting from the usage of one marker: marker position [314px,318px], surface equalling 660px

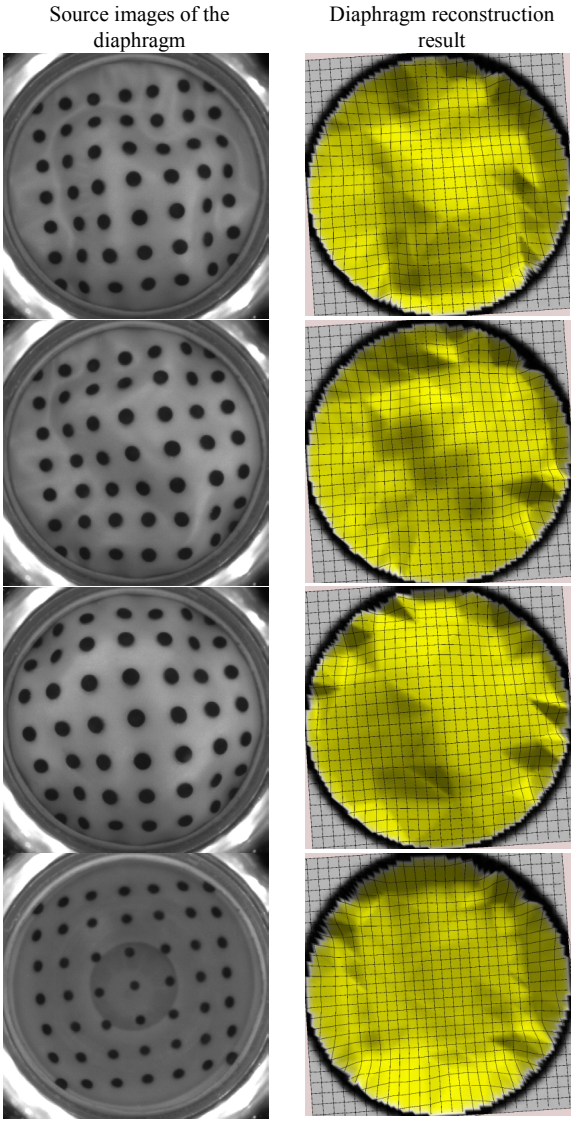


Fig. 5 Reconstructing the view of the flaccid diaphragm based on one photo using the technique provided in [18]

of the view of the diaphragm is shown in Fig. 5. The accuracy in the reconstruction, and thus determining the *SV* depends on [1],[2] by the number and arrangement of the markers as well as the precision of the determined distances. For this reason, the article placed great emphasis on the examination of the influence of the marker size on the value of the determined distance. The acquired dependencies and recommendations to use method [1],[2] was included in part V and VI of this study.

IV. Measurement System Configuration

The influence of the size of the marker on the value of the determined distance was studied using the Optitrack v100: Slim camera. The camera was equipped with a tripod, a lens with a brightness of $F = 2.0$ with a fixed focus of $f = 16$ mm, visible light filter (high-pass filter, $\lambda \geq 850$nm), infrared illuminator with the wavelength $\lambda = 850$ nm and a illuminator driver. The operation of the controller and the IR

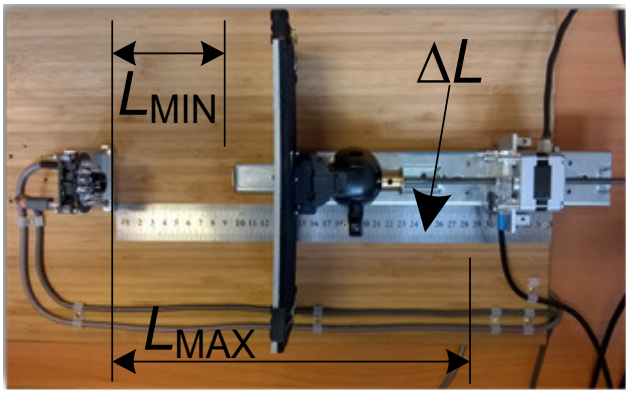


Fig. 6 View of the measuring system

lamp together are discussed in [19] – [21]. The experiment was performed in the configuration shown in Fig. 6. The distance marked in Fig. 6 equaled $L_{MIN} = 0.105$m, $L_{MAX} = 0.175$m, $\Delta L = 0.001$m. In the study, the focus of the lens was set to a distance of $L = 0.14$m, measuring from the image sensor plane of the camera. This position was taken as a reference point d_0 (zero position). All distance measurements were performed from this point. Measurements were performed for the range that included the permissible displacement of the front flaccid diaphragm occurring in the extracorporeal pneumatic heart prosthesis: $d_0 \pm 0.035$m. Studying the influence of the size of the marker area on the scope of measured distances was divided into two stages. The first stage consisted of acquiring images of marker views. Images were acquired in grayscale with a resolution of 640px × 480px. In the tests the distances were determined to the white plane, Fig. 6, bearing the marker. The marker was a black circle with a diameter from 0.003m to 0.009m. Exemplary views of the marker along with the result of its processing are shown in Fig. 7. The position of the plane with the marker was determined by setting it relative to d_0 with an accuracy of ±0.00001m. For this purpose a 39BYGL215A stepper motor was used. The engine with the pusher was placed on a crane. The tip of the

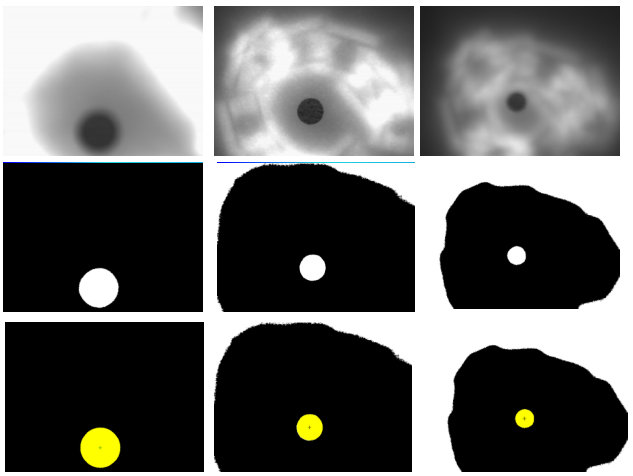


Fig. 7 Marker views (by rows): source image, image after segmentation, image result with the indicated center and analyzed marker surface

pusher was attached to a movable plane, Fig. 6. The straight edge was mounted parallel to the crane, which was used for visually checking the accomplishments of each push given.

The second stage consisted of carrying out activities related to the processing of raster images according to the method provided in [1],[2]. The analysis was then carried out with the aim of searching the connection existing between the initial value of the marker, the range of change observed on the surface of the marker and the obtained measurement range.

V. Results of Research

Studying the effect of the marker area size for distance measurement was performed for markers with a diameter of 0.003m, 0.004m, 0.005m, 0.006m, 0.007m, 0.008m, 0.009m. The experiment was performed in the configuration shown in Fig. 6. All distance measurements were carried out with the same lighting conditions. For each *i*-th marker the marker area was first defined in neutral position d_0. Then the screen was moved towards the camera with a step of 0.001m. This was repeated until achieving a displacement of 0.035m. The displacement was realized by a stepper motor controlled by a microcomputer system. For each of the predefined positions the marker made 100 measurements of its surface area. The test results were accepted as the arithmetic average determined from a series of measurements. After the completion of this experiment the screen with the marker would come back to the zero position. In the next step, measurements were carried out, in which the screen with the marker would move away from the camera image sensor plane. As before, the change in the area of the tested marker was recorded with a step of 0.001m. The last measurement was performed with removing the marker relative to the position for which the sharpness was fixed at 0.035m. These actions were implemented for all seven markers tested. The results are presented in Table I. The surface areas of the markers presented in Table I were determined through operations on the image. For this purpose, for each measurement the following was carried out:

- image defuzzification;
- identifying the location of the spot on the image that matches the location of the marker;
- identifying the location of the center of the marker;
- determining the area of the marker in pixels.

The range of measured distances, in the studied DFD technique, is due to the nature of changes in the surface area of the marker view visible in the image after defuzzification. The resulting variability of the surface areas of the markers in the measuring range of $d0 \pm 0.035$m is shown in Fig. 8. It shows that the least variability of 2383px (reference value), was obtained for the marker with a diameter of 0.003m. The marker with the given diameter, although most promising due to the possibility of a high density of markers on the diaphragm surface does not give the basis (only based on the

TABLE I.
THE MEASUREMENT RESULTS OF THE MARKER SURFACE FOR THE DISTANCE RANGE OF $D_0 \pm 0.035$M WITH A STEP OF 0.005M

	Displacement in relation to position zero in [m]													
	-0.035	-0.030	-0.025	-0.020	-0.015	-0.010	-0.005	0.005	0.010	0.015	0.020	0.025	0.030	0.035
0.003m	3598	3397	3170	2922	2682	2449	2247	1934	1806	1685	1569	1449	1334	1215
0.004m	6793	6270	5746	5251	4787	4370	4013	3443	3224	3007	2796	2598	2407	2222
0.005m	10334	9495	8706	7954	7263	6633	6090	5246	4914	4615	4335	4066	3808	3561
0.006m	14431	13239	12122	11106	10187	9356	8619	7402	6897	6424	5988	5571	5171	4785
0.007m	19719	18077	16556	15174	13910	12789	11778	10109	9442	8823	8249	7716	7214	6735
0.008m	25647	23454	21464	19681	18064	16611	15313	13095	12175	11322	10526	9778	9088	8454
0.009m	39891	36419	33311	30594	28147	25936	23935	20496	19033	17721	16533	15442	14480	13593

A displacement equal to zero was registered for markers: 2070, 3707, 5632, 7962, 10894, 14156, 22122.

knowledge of the measured surface area) to perform an accurate distance measurement. The change in the surface marker with a diameter of 0.003m is determined as a function of distance dependence $f(x) = 6.82x^2 - 279.42x + 3906.4$. A similar result was observed for the marker with a diameter of 0.004m and 0.005m. The variability of these markers equaled 4571px (increase of approx. 2 times) and 6773px (increase of approx. 3 times) accordingly, and the nature of changes are defined by dependencies $f(x) = 16.39x^2 - 579.83x + 7326.20$ for the 0.004m diameter marker and $f(x) = 26.81x^2 - 898.10x + 11{,}148$ for the 0.005m diameter marker. Better quality results were obtained for markers with a diameter of 0.006m, 0.007m, 0.008m and 0.009m. The nature of the changes in their surface areas are defined by third degree polynomials: $f(x) = -1.54x^3 + 70.44x^2 - 1{,}445.7x + 15{,}841$ for 0.006m; $f(x) = -1.93x^3 + 93.99x^2 - 1{,}966.5x + 21{,}631$ for 0.007m; $f(x) = -2.57x^3 + 120.74x^2 - 2{,}541x + 28{,}071$ for 0.008m and $f(x) = -3.56x^3 + 179.39x^2 - 3{,}887x + 43{,}535$ for 0.009m. For the given markers change ranges were determined: 9,646px (increase approx. 4 times), 12,984px (increase approx. 5.5 times) 17,193px (increase of approx. 7 times) and 26,298px (increase of approx. 11 times). The obtained results show that the best of the tested markers to measure distance was the marker with a diameter of 0.009m. The marker with such a diameter, however, is not suitable for use in the task of determining the shape of the diaphragm (flaccid diaphragm) of a pulsatile pneumatic heart assist pump. Due to its size it does not provide the assurance of performing a high density of markers on the surface of the diaphragm, in order to precisely reproduce its shape in a computerized measurement system. Accurate projection of

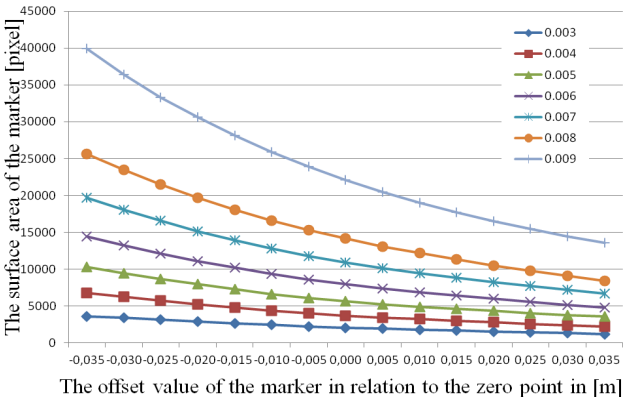


Fig. 8 Measurement results on the variability of surface areas of the markers

the diaphragm shape requires displaying the greatest possible number of markers on its surface, Fig. 5. Hence the simple conclusion that the markers should be as small as possible. Therefore, the analysis of the test results were repeated. At the time it was noticed that the implementation of the standardization of the results with respect to the surface area of the marker captured at point d_0 introduces a significant change in the acquired variable ranges of the marker surface areas, Table II. The graphic visualization of the results are shown in Fig. 9. As a result of carrying out the standardization of the results it turned out that the diameter of the marker had no significant impact on the accuracy of the distance measurement being performed. The shape of the "Mean Value" curve presented in Fig. 10 was described by the polynomial $f(x) = -0.025x^3 + 0.124x^2 - 0.238x + 0.140$, where x is the normalized surface area of the marker.

TABLE. II.

THE MEASUREMENT RESULTS OF THE MARKER SURFACE FOR THE DISTANCE RANGE OF $D_0 \pm 0.035$M WITH A STEP OF 0.005M AFTER STANDARDIZATION

	Displacement in relation to position zero in [m]													
	-0.035	-0.030	-0.025	-0.020	-0.015	-0.010	-0.005	0.005	0.010	0.015	0.020	0.025	0.030	0.035
0.003m	1.738	1.641	1.531	1.411	1.295	1.183	1.085	0.934	0.872	0.814	0.758	0.700	0.644	0.587
0.004m	1.832	1.691	1.550	1.416	1.291	1.178	1.082	0.928	0.869	0.811	0.754	0.700	0.649	0.599
0.005m	1.834	1.685	1.545	1.412	1.289	1.177	1.081	0.931	0.872	0.819	0.769	0.721	0.676	0.632
0.006m	1.812	1.662	1.522	1.394	1.279	1.175	1.082	0.929	0.866	0.806	0.752	0.699	0.649	0.601
0.007m	1.810	1.659	1.519	1.392	1.276	1.173	1.081	0.927	0.866	0.809	0.757	0.708	0.662	0.618
0.008m	1.811	1.656	1.516	1.390	1.276	1.173	1.081	0.925	0.860	0.799	0.743	0.690	0.642	0.597
0.009m	1.803	1.646	1.505	1.383	1.272	1.172	1.082	0.926	0.860	0.801	0.747	0.698	0.654	0.614

For a displacements equaling zero all sizes of determined surface markers took on the value of 1.

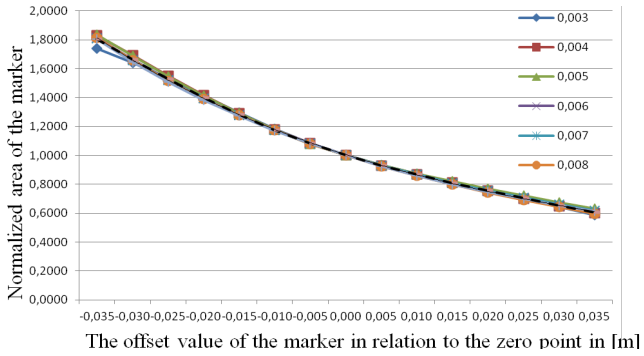


Fig. 9 The measurement results after normalization

For the given equation the average distance measurement error did not exceed ±0.00035m.

VI. Conclusion

The article includes the course and results of the new DFD technique [1],[2]. The purpose of the experiments was to determine the influence of the marker on the result and the scope of the measured distance. Tests were carried out for seven markers with a diameter of 0.003m to 0.009m.

The method of measuring the distance presented in [1],[2] have been developed to determine, in real time, the shape of the flaccid diaphragm [18] of the pulsating pneumatic heart assist pump (artificial heart), Fig. 5. Therefore, it was particularly important to evaluate the effect of the measurement method in the range of movements that the diaphragm is subject to in the prosthetic heart model.

The achieved results show that in the range of d0 ±0.035m the tested technique allows to obtain high accuracy measurements. The measurement error during the test did not exceed ± 0.00035m [1].

Based on achieved results the following recommendations were formulated:

1) where the standardization is not used it is recommended to use markers with a diameter no less than 0.006m;
2) the normalization of the determined surface area of the marker enables the use of markers having a diameter of no less than 0.006m.
3) in order for the distance measurement results to be independent from the diameter of the marker it is recommended to use the standardization of the marker surface area and determine the distance to the object from the polynomial $f(x) = -0.025x^3 + 0.124x^2 - 0.238x + 0.140$, where x is the standardized surface area of the marker.

References

[1] K. Murawski, "Method of Measurement the Distance to an Object Based on One Shot Obtained from a Motionless Camera with a Fixed-Focus Lens", *Acta Physica Polonica A*, 127, 6, pp. 1591 – 1595, 2015. http://dx.doi.org/10.12693/APhysPolA.127.1591

[2] K. Murawski, "Method of measuring the distance using the cameras", *Patent Application No*. P.408076, 2014, (in Polish).

[3] H. Wang, J. Hu, "Active stereo method for three – dimensional shape measurement", *Optical Engineering*, 51, 6, pp. 1 – 8, 2012. http://dx.doi.org/10.1117/1.OE.51.6.063602

[4] SY. Chen, YF. Li, "Finding Optimal Focusing Distance and Edge Blur Distribution for Weakly Calibrated 3-D Vision", *IEEE Transactions on Industrial Informatics*, 9, 3, pp. 1680-1687, 2013. http://dx.doi.org/10.1109/TII.2012.2221471

[5] F. Bonin-Font, A. Burguera, A. Ortiz, G. Oliver, "A Monocular Mobile Robot Reactive Navigation Approach Based on the Inverse Perspective Transformation", *ROBOTICA*, 31, pp. 225 – 249, 2013. http://dx.doi.org/10.1017/S0263574712000252

[6] A. de La Bourdonnaye, R. Doskočil, V. Křivánek, "Practical Experience with Distance Measurement Based on Single Visual Camera", *Advances in Military Technology*, 7, 2, 49 – 56, 2012.

[7] http://tdserver1.fnal.gov/darve/mu_cool/pressuretest/Basics_of_Photogrammetry.pdf, (2015).

[8] K. Yue, Z. Li, M. Zhang, S. Chen, "Transient full-field vibration measurement using spectroscopical stereo photogrammetry", *OPTICS EXPRESS*, 18, no. 26, pp. 26866 – 26871, 2010. http://dx.doi.org/10.1364/OE.18.026866

[9] Y. Morimoto, A. Masaya, M. Fujigaki, D. Asai, *Applied Measurement Systems*, chapter 7, 137, ISBN 978-953-51-0103-1, 2012.

[10] R. B. Rusu, A. Aldoma, S. Gedikli, M. Dixon, "3D Point Cloud Processing: PCL", *Tutorial at IEEE/ RSJ International Conference on Intelligent Robots and Systems* (*IROS*), 2011.

[11] A. Saxena, H. Koppula, R. Newcombe, X. Ren, "RGB-D: Advanced Reasoning with Depth Cameras", *Workshop in conjunction with Robotics: Science and Systems* (*RSS*), 2013.

[12] Red.: J. Sarna, R. Kustosz, E. Woźniewska, M. Gonsior, A. Jarosz, K. Szymańska, D. Hansel, E. Krzak, "Program Polskie Sztuczne Serce", ISBN 978-83-63310-16-5, 2013.

[13] T. Pustelny, G. Konieczny, Z. Opilski, M. Gawlikowski, "Measuring systems for pulsatile heart assist pumps ReligaHeart® - measuring system movement of the diaphragm", *Polish artificial heart, the development of design, qualification tests, preclinical and clinical*, ISBN 978-83-63310-12-7, pp. 22 – 36, 2013, (in Polish).

[14] P. Gibinski, G. Konieczny, E. Maciak, Z. Opilski, T. Pustelny, "Acoustic device for measuring instantaneous blood volume in cardiac support chamber i.e. pneumatic heart assist driving chamber, has sensor supporting heart in openings, and audio amplifier connected with volume unit of blood-cell support", *Patent No*. PL394074-A1.

[15] G. Konieczny, T. Pustelny, P. Marczyński, "Quasi-Dynamic Testing of an Optical Sensor for Measurements of the Blood Chamber Volume in the POLVAD Prosthesis", *Acta Physica Polonica A*, 124, 3, pp. 483-485, 2013. http://dx.doi.org/10.12693/APhysPolA.124.483

[16] D. Komorowski, M. Gawlikowski, "Preliminary investigations regarding the blood volume estimation in pneumatically controlled ventricular assist device by pattern recognition", *Computer recognition systems 2*, ASC 45, pp. 558 – 565, 2007. http://dx.doi.org/10.1007/978-3-540-75175-5_70

[17] R. Kustosz, A. Jarosz, M. Gawlikowski, A. Kapis, M. Gonsior, "The role and perspectives of development of the Polish air pump heart assist on the market of heart prosthetic", *Polish artificial heart, the development of design, qualification tests, preclinical and clinical*, ISBN 978-83-63310-12-7, 2013, (in Polish).

[18] K. Murawski, T. Pustelny, M. Murawska, "System and method of determining the shape of diaphragm of pneumatic extracorporeal heart assist pump", *Patent Application No*. P.414104, 2015, (in Polish).

[19] K. Murawski, D. Białas, M. Rękas, "Measurement of Corneal Neovascularisation with the use of Image Processing Techniques", *Acta Physica Polonica A*, vol. 127, 6, pp. 1732 – 1736, 2015. http://dx.doi.org/10.12693/APhysPolA.127.1732

[20] K. Różanowski, K. Murawski, "An Infrared Sensor for Eye Tracking in a Harsh Car Environment", *Acta Physica Polonica A*, 122, 5, pp. 874 – 879, 2012. http://dx.doi.org/10.12693/APhysPolA.122.874

[21] K. Murawski, R. Różycki, P. Murawski, A. Matyja, M. Rekas, An Infrared Sensor for Monitoring Meibomian Gland Dysfunction, Acta Physica Polonica A, 124, 3, 517 – 520, 2013. http://dx.doi.org/10.12693/APhysPolA.124.517

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 823–830
DOI: 10.15439/2016F85
ACSIS, Vol. 8. ISSN 2300-5963

Secret key agreement based on a communication through wireless MIMO fading channels

Victor Yakovlev and Valery Korzhik
State University of Telecommunications
Saint-Petersburg, Russia
Email: viyak@bk.ru, val-korzhik@yandex.ru

Pavel Mylnikov
Research & Design Institute
for Information Technology,
Signaling and Telecommunications
on Railway Transport (JSC NIIAS)
Saint-Petersburg, Russia
Email: paul.mylnikov@gmail.com

Guillermo Morales-Luna
Computer Science
CINVESTAV-IPN
Mexico City, Mexico
Email: gmorales@cs.cinvestav.mx

***Abstract*—The method of key sharing between a mobile unit and a base station through a wireless MIMO-based fading channel is investigated. The description of a key distribution protocol is given. The expression to estimate the correct key bit agreement based on the use of guard interval is proved. Statistical properties of the key string are tested using the NIST criteria. Impossibility of key string eavesdropping by illegal users is guaranteed due to small values of correlation between legal and eavesdropper carrier phases. Numerical examples show that the MIMO system with 8 antennas is able to agree 256 bits with a reliability value 0.99 for SNR equal to 35 dB. This experiment confirms that the MIMO scenario is especially effective for secret key distribution.**

***Index Terms*—MIMO fading channel, key distribution, multi-phase detection, correlation, NIST tests.**

I. Introduction

SECURE transmission is still a concern for wireless mobile devices due to broadcast nature of signals. Although traditional secure systems employ private or public-key cryptography independently of physical transmission, there is a growing interest in *physical layer security* methods that exploit noisy telecommunication channels and channels with multi-path wave propagation.

In a pioneered paper [1], Wyner introduced the *wire-tap channel concept* with two types of channels: the main channel (less noisy) and the wire-tap channel (more noisy). Wyner's theorem states that under some (not very strong conditions) there exist such encoding and decoding procedures that reliable transmission on the main channel and zero information leakage on the wire-tap channel can be provided with an increasing of the block lengths if the transmission rate is less than the so called *secrecy capacity Cs.* In the post-Wyner's period there appear a lot of paper devoted to a generalization of wire-tap channel models [2], [3], [4], [5], [6] and to a specification of the amount of information leaking to eavesdropper [7]. But unfortunately it is still unknown constructive encoding and decoding procedures providing a transmission rate close to secrecy capacity. Next advance in the physical security area occurs due to Maurer's paper [8] at the cost of public discussion between legitimate users. Such approach allows to share secret keys between legitimate users even in the case when the wire tapper observes a "better" channel than one used by the legitimate user but only if the wire tapper is passive (that is in another words if legal channel is authenticated). After a common key sharing the legitimate correspondents can use ideal Shannon's one-time pad cipher [9].

The idea of a common key sharing and the execution of an ideal cipher is very positive especially in the so called *post-quantum* period when it is assumed that many cryptographic algorithms can be broken by a quantum computer [10]. But such approach requires to share a very long secret key string before ideal encryption. Moreover, in order to provide a secure key sharing that means to get a negligible amount of Shannon information leaking to an eavesdropper about this key, it is necessary to be sure that signal-to-noise ratio at the input of wire tapper receiver is fixed and known for legitimate users. In order to avoid such strong requirement it has been proposed to execute (for mobile users) a multipath wave propagation in some wireless channels [11], [12]. Unfortunately if mobile unit stops it may result in a very slow and small channel fluctuation. In order to take for granted some given randomness level it would be better to create this randomness artificially by means of legitimate users. In [13] it has been proposed a method using smart antenna excited randomly by electronic means. More detail investigation of such approach was undertaken in [14]. But such approach requires a special construction of a *Variable Directional Antenna.*

The explosion of interest to multiple-input multiple-output (MIMO) systems soon led to a realisation that exploiting the available spatial dimensions can also enhance the secrecy capabilities of wireless channels [15].

One of advantage of MIMO system for key sharing is its property that a presence of many antennas results in a better randomisation even for very small transfer of mobile units. It is worth to note that in contrast to communication system where a presence of MIMO devices results in interference of signals at the receiving antennas, key sharing occurs avoidable of such interference because in that case it is necessary to form any but only coinciding key bits. The last property is provided thanks to the *Reciprocity Theorem of radio wave propagation* between

978-83-60810-90-3/$25.00

transmitting and receiving sides of MIMO-based link. Further investigation of MIMO-based *key distribution protocols (KDP)* was undertaken by authors of the papers [16], [17]. But a final solution of this problem is very far from a termination. First of all it is requested to increase the key generation rate providing simultaneously high secrecy and good statistical properties of the shared keys that should be close to truly random data. Namely these questions form the main subject of our investigations undertaken in the current paper.

The reminder of the article is organised as follows. Section II describes the model of MIMO channel with point of view key sharing protocol. In Section III algorithm of key distribution is presented jointly with estimation of key bits reliability and key rate generation. Section IV discusses system parameters optimisation. Section V concludes the paper and formulates open problems for further investigations. The appendix presents the proof of the relation for the error probability given in Section III.

II. Mathematical model of MIMO-based channel

We assume that a key distribution protocol (KDP) is performed between a mobile unit A and a base station B that have the same number of antennas N_A and that the signal power radiated of each antenna is equal to P_S/N_A.

For a *frequency-flat fading MIMO channel*, the commonly used discrete-time input-output relation for test-signal is given by

$$y = Hs + z \tag{1}$$

where H is a square $(N_A \times N_A)$-matrix, s is a transmitted test-signal $(N_A \times 1)$-vector, y is a received signal $(N_A \times 1)$-vector, and z is an additive noise $(N_A \times 1)$-vector of the MIMO channel output.

Due to the Reciprocity Theorem the relation for back channel is

$$y' = H^T s + z'.$$

However the elements of the matrix H can change during the test signal transmission, generally speaking, and therefore in order to provide approximated equally channel matrices in direct and back channels it is necessary that the following inequalities hold [18]:

$$\Delta t << T_c \quad , \quad \Delta f << B_c$$

where Δt is the delay in transmission between direct and back test signals, Δf is the frequency (Doppler) shift, T_c is the coherent time and B_c is the coherent band width for the MIMO channel.

In order to specify the values of the matrix H, it is necessary to describe the channel model in detail.

Let us consider the multi-path MIMO channel model with Raleigh fading according to [19] and presented in Fig. 1.

We denote the number of rays as L and denote by β_l the attenuation in the l-th ray, by ϕ_l, ψ_l the transmitted and received angles, respectively, by Φ, Ψ the antenna diagram angles, and by ω_l the frequency shift due to mobile units transfer (Doppler effect).

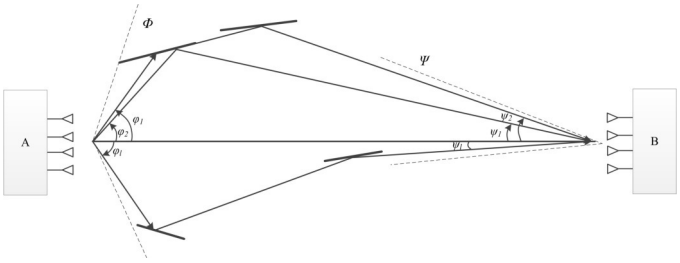


Fig. 1. General model of MIMO channel between mobile unit and base station.

Then the matrix $H(t)$ of the test signal at time t can be presented as

$$H(t) = \sum_{l=1}^{L} \beta_l \left(\mathbf{a}_{Rl}\mathbf{a}_{Tl}^T\right) e^{-j\omega_l t}$$

being

- $\beta_l = a_l e^{j\theta_l}$: signal attenuation resulting by reflection,
- $\mathbf{a}_{Rl}, \mathbf{a}_{Tl}$: response vectors at the receiver and at the transmitter respectively.

$$\begin{aligned} \mathbf{a}_{Rl} &= \left[1 \;\; e^{-j\Omega_{Rl}} \;\; \dots \;\; e^{-j(N_A-1)\Omega_{Rl}}\right]^T, \\ \mathbf{a}_{Tl} &= \left[1 \;\; e^{-j\Omega_{Tl}} \;\; \dots \;\; e^{-j(N_A-1)\Omega_{Tl}}\right]^T, \end{aligned}$$

- $\Omega_{Rl} = \frac{2\pi}{\lambda} d_R \sin\phi_l$: angular receiver frequency,
- $\Omega_{Tl} = \frac{2\pi}{\lambda} d_T \sin\psi_l$: angular transmission frequency,
- λ: wave length corresponding to carrier frequency,
- d_R: diversity interval for receiving antennas, and
- d_T: diversity interval for transmission antennas.

A typical example of the above channel model is the octal-rays model of the railway telecommunication system having 8 antennas with distances $\frac{\lambda}{2}$, between them, mobile object speed 100 km/h, $\Phi = 28°$, $\Psi = 180°$, and carrier frequency 2.6 GHz.

Investigation of such model has been undertaken in many papers (e.g. [20], [21] and others) and results of our investigations show that the entries of matrix H can be correctly approximated by zero mean Gaussian distribution with equal variances.

The space correlation is determined only by mutual antenna locations. Then space-time correlation can be presented following to the results of [22], [18] as

$$R_H(t) = R_H \cdot \rho(t),$$

where R_H is the matrix of space correlation between antennas, and ρ is the time correlation function of antenna location. For Jakes fading model [20], the function ρ can be determined as

$$\forall t: \;\; \rho(t) = J_0(2\pi f_D(t)),$$

where f_D is the Doppler spread and J_0 is the Bessel function of zero order.

III. KDP BASED ON MIMO CHANNEL MODEL

The KPD is described in the following steps:

1) User B (base station) sends the test signal to the mobile unit A executing all antennas.
2) A calculates some parameter of the received signals.
3) Just after step 2, A sends the same test signal to B.
4) B calculates the same (selected in advance by both users) parameters of the received signals.
5) Both users A and B form the key bits from the found parameters using a quantisation procedure.

It is worth to note that the knowledge of MIMO-based channel model parameters can be ignored in KDP design if during the time of its execution these parameters are approximately constant. But this knowledge it is necessary to estimate a reliability of KDP (the probability of key bits coincidence for both correspondents), the statistical properties of key strings and its security (in terms of information leakage about this key to eavesdropper who can be located in some vicinity of users A and B).

We can see from relation (1) that each coordinate y_i of the vector $\mathbf{y}$ is a complex Gaussian random value with amplitude $\mu_i = \sqrt{\Re(y_i)^2 + \Im(y_i)^2}$ (here $\Re$ and $\Im$ are the maps that take the real and the imaginary parts of a complex number) and phase $\theta_i = \tan^{-1}\left(\frac{\Im(y_i)}{\Re(y_i)}\right)$, and these variables have Rayleigh distribution and uniform distribution, respectively. It has been proved in [23] that phases are less correlated versus distance between legal users and eavesdroppers than amplitude. Therefore our selection as parameter for the key bit generation, namely the phase quantisation procedure into q integers, is determined as:

$$\begin{array}{rl} \text{if} & \theta_i \in \left[\frac{2\pi(q-1)}{Q}, \frac{2\pi q}{Q}\right) \text{ with } 1 \le q \le Q \\ \text{then} & f_Q(\theta_i) = q, \end{array} \tag{2}$$

where Q is the number of quantisation levels. Then the probability of an integer q equals $\frac{1}{Q}$. Since the channel noise results in a transition of q to $q' \neq q$ and it is more likely closer to the bounds of the decision areas in (2), we propose to introduce guard intervals between decision areas as key symbols may be erased. Then the decision function (2) can be modified as:

$$\begin{array}{rl} \text{if} & \theta_i \in \left((q-1)\Omega - \frac{\gamma}{2}, (q-1)\Omega + \frac{\gamma}{2}\right) \cup \\ & \left(q\Omega - \frac{\gamma}{2}, q\Omega + \frac{\gamma}{2}\right) \\ \text{then} & f_{QG}(\theta_i) = \textit{erasure}; \end{array}$$

$$\begin{array}{rl} \text{if} & \theta_i \in \left[(q-1)\Omega + \frac{\gamma}{2}, q\Omega - \frac{\gamma}{2}\right) \\ \text{then} & f_{QG}(\theta_i) = q; \end{array}$$

with $1 \le q \le Q$, $\Omega = \frac{2\pi}{Q}$ and γ a threshold, $\gamma \in \left[0, \frac{1}{2}\Omega\right)$.

Let us consider one of the decision areas (or *sector*) in Fig. 2, determined by $\mathbf{y}_i = (\mu_i, \theta_i)$, $\mathbf{x}_i = Hs = (a_i, \phi_i)$, $\mathbf{z}_i = (b_i, \psi_i)$. Under the decision taken about the phase ϕ_i when $\mathbf{y}_i$ is received, the following events may occur:

- $\mathbf{y}_i$ is in the same area that the vector $\mathbf{x}_i$ (correct decision area with angle $\Omega - \gamma$),

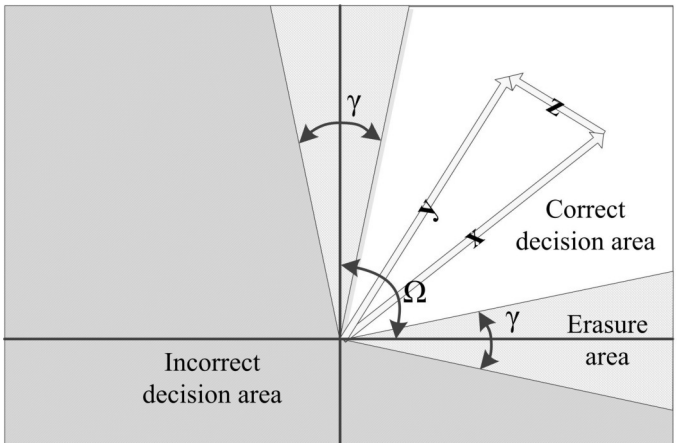


Fig. 2. Sectors of decision area.

- $\mathbf{y}_i$ belongs to the guard interval (erasure area with angle γ),
- $\mathbf{y}_i$ appears outside of both previous areas (incorrect decision).

Let us denote the probabilities of previous events by P_{cor}, P_{er}, P_{error}, respectively.

It is proved in the Appendix that

$$\begin{aligned} P_{error} &= \frac{1}{2\pi\Omega}\int_0^{\Omega} F(\phi)\, d\phi \\ F(\phi) &= \int_{\phi+\frac{\gamma}{2}}^{2\pi-(\phi+\frac{\gamma}{2})} \left[1 + \left[\frac{h\tan\left(\phi+\frac{\gamma}{2}\right)}{\sin\psi}\right]^2\right]^{-1} d\psi \\ P_{er} &= 1 - P_{cor} - P_{error} \end{aligned} \tag{3}$$

where P_{cor} is given by eq. (16) in the Appendix after combining (11)–(15).

In Fig. 3 there are plotted the dependences of P_{cor}, P_{er}, P_{error} with respect to signal to noise ratio for $Q = 8$ and different *guard intervals* (GI) that were calculated after numerical computations of corresponding integrals. We observe that it is possible to trade off P_{error} to the value of the guard interval but it affects also on P_{er}. Hence there appears the problem of KDP parameters optimisation, given some final requirements, as key generation rate maximisation for given SNR and the number of antennas N_A in MIMO system. (We remark that it is not obtained a precise expression for the corresponding probabilities but some bounds, namely an upper bound for P_{error}, a lower bound for P_{er} and lower bound for P_{cor} because it was not taken into account that correct key bits can be obtained sometimes even if both legal users got incorrect phase. But such incorrectness is acceptable).

From Fig. 3, it can be seen that a guard interval (GI) allows to decrease the error probability but simultaneously the probability of erasure increases. In reality a final decision about key bit has to be taken no by the single user B but by

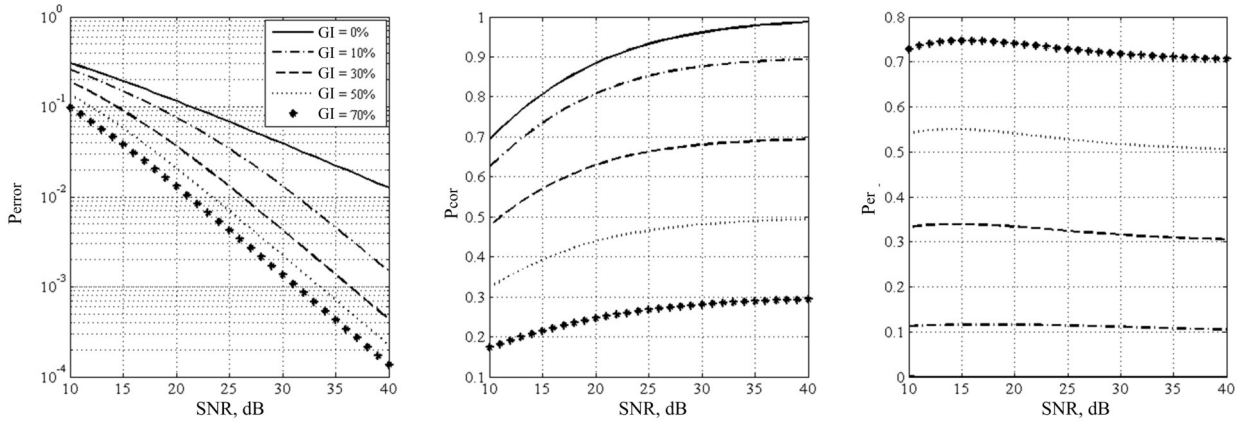


Fig. 3. Curves of P_{error}, P_{cor}, P_{er} against the values of SNR for $Q = 8$.

both users A and B. Thus the final probabilities P_{cor}, P_{er}, P_{error} should be changed as follows:

$$\begin{aligned} P^*_{cor} &= P_{cor}(A) \cdot P_{cor}(B) \\ P^*_{error} &= P_{error}(A) \cdot P_{cor}(B) + \\ &\quad P_{error}(B) \cdot P_{cor}(A) + \\ &\quad P_{error}(A) \cdot P_{error}(B) \\ P^*_{er} &= 1 - P^*_{cor} - P^*_{error} \end{aligned}$$

The KDP should be also slightly corrected under the condition of symbol erasures. Namely the numbers of the erased symbols have to be transmitted from both users to opposite ones in addition and next it is transmitted extra signal if it is necessary.

A relation for the probability $P_{n_0}(k)$ of the key bit string sharing of length n_0 is:

$$P_{n_0}(k) = \left(\frac{P^*_{cor}}{1 - P^*_{er}} \right)^{\frac{n_0}{\log Q}} \tag{4}$$

where Q is the number of quantisation levels. The key bit stream rate for the use of all N_A antennas is

$$R = N_A \log_2 Q (1 - P^*_{er}) \frac{bit}{sample}. \tag{5}$$

Then the following optimisation problem arises:

$$(\gamma^*, Q^*, N^*_A, (h^2)^*) = \underset{\gamma, Q, N_A, h^2}{\arg\max}\ R \tag{6}$$

subject to the restrictions

$$\begin{gathered} P_{n_0}(k) \geq P_{n_0}(k)_{\text{requested}}; \\ \gamma \in \left[0, \frac{\pi}{Q}\right); \\ Q \in [2, Q_{\max}]; \\ N_A \in [1, (N_A)_{\max}]; \\ h^2 \in [1, (h^2)_{\max}]; \end{gathered}$$

where the values $P_{n_0}(k)_{\text{requested}}$, $Q_{\max}$, $(N_A)_{\max}$, $(h^2)_{\max}$ have to be conditioned by the general requirements of the MIMO system design.

The solution of problem (6) has been found by the *branch and bound algorithm* [24].

In Fig. 4 there are presented the dependences R from SNR under the conditions $n_0 = 256$, $P_{n_0}(k)_{\text{requested}} = 0.9$ and 0.99, and $N_A = 1, 2, 4, 8, 16$.

In Table I the optimal values for $a = \frac{\gamma}{\Omega}$ and Q are displayed maximising the rate R for a given SNR.

An analysis of the curves in Fig. 4 shows that the key generation rate R increases with an increasing of the number of antennas in MIMO massive. Key generation rate increases obviously as SNR increases. For every SNR value there exist optimal number of phase quantisation levels and value of guard interval providing the requested probability of correct key sharing for both legal users.

IV. Investigation of Key Stream Statistic and Information Leaking to Eavesdropper

The statistics of the key stream distributed due to KDP is very important because if it is very far from uniform distribution it may result in effective attacks for cipher breaking. In order to investigate the key stream statistic after phase quantisation from different antennas they will be combined in a serial sequence containing bits from all antennas and this sequence investigated by statistical tests. In Fig. 5 there are presented the empirical density distributions for the length of binary strings equal to 1 and 16.

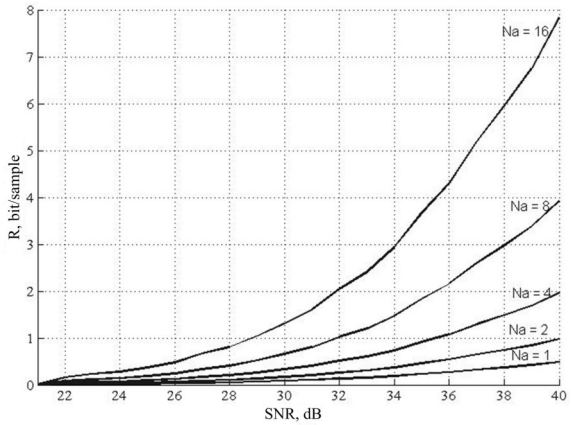
We see there that also a balance of zeros and ones (as follows from Fig. 5 a) is good, but the multi-variate distribution (Fig. 5 b) has anomalous peaks. In order to improve the statistics of the key string it was undertaken some deterministic transforms of two types recommended in [25]. The first type is so called *transposition of symbols* and the second transform is *adjacent bit XOR-ing*. The results of testing after such transforms are presented in Table II in which were used some NIST STS tests [26].

We see that after both transforms the key bit sequence passes the most of NIST tests.

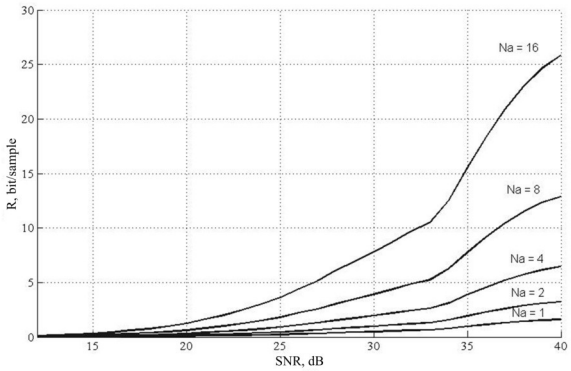
Now let us face with eavesdropping problem and assume that the following parameters hold [21]:

TABLE I
OPTIMAL PARAMETERS γ AND Q PROVIDING THE MAXIMUM RATE R FOR A GIVEN SNR.

$P_{n_0}(k)_{\text{requested}}$		SNR (dB)															
		10	12	14	16	18	20	22	24	26	28	30	32	34	36	38	40
0.9	Q^*	–	2	2	2	2	2	2	2	2	2	2	2	4	4	4	4
	a^*	–	0.89	0.86	0.82	0.77	0.71	0.64	0.56	0.47	0.38	0.30	0.22	0.37	0.24	0.15	0.10
0.99	Q^*	–	–	–	–	–	–	2	2	2	2	2	2	2	2	2	2
	a^*	–	–	–	–	–	–	0.89	0.86	0.82	0.77	0.71	0.64	0.57	0.48	0.39	0.30

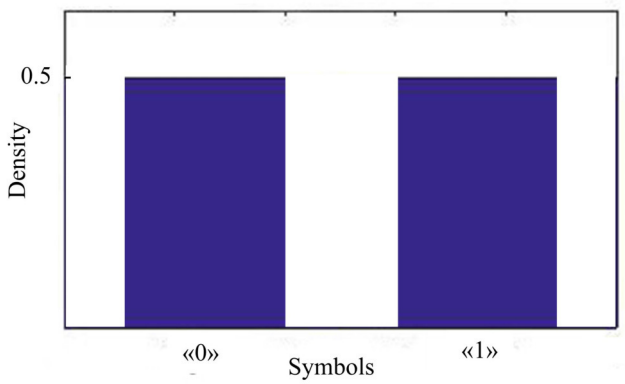


a) $P_{n_0}(k)_{\text{requested}} = 0.9$

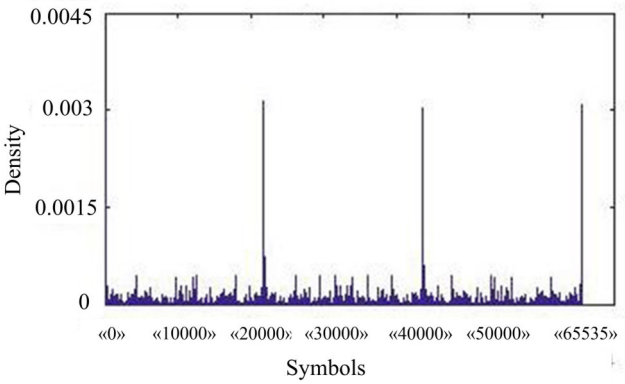


b) $P_{n_0}(k)_{\text{requested}} = 0.99$

Fig. 4. The key sharing bit rate of the length 256 bits with $P_{n_0}(k)_{\text{requested}} \in \{0.9, 0.99\}$ for optimisation of the system parameters.



a) String length 1



b) String length 16

Fig. 5. Empirical probability density distributions for two string lengths.

- frequency carrier: 2600 MHz;
- MIMO massive: 8×8;
- distance between MIMO antennas at the departure unit: 0.5λ
- distance between MIMO antennas at the arrival unit: 0.5λ
- number of rays: 8;
- departure ray angle: 28°;
- arrive ray angle: 180°;
- speed of mobile unit motion: 50 km/h;
- number of simulated channel matrices: 1000.

The mutual correlation between the phases of legal user B and an eavesdropper located at a distance d from B (in terms of wave length factors) is presented in Fig. 6.

We see from this figure that in line with similar results presented in [14] the correlation has not a monotonically decreasing dependence from d but it has a randomly-looking dependence. But in contrast to [14] it has significantly less values from all distances between $(0.1\lambda, \ldots, 20\lambda)$. This is a consequence of the multi-phase functional used for key bit generation and another channel model. Thus, we can believe that it can be neglected an opportunity of key eavesdropping

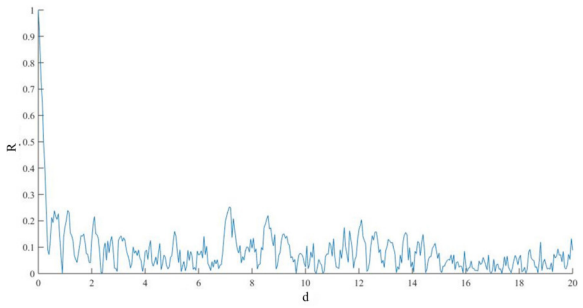


Fig. 6. Mutual correlation between phases of legal user and eavesdropper against distance between them in terms of wave length factors.

TABLE II
EXPERIMENTAL TESTING BASED ON NIST STS OF THE KEY SEQUENCES AFTER TWO TYPES OF TRANSFORMS.

Nr.	Name of test	Transposition	TXOR
1	The Frequency (Monobit) Test	10/10	10/10
2	Frequency Test within a Block	10/10	10/10
3	The Runs Test	9/10	10/10
4	Tests for the Longest-Run-of-Ones in a Block,	0/10	10/10
5	The Binary Matrix Rank Test	10/10	10/10
6	The Discrete Fourier Transform (Spectral) Test	0/10	9/10
7	The Non-overlapping Template Matching Test	1/10	8/10
8	The Overlapping Template Matching Test	0/10	10/10
9	Maurer's "Universal Statistical" Test	10/10	10/10
10	The Linear Complexity Test	5/5	7/7
11	The Serial Test	5/5	7/7
12	The Approximate Entropy Test	0/10	10/10
13	The Cumulative Sums (Cusums) Test	3/10	10/10
14	The Random Excursions Test	1/10	10/10
15	The Random Excursions Variant Test	10/10	10/10

TXOR: Transposition plus XOR of adjacent bits.
The numerators of fractions are number of "passed" tests and the denominators are the total number of tests.

in large area of eavesdropper locations.

(It is worth to note that if phase had Gaussian distribution and even for binary quantisation values it would be results in the error probability for eavesdropper about one key bit near 0.47 [14] that it is very close to "break of eavesdropper channel".)

V. CONCLUSION

We considered a method of key sharing for wireless secret communication based on MIMO concept with the use of multi-phase functionals that seems to be especial effective for mobile unit and multi-path fading channels.

It has been proved that following to the proposed key distribution protocol it can be provided a key sharing of size 256 bits and with probability of its reliable performance about 0.99 for SNR equal to 35 dB, and 16 antennas after execution of about 74 test signals on average. It was also shown that the key sequence after simple transforms is very close to i.i.d. practically satisfying all NIST tests. Interception of key stream by eavesdropper is prevented by a very small correlation between phases at legal users and eavesdropper if distance between them is not lesser than 0.1λ.

We believe that a future work that can be undertaken is in the use of error correcting codes in order to maximise the key distribution rate and to short a delay in key delivering.

REFERENCES

[1] A. Wyner, "Wire-tap channel concept," *Bell System Technical Journal*, vol. 54, pp. 1355–1387, 1975.
[2] A. B. Carleial and M. E. Hellman, "A note on Wyner's wiretap channel (corresp.)," *IEEE Trans. Information Theory*, vol. 23, no. 3, pp. 387–390, 1977. [Online]. Available: http://dx.doi.org/10.1109/TIT.1977.1055721
[3] S. K. Leung-Yan-Cheong and M. E. Hellman, "The Gaussian wire-tap channel," *IEEE Trans. Information Theory*, vol. 24, no. 4, pp. 451–456, 1978. [Online]. Available: http://dx.doi.org/10.1109/TIT.1978.1055917
[4] I. Csiszár and J. Körner, "Broadcast channel with confidential messages." *IEEE Transactions on Information Theory*, vol. 24, no. 2, pp. 339–348, 1978.
[5] L. H. Ozarow and A. D. Wyner, "Wire-tap channel II," 1985, pp. 33–50.
[6] J. Barros and M. R. D. Rodrigues, "Secrecy capacity of wireless channels," in *IEEE International Symposium on Information Theory*. IEEE, 2006, pp. 356–360.
[7] V. I. Korzhik and V. Yakovlev, "Nonasymptotic estimation for efficiency of code jamming for the wire-tape channel concept (In Russian)," *IEEE Transactions on Information Theory*, vol. 17, no. 4, pp. 223–228, 1981.
[8] U. Maurer, "Secret key agreement by public discussion from common information." *IEEE Transactions on Information Theory*, vol. 39, no. 3, pp. 733–742, 1993.
[9] C. E. Shannon, "Communication theory of secrecy systems," *Bell Systems Technical Journal*, vol. 28, no. 4, pp. 656–715, 1949.
[10] D. Micciancio and O. Regev, "Lattice-based cryptography," in *Post-quantum Cryptography*, D. J. Bernstein and J. Buchmann, Eds. Springer, 2008.
[11] A. M. Sayeed and A. Perrig, "Secure wireless communications: Secret keys through multipath," in *ICASSP*, 2008, pp. 3013–3016.
[12] Y. Liu, S. C. Draper, and A. M. Sayeed, "Secret key generation through ofdm multipath channel," in *CISS*, 2011, pp. 1–6.
[13] T. Aono, K. Higuchi, T. Ohira, B. Komiyama, and H. Sasaoka, "Wireless secret key generation exploiting reactance-domain scalar response of multipath fading channels," *IEEE Transactions on Antennas and Propagation*, vol. 53, no. 11, pp. 3776–3784, 2005.
[14] V. Yakovlev, V. I. Korzhik, Y. Kovajkin, and G. Morales-Luna, "Secret key agreement over multipath channels exploiting a variable-directional antenna," *Int. Jour. Adv. Computer Science & Applications*, vol. 3, no. 1, pp. 172–178, 2012.
[15] J. W. Wallace and R. K. Sharma, "Automatic secret keys from reciprocal MIMO wireless channels: measurement and analysis." *IEEE Trans. Information Forensics and Security*, vol. 5, no. 3, pp. 381–392, 2010. [Online]. Available: http://dblp.uni-trier.de/db/journals/tifs/tifs5.html#WallaceS10
[16] Z. Li, W. Trappe, and R. Yates, "Secret communication via multi-antenna transmission," in *Information Sciences and Systems, 2007. CISS '07. 41st Annual Conference on*, March 2007, pp. 905–910.
[17] S. Shafiee and S. Ulukus, "Achievable rates in Gaussian MISO channels with secrecy constraints," in *International Symposium on Information Theory, 2007. ISIT 07.*, June 2007.
[18] E. Biglieri, R. Calderbank, A. Constantinides, A. Goldsmith, A. Paulraj, and H. V. Poor, *MIMO Wireless Communications*. New York, NY, USA: Cambridge University Press, 2007.
[19] H. Yigit and A. Kavak, "Analytical derivation of 2×2 MIMO channel capacity in terms of multipath angle spread and signal strength," *Frequenz*, vol. 66, no. 1, pp. 97–100, 2012.
[20] W. C. Jakes and D. C. Cox, *Microwave Mobile Communications*. Wiley-IEEE Press, 1994.

[21] K. Guan, Z. Zhong, and B. Ai, "Assessment of LTE-R using high speed railway channel model." in *CMC*, D. Yuan, M. Cao, C.-X. Wang, and H. Huang, Eds. IEEE Computer Society, 2011, pp. 461–464. [Online]. Available: http://dblp.uni-trier.de/db/conf/ieeecmc/ieeecmc2011.html#GuanZA11

[22] M. Bakulin, L. Varukina, and V. Krejdelin, *Tehnologija MIMO: principy i algoritmy*. Gorjachaja linija–Telekom, 2014.

[23] V. Yakovlev, V. Korzhik, and Y. Kovajkin, "Key sharing protocol for wireless local area networks based on the use of randomly excited antenna with variable diagram under the condition of multipath wave propagation. part 1. channel model for key sharing based on the use of smart antenna," in *Problemy informacionnoi bezopasnosti. Komp'juternye sistemy, SPb.: SPbGTU*, June 2011.

[24] A. H. Land and A. G. Doig, "An automatic method of solving discrete programming problems," *Econometrica*, vol. 28, no. 3, pp. 497–520, 1960. [Online]. Available: http://jmvidal.cse.sc.edu/library/land60a.pdf

[25] B. Schneier, *Applied Cryptography (2Nd Ed.): Protocols, Algorithms, and Source Code in C*. New York, NY, USA: John Wiley & Sons, Inc., 1995.

[26] L. E. Bassham, III, A. L. Rukhin, J. Soto, J. R. Nechvatal, M. E. Smid, E. B. Barker, S. D. Leigh, M. Levenson, M. Vangel, D. L. Banks, N. A. Heckert, J. F. Dray, and S. Vo, "Sp 800-22 rev. 1a. a statistical test suite for random and pseudorandom number generators for cryptographic applications," Gaithersburg, MD, USA, Tech. Rep., 2010.

[27] I. S. Gradshteyn and I. M. Ryzhik, *Table of integrals, series, and products*, 7th ed. Elsevier/Academic Press, Amsterdam, 2007, translated from the Russian, Translation edited and with a preface by Alan Jeffrey and Daniel Zwillinger, With one CD-ROM (Windows, Macintosh and UNIX).

APPENDIX

Proof of the formula (3)

Let us consider one of the decision areas, namely $\angle\text{TOL} = \Omega$ (see Fig. 7). Assume that the vector $\mathbf{y}$ is in the sector $\angle\text{SOX}$ (area D_1') and that it is a sum of the vectors $\mathbf{x}$ and $\mathbf{z}$. Then the error for taking a decision on the phase of $\mathbf{y}$ occurs if $\mathbf{y}$ lays on the axis OS or at left to it. This event will occur if and only if:

$$\angle\text{BOA} \geq \angle\text{SOA} = \Omega - \phi + \frac{\gamma}{2}. \tag{7}$$

Let us draw the perpendicular from the point B on the axis OX. Then $\text{BC} = b\sin(\pi - \psi) = b\sin\psi$, $\text{AC} = -b\cos\psi$. We have

$$\angle\text{BOA} = \tan^{-1}\left(\frac{\text{BC}}{a - \text{AC}}\right) = \tan^{-1}\left(\frac{b\sin\psi}{a + b\cos\psi}\right). \tag{8}$$

where a is the amplitude of vector $\mathbf{x}$, and b is the amplitude of vector $\mathbf{z}$. By substituting (8) into (7) we get

$$\frac{b\sin\psi}{a + b\cos\psi} \geq \tan\left(\Omega - \phi + \frac{\gamma}{2}\right).$$

If $a >> b$ (which is very likely) then the term $b\cos\psi$ can be neglected and it results in the following condition to produce error:

$$u := \frac{b}{a} \geq \frac{\tan\left(\Omega - \phi + \frac{\gamma}{2}\right)}{\sin\psi} =: \ell_{\Omega-\phi,\psi}.$$

Let us denote by $P(u)$ the probability density of the random variable u. Then the error probability (P_{error}), provided that the received vector $\mathbf{y}$ lies at the left of OS, can be expressed by the following formula, on the assumption that phases ϕ and ψ are distributed uniformly:

$$P^{I}_{error} = \frac{1}{\Omega}\int_0^{\Omega} d\phi\,\frac{1}{2\pi}\int_{\Omega-\phi+\frac{\gamma}{2}}^{\pi} d\psi \int_{\ell_{\Omega-\phi,\psi}}^{+\infty} f(u)\,du \tag{9}$$

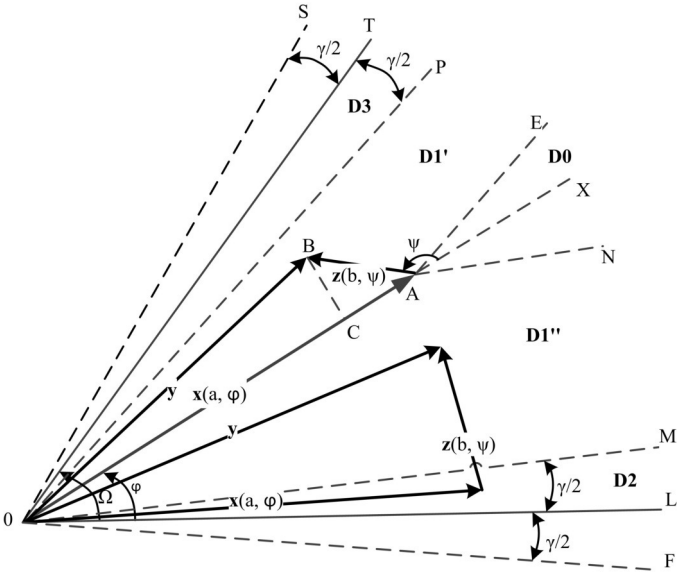


Fig. 7. Areas for taken of decision after quantisation.

(the superindex I emphasises that it is true whenever the received vector occurs to the left of line OS).

For the area $\text{D1}'' = \angle\text{LOX}$, we can repeat the derivation of (9) in order to get

$$P^{II}_{error} = \frac{1}{\Omega}\int_0^{\Omega} d\phi\,\frac{1}{2\pi}\int_{\pi}^{2\pi-\phi-\frac{\gamma}{2}} d\psi \int_{\ell_{\phi,\psi}}^{+\infty} f(u)\,du \tag{10}$$

Let us specify the formula for the probability of correct decision after quantisation and introducing of guard interval. We can see such areas at Fig. 7. Having received the vector $\mathbf{y}$ we get the following cases:

1) $\mathbf{x} \in \text{D1}, \mathbf{y} \in \text{D1}$, with $\text{D1} = \text{D1}' \cup \text{D1}''$. After repeating the procedure to obtain (9) and (10), we get

$$P_{cor\,\text{D1}} = \frac{1}{\Omega}\int_{\frac{\gamma}{2}}^{\Omega-\frac{\gamma}{2}} d\phi\,\frac{1}{2\pi}\int_{\phi-\frac{\gamma}{2}}^{2\pi-\phi+\frac{\gamma}{2}} d\psi \cdot \int_0^{\ell_{\phi,-\psi}} f(u)\,du \tag{11}$$

2) $\mathbf{x} \in \text{D2} \cup \text{D3}, \mathbf{y} \in \text{D1}$, with $\text{D1} = \text{D1}' \cup \text{D1}''$. After repeating the procedure to obtain (9) and (10), we get

$$P_{cor\,\text{D2}\cup\text{D3}} = \frac{1}{\Omega}\int_{\Omega-\frac{\gamma}{2}}^{\frac{\gamma}{2}} d\phi\,\frac{1}{2\pi}\int_{\phi-\frac{\gamma}{2}}^{2\pi-\phi+\frac{\gamma}{2}} d\psi \cdot \int_{\ell_{-(\Omega-\phi),\psi}}^{\ell_{\phi,-\psi}} f(u)\,du \tag{12}$$

3) $\mathbf{x} \in \text{D1}, \mathbf{y} \in \text{D0}$. In such situation the vector $\mathbf{x}$ transfers to the area of correct decision from the area of erasure due to noise.

$$P_{cor\,\text{D0}} = \frac{1}{\Omega}\int_{\frac{\gamma}{2}}^{\Omega-\frac{\gamma}{2}} d\phi\,\frac{1}{2\pi}\int_0^{\Omega-\gamma} d\psi \int_0^{+\infty} f(u)\,du = \frac{(\Omega-\gamma)^2}{2\pi\Omega} \tag{13}$$

4) $\mathbf{x} \in$ D2, $\mathbf{y} \in$ D0 (in this case, D2 and D0 are intersected)

$$P_{cor\,\mathrm{D2}\to\mathrm{D0}} = \frac{1}{\Omega}\int_0^{\frac{\gamma}{2}} d\phi\,\frac{1}{2\pi}\int_{\frac{\gamma}{2}}^{\Omega-\frac{\gamma}{2}} d\psi\,\cdot \int_{\ell_{-\phi,\psi}}^{+\infty} f(u)\,du \tag{14}$$

5) $\mathbf{x} \in$ D3, $\mathbf{y} \in$ D0. It is easy to see that

$$P_{cor\,\mathrm{D3}\to\mathrm{D0}} = P_{cor\,\mathrm{D2}\to\mathrm{D0}} \tag{15}$$

Combining (11)–(15) we get:

$$P_{cor} = P_{cor\,\mathrm{D1}} + P_{cor\,\mathrm{D2}\cup\mathrm{D3}} + P_{cor\,\mathrm{D0}} + 2\,P_{cor\,\mathrm{D2}\to\mathrm{D0}} \tag{16}$$

It is very easy to see that

$$P_{erasure} = 1 - P_{cor} - P_{error}.$$

In order to prove the relations (9)–(14) in a closed form it is necessary to derive the probability density function of the random variable $u = \frac{b}{a}$, where a and b have the Rayleigh distribution and they are mutually independent. This means that

$$f(a) = \begin{cases} \frac{a}{\sigma_a^2} e^{-\frac{a^2}{2\sigma_a^2}} & \text{if } a \geq 0 \\ 0 & \text{if } a < 0 \end{cases}$$

$$f(b) = \begin{cases} \frac{b}{\sigma_b^2} e^{-\frac{b^2}{2\sigma_b^2}} & \text{if } b \geq 0 \\ 0 & \text{if } b < 0 \end{cases}$$

After a simple transform, we get the following relation

$$f(u) = \frac{u}{\sigma_a^2\sigma_b^2}\int_0^{+\infty} a^3 e^{-ra^2}\,da, \tag{17}$$

where $r = \frac{1}{2\sigma_a^2} + \frac{u^2}{2\sigma_b^2}$.

The integral (17) can be expressed in a closed form [27]. Then for the definite integral (17) we get

$$f(u) = \frac{2\sigma_a^2\sigma_b^2\,u}{(u\sigma_a^2 + \sigma_b^2)^2}. \tag{18}$$

By denoting $\delta^2 = \frac{\sigma_b^2}{\sigma_a^2}$ then we get from (18)

$$f(u) = \frac{2\delta^2\,u}{(\delta^2 + u^2)^2}. \tag{19}$$

Substituting (19) into (9) and (10) we obtain

$$P_{error} = \frac{1}{2\pi\Omega}\int_0^{\Omega} d\phi\int_{\phi+\frac{\gamma}{2}}^{2\pi-\phi-\frac{\gamma}{2}} d\psi\,\cdot \int_{\ell_{\phi,\psi}}^{+\infty} \frac{2\delta^2\,u}{(\delta^2+u^2)^2}\,du \tag{20}$$

Last integral in (20) can be expressed in a closed form:

$$\int \frac{2\delta^2\,u}{(\delta^2+u^2)^2}\,du = -\frac{\delta^2\,u}{\delta^2+u^2}.$$

Then we get for the definite integral

$$\int_V^W \frac{2\delta^2\,u}{(\delta^2+u^2)^2}\,du = \left[\frac{1}{1+(hV)^2} - \frac{1}{1+(hW)^2}\right] \tag{21}$$

where $h = \frac{1}{\delta}$ is the signal-to-noise ratio.

Substituting (21) into (20) we get finally

$$P_{error} = \frac{1}{2\pi\Omega}\int_0^{\Omega} d\phi\,\cdot \int_{\phi+\frac{\gamma}{2}}^{2\pi-\phi-\frac{\gamma}{2}} \left[1 + \left[\frac{h\tan\left(\phi+\frac{\gamma}{2}\right)}{\sin\psi}\right]^2\right]^{-1} d\psi$$

Q.E.D

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 831–835
DOI: 10.15439/2016F231
ACSIS, Vol. 8. ISSN 2300-5963

Evaluation of an Optimized K-Means Algorithm Based on Real Data

Cosmin M. Poteraş
Faculty of Automation, Computers and Electronics
University of Craiova, Romania
Email: cpoteras@software.ucv.ro

Mihai L. Mocanu
Faculty of Automation, Computers and Electronics
University of Craiova, Romania
Email: mmocanu@software.ucv.ro

***Abstract*—In a previous paper [1] we introduced an optimized version of the K-Means Algorithm. Unlike the standard version of the K-Means algorithm that iteratively traverses the entire data set in order to decide to which cluster the data items belong, the proposed optimization relies on the observation that after performing only a few iterations the centroids get very close to their final position causing only a few of the data items to switch their cluster. Therefore, after a small number of iterations, most of the processing time is wasted on checking items that have reached their final cluster. At each iteration, the data items that might switch the cluster due to centroids' deviation will be re-checked. The prototype implementation has been evaluated using data generated based on an uniform distribution random numbers generator. The evaluation showed up to 70% reduction of the running time. This paper will evaluate the optimized K-Means against real data sets from different domains.**

I. Introduction

THE MORE data continues to grow in both quantity (volume) and diversity, the more challenging clustering algorithms become. Clustering refers to identifying data items' common characteristics (features or attributes) and grouping the data items according to a quantitative estimation of these characteristics. The resulted groups are usually called clusters and they have to be strongly differentiated by their underlying characteristics.

A wide range of domains have successfully employed clustering. Paper [2] made use of clustering for analysing markets as well as recommendations. Papers [3] [4] apply clustering in medicine. Paper [5] uses clustering to analyse news articles and their comments for e-business related purposes. Paper [6] employs clustering for predicting students academic results, while paper [7] applies clustering to human activity recognition.

Unless a mathematical model is available, choosing the most suitable clustering algorithm might prove a hard decision. Arguments that might lead the decision can range from complex experimental results to our own intuition.

Among other challenging open clustering-related issues like: heterogeneity, volume or scalability, that are worth putting research efforts into, raised by clustering, the execution time plays a very important role.

Our proposed optimization focuses on improving the execution time of the K-Means algorithm while keeping the same output.

In use for more than four decades, the K-Means algorithm has been applied in a wide area of fields, ranging from artificial intelligence to image processing or from neural networks to machine vision, or more specifically in unsupervised learning, pattern recognition, classification analysis a.o.

The K-Means algorithm uses a set of cluster centers (cluster centroids) and distributes the data items to the cluster with the closest centroid in terms of Euclidean distance. Picking up the best initial centroids is still an open issue. Different centroids lead to different output and has an important influence on the performance of the algorithm. Choosing the right centroids is beyond the scope of this paper.

The standard K-Means algorithm implies successive exploration of the entire data space with the goal of distributing data items to clusters. At the end of every iteration, the centroids are re-computed by averaging the data items inside the cluster. The next iteration will make use of the newly computed centroids and re-distribute the data items. The loop continues until the centroids no longer change or until a maximum number of iterations has been reached.

The optimization introduced in [1] and presented also in this paper is based on an easily noticeable fact: after a small number of iterations, most of the data items no longer change their cluster, and at the same time, the centroids' deviation reduces significantly. So, why exploring the entire data space if only a small number of data items are subjects to changing the cluster?

Our solution aims of drawing a line between the data items that will certainly not change their cluster, avoiding their exploration in the next iteration(s), and the data items that might switch their cluster which obviously are to be checked.

Exploring the data space only partially on every iteration will not affect the centroid computation. The influence of the data items that are not subject to changing the cluster, on the future centroids, will be preserved in the next iteration(s).

That being said, it becomes obvious that the optimization does not affect the output in any way.

The paper's structure is as follows: section II presents previous attempts for reducing the execution time of the K-Means algorithm, section III describes our proposed optimization for the K-Means algorithm, section IV experimentally evaluates the algorithm against real data sets, (unlike paper [1] where the evaluation is performed against a randomly generated data

978-83-60810-90-3/$25.00

set with uniform distribution), while section V concludes the paper.

II. Related Work

The K-Means algorithm was subject to many research studies covering a wide range of optimization approaches, from computational complexity reduction to parallel and distributed implementations.

In paper [8] the authors propose an optimization that relies on the assumption that if a data item got closer to the centroid on the previous iteration, it will not change the cluster. The assumption allowed the implementation to reduce the amount of computations necessary for computing new centroids.

Parallel and distributed solutions have been discussed in [9][10][11] by treating important topics specific to this kind of environments: synchronization, communication overhead, data availability, architecture (peer-to-peer, client-server, a. o.). The parallel and distributed implementations showed considerable improvement when dealing with very big data sets.

GPUs proved to be a good host for highly-parallel implementations of the K-Means algorithm. Such platforms are addressed in papers [12][13]

III. Optimizing K-Means

In this section we will introduce both the standard K-Means algorithm and the optimized version of it, as proposed in paper [1]. The same optimization strategy will be discussed here, for a better understanding and reading experience.

Algorithm 1 presents the main phases of the standard K-Means algorithm.

Algorithm 1 Standard K-Means

1. Load initial centroids
2. Visit all data items and distribute them to the cluster with the closest centroid
3. For each cluster compute the average of all data items and set the result as the cluster's centroid
4. If the exit criteria (no centroid changes or the maximum number of iteration is reached) are not met, go to step 2
5. Exit.

At a closer look, we can immediately identify step number 2 as the one requiring the most execution time. The time spent on step 2 increases proportionally with the size of the data set, as the entire data set is explored at every iteration.

Figure 1 illustrates an example of centroids evolution of a standard K-Means algorithm. Data items are represented as 2D points. Centroids A, B and C start from their initial positions A_1, B_1, C_1, and successively traverse positions A_i, B_i, C_i, where $i = 1..6$. A_6, B_6 and C_6 are the final positions of the centroids. One can easily notice that after only few iterations, the centroids get very close to their final position, which means that after a few iterations, the number of data items that are subject to changing the cluster reduces considerably.

This observation plays a key role in improving step 2 of algorithm 1; it states that after a small number of iterations,

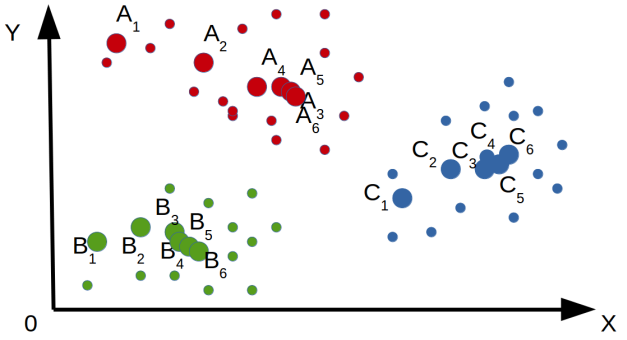


Fig. 1. Centroids Deviation

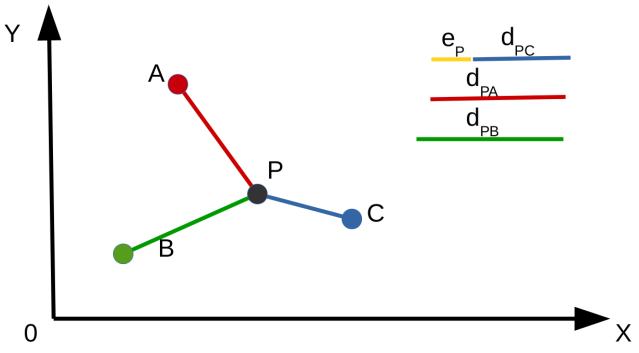


Fig. 2. Point P in an Arbitrary Iteration

the number of points that must be visited (that might change the cluster) is reduced considerably. This lead us to defining specific criteria for splitting the data set in two collections: the former would be made of all data items that are subject to changing the cluster (let's call that the border collection) and the latter would be made of all other points (not changing the cluster). Before providing the mathematical criteria behind the two data item collections, let's examine figure 2.

Figure 2 is a snapshot of an arbitrary iteration around an arbitrary point P. Point P is part of cluster C as a consequence of the fact that the distance from P to C (d_{PC}) is less than the distance to A (d_{PA}) and the distance to B (d_{PB}). We are interested in evaluating the "safety distance" of point P, that is, the distance that P is missing from jumping to the next closest cluster (the cluster represented by centroid A). Let's call this distance e_P (distance to the cluster's edge). We can state that P is e_P-away from the next closest cluster.

$$e_P = min(d_{PA} - d_{PC}, d_{PB} - d_{PC}) \quad (1)$$

At the end of the iteration, centroids A, B and C would be re-computed causing them to jump to new positions A', B', C'. The worst scenario for P is the following: A moves closer to P by $|AA'|$, B also moves closer by $|BB'|$ while C moves away from P by $|CC'|$. The conditions that e_P must fulfil in order for P to remain in cluster C is:

$$e_P > |CC| + |AA| \quad (2)$$

and

$$e_P > |CC| + |BB| \quad (3)$$

For the sake of computation reduction, we can merge conditions 2 and 3 into the following condition:

$$e_P > 2 * max(|AA|, |BB|, |CC|) \quad (4)$$

Inequality 4 help us determining whether a certain point might change the cluster or not, but doesn't save us from visiting the entire data set. The solution is to group the data items by the value of e_P. We can split the range of values that e can take into intervals. Each group would be associated an interval. As long as e_P is greater than the interval's lower bound and lower or equal to the interval's upper bound, P would become part of the group associated with that interval. This allows us to change step 2 of algorithm 1 so that instead of visiting all points, we could visit the groups and check inequality 4. If inequality 4 returns false, it means the points inside that group have to be re-visited. Otherwise, all points inside the group will hold the cluster.

At the end of each iteration, new centroids are to be computed, which might cause centroids to deviate from their current position. To keep the groups synchronized, we will need to also update the lower and upper bounds of the associated intervals. That is, we will assume the worst case scenario presented above, which means the interval bounds would shift towards 0 by at most twice the maximum centroid deviation. More precisely, the bounds of the intervals would be reduced by $2 * max(|AA|, |BB|, |CC|)$.

Algorithm 2 resumes the solution above.

Choosing the $WIDTH$ constant has a big impact on the performance of the algorithm. The number of groups could explode if the value of $WIDTH$ is to small. The other way around, if the value of $WIDTH$ is to big, the intervals for e_P would be very wide causing them to be marked for re-visiting very often. Both extreme scenarios might reduce considerably the improvement brought by algorithm 2.

It is very hard to accurately define a general approach for choosing the right value of $WIDTH$. It depends a lot on the data distribution. However, in our research, we defined $WIDTH$'s value as the average distance between adjacent data elements. Such a value for $WIDTH$, would increase the chances of balancing the groups. Algorithm 3 explains the procedure for defining the value of $WIDTH$, as it was used in our researches:

Therefore, the data set must be analysed prior to defining $WIDTH$, but if the data set is to big, this might require more time than we gain through the proposed optimization. A good compromise would be to analyse only a sample of the data set.

It can be easily noticed that the optimized K-Means has the same output as the standard K-Means. The quality of the clusters is preserved.

Algorithm 2 Optimized K-Means

1. Define constant $WIDTH$
2. Define group intervals
$I_i = (i * WIDTH, (i+1) * WIDTH]$
3. Mark the entire data set to be visited
4. For each point to be visited
5. $e = min(d_{PC_l} - d_{PC_w})$ where C_w is the center of the closest (winner) cluster and $C_l, l = 1..k, l <> w$ stands for all other centroids
6. Map all points with $i * WIDTH < e \leq (i+1) * WIDTH$ to interval $(i * WIDTH, (i+1) * WIDTH]$ where i is a positive integer
7. Compute new centroids C_j, where $j = 1..k$ and their maximum deviation $D = max(|C_j C_j|)$
8. If $D = 0$ or the maximum number of iterations was reached, move to 11
9. Update I_i s boundaries by subtracting $2 * D$ (points owned by this interval got closer to the edge by $2 * D$)
10. Pick up all points that are mapped to an interval whose lower bound became less than or equal to 0, mark them for re-visiting, then go to 4
11. Exit

Algorithm 3 Defining the WIDTH

1. Extract a 5-10% sample of the dataset
2. Traverse each data element in the dataset and compute the distance to it's neighbours.
3. While traversing, average the distances to neighbours that do not differ by more than 50% than the current average value.
4. Assign $WIDTH$ the result of 3.

IV. Experimental evaluation

Experiments were carried out for three publicly available real data sets posted on UC Irvine Machine Learning Repository [14]. The execution environment was made of a Intel(R) Core(TM) i5-3320M CPU @ 2.60GHz, with 8GB of RAM memory, Ubuntu 14.04 operating system. We have selected three data sets that will be described below. A number of 2, 4, 8 and 12 centroids were randomly selected. Multiple runs were carried out for each scenario. The execution times shown below represent the average of all measured execution times for each scenario.

A. US Census (1990) Data Set

The US Census 1990 [15] is a raw 1% sample data set (2458285 records) of the official US Census 1990, obtained of from the (U.S. Department of Commerce) Census Bureau website. For our experiments we've clustered the records by the age information. The $WIDTH$ used for experiments, computed according to algorithm 3, was 1.

Data distribution is shown in figure 3

Results are shown in table I

One could expect such results considering that the data distribution on a wide segment (ages 0-70) is almost uniform. The improvement raises to 61.39% for two centroids, but it

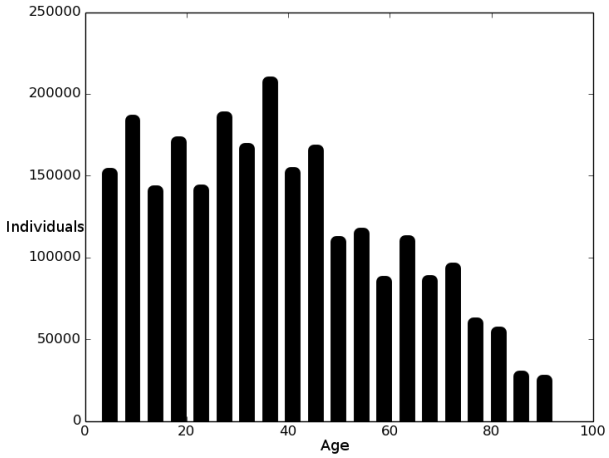


Fig. 3. US Census 1990 Data Distribution

TABLE I
RESULTS FOR US CENSUS DATA (1990) DATA SET

	Number of Centroids			
	2	4	8	12
Time(s) - Standard K-Means	1611	3016	6079	7846
Time(s) - Optimized K-Means	622	1843	4998	6539
Improvement (%)	61.39	38.89	17.78	16.66

TABLE II
RESULTS FOR 3D ROAD NETWORK DATA SET

	Number of Centroids			
	2	4	8	12
Time(s) - Standard K-Means	469	2543	9274	17894
Time(s) - Optimized K-Means	196	833	3438	8104
Improvement (%)	58.16	67.24	62.92	54.70

drops as the number of centroids grow, to as low as 16.66% in case of 12 centroids.

B. 3D Road Network Data Set

The 3D Road Network data set [16] gives the elevation information for a 2D road network in North Jutland, Denmark. There are 434874 road segments. The road segments were clustered using both standard and optimized K-Means algorithms by their elevation. After applying the algorithm 3, the resulted value for $WIDTH$ was 0.07.

Data distribution is shown in figure 4. It is not a uniform distribution, but the differences between adjacent data intervals are smooth.

Results are shown in table II

Results for the 3D Road Network Data set are very encouraging. Improvement was between 54% and 67% for a range of 2 to 12 centroids.

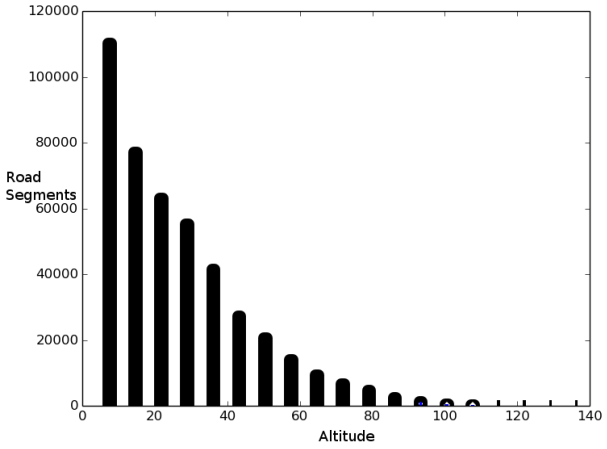


Fig. 4. 3D Road Network Data Distribution

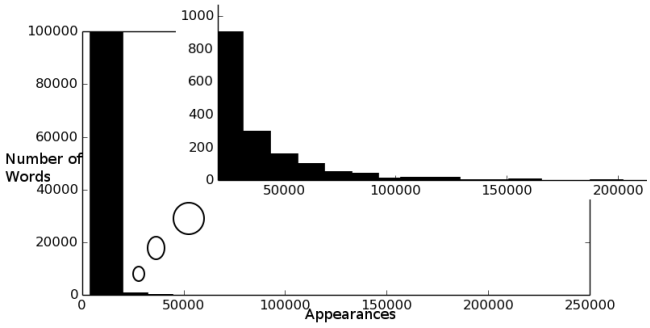


Fig. 5. Bag of Words Data Distribution

C. Bag of words Data Set

Bag of words [17] is a text collection obtained from five different sources. We've run our experiments against the New York Times collections which is made of words extracted from New York Times articles. The vocabulary is made of unique words resulted after removing stop-words and truncating the collection by only keeping words that occurred more than ten times. For each word, the number of occurrences was extracted, resulting a collection of 102660 words. The words were clustered using both standard and optimized K-Means algorithms by their number of appearances. Algorithm 3 indicated a value of 33.65 for the $WIDTH$ constant.

Data distribution is shown in figure 5

The distribution shows a concentration of words by their appearance in the interval [0 - 12000] where approximately 100000 words fit, so more than 97% of the data set.

Results are shown in table III

For the Bag of Words Data Set, the Optimized K-Means, together with algorithm 3 for determining the value of the $WIDTH$ constant, proved to be totally inefficient. They brought an increase of the execution time between 17.7% and 67.4%. The main reason for this increase seems to be the data distribution, which causes algorithm 3 to determine

TABLE III
RESULTS FOR BAG OF WORDS DATA SET

	Number of Centroids			
	2	4	8	12
Time(s) - Standard K-Means	125	339	1274	2697
Time(s) - Optimized K-Means	148	399	1768	4515
Improvement (%)	-18.6	-17.7	-38.7	-67.4

an inappropriate value for the $WIDTH$. The bigger the distance between adjacent words (in terms of the number of appearances), the greater the value of $WIDTH$ would be. A greater value for $WIDTH$ would result into wider word groups, which once invalidated, result into re-visiting a bigger number of words. Also, the wider the group, the greater the chances are for that group to be invalidated and re-visited, even on small deviations.

V. CONCLUSIONS

In this paper we evaluated an optimization for the K-Means algorithm proposed in a previous paper [1], by using real publicly available data sets. Two of the data sets (US Census (1990) and 3D Road Network) shown important improvements, very close the performance resulted by running the algorithm on the randomly generated uniform data set used in [1]. The common feature of the two data sets is that there is no sharp trend on their distribution chart. This is a very important feature when trying to calibrate the Optimized K-Means algorithm (computing the $WIDTH$ constant). The third data set used (Bag of Words), as opposed to the ones mentioned above, showed an unacceptable loss of performance. The main reason for this loss is related to the data distribution, namely, more than 97% of the data elements are concentrated into less than 5% of their distance range (number of word appearances range). Calibrating the algorithm under these circumstances requires a more intensive analysis of the data set. Our future research efforts will focus on improving algorithm 3 to cover a wider range of data distributions.

REFERENCES

[1] Cosmin Marian Poteraş, Marian Cristian Mihăescu, Mihai Mocanu: An optimized version of the K-Means clustering algorithm, Proceedings of the 2014 Federated Conference on Computer Science and Information Systems, ACSIS, Vol. 2, pages 695–699, 2014, DOI: 10.15439/2014F258.

[2] Dolnicar, S: Using cluster analysis for market segmentation—typical misconceptions, established methodological weaknesses and some recommendations for improvement, Australasian Journal of Market Research, 2003, 11(2), 5–12.

[3] Ng, H. P.; Ong, S. H.; Foong, K. W. C.; Goh, P. S.; Nowinsky, W. L.: Medical Image Segmentation Using K-Means Clustering and Improved Watershed Algorithm, 7th IEEE Southwest Symposium on Image Analysis and Interpretation, March 26-28, 2006, Denver, Colorado, pages 61-66

[4] Agnieszka Wosiak, Danuta Zakrzewska: On Integrating Clustering and Statistical Analysis for Supporting Cardiovascular Disease Diagnosis, Proceedings of the 2015 Federated Conference on Computer Science and Information Systems, ACSIS, Vol. 5, pages 303–310 (2015) DOI: 10.15439/2015F151, http://dx.doi.org/10.15439/2015F151

[5] Hongwei Xie, Li Zhang; Jingyu Sun, Xueli Yu: Application of K-means Clustering Algorithms in News Comments - The International Conference on E-Business and E-Government, May 2010, Guangzhou, China, pages 451-454

[6] kK Oyelade, O. J, Oladipupo, O. O, Obagbuwa, I. C: Application of K-Means Clustering algorithm for prediction of Students' Academic Performance, (IJCSIS) International Journal of Computer Science and Information Security, Vol. 7, No. 1, 2010, pages 292 - 295

[7] Szymon Wawrzyniak, Wojciech Niemirom: Clustering Approach to the Problem of Human Activity Recognition using Motion Data, Proceedings of the 2015 Federated Conference on Computer Science and Information Systems, ACSIS, Vol. 5, pages 411–416 (2015), DOI: 10.15439/2015F424, http://dx.doi.org/10.15439/2015F424

[8] Souptik Datta, Chris Giannella, Hillol Kargupta: K-Means Clustering Over a Large, Dynamic Network, Proceedings of the Sixth SIAM International Conference on Data Mining, April 20-22, 2006, Bethesda, MD, USA. SIAM 2006 ISBN 978-0-89871-611-5, pages 153—164.

[9] Yufang Zhang, Zhongyang Xiong, Jiali Mao, Ling O: The Study of Parallel K-Means Algorithm, Proceedings of the 6th World Congress on Intelligent Control and Automation, June 21–23, 2006, Dalian, China, pages 5868–5871.

[10] Jing Zhang, Gongqing Wu, Xuegang Hu, Shiying Li, Shuilong Hao: A Parallel K-means Clustering Algorithm with MPI, 4th Internation Symposium on Parallel Architectures, Algorithms and Programming, ISBN 978-0-7695-4575-2, pages 60-64, 2011.

[11] Jitendra Kumar, Richard T. Mills, Forrest M. Hoffman, William W. Hargrove: Parallel k-Means Clustering for Quantitative Ecoregion Delineation Using Large Data Sets, Proceedings of the International Conference on Computational Science, ICCS 2011, Procedia Computer Science 4 (2011) 1602–1611.

[12] Reza Farivar, Daniel Rebolledo, Ellick Chan, Roy Campbell: A Parallel Implementation of K-Means Clustering on GPUs, Proceedings of the International Conference on Parallel and Distributed Processing Techniques and Applications, PDPTA 2008, Las Vegas, Nevada, USA, July 14-17, 2008, 2 Volumes. CSREA Press 2008 ISBN 1-60132-084-1, pages 340–345.

[13] Mario Zechner, Michael Granitzer: Accelerating K-Means on the Graphics Processor via CUDA, The First International Conference on Intensive Applications and Services INTENSIVE 2009, 20–25 April, Valencia, Spain, pages 7–15, ISBN 978-1-4244-3683-5.

[14] M. Lichman: UCI Machine Learning Repository, University of California, Irvine, School of Information and Computer Sciences, 2013, http://archive.ics.uci.edu/ml

[15] The USCensus1990raw data set, U.S. Department of Commerce Census Bureau - http://dataferrett.census.gov/

[16] Manohar Kaul: Building Accurate 3D Spatial Networks to Enable Next Generation Intelligent Transportation Systems, Proceedings of International Conference on Mobile Data Management (IEEE MDM), June 3-6 2013, Milan, Italy

[17] http://archive.ics.uci.edu/ml/datasets/Bag+of+Words

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 837–840
DOI: 10.15439/2016F45
ACSIS, Vol. 8. ISSN 2300-5963

Automatic Mapping of MySQL Databases to NoSQL MongoDB

Liana Stanescu
University of Craiova Faculty of Automation, Computers and Electronics Bvd. Decebal 106
Craiova, Romania
Email: stanescu@software.ucv.ro

Marius Brezovan
University of Craiova Faculty of Automation, Computers and Electronics Bvd. Decebal 106
Craiova, Romania
Email: mbrezovan@software.ucv.ro

Dumitru Dan Burdescu
University of Craiova Faculty of Automation, Computers and Electronics Bvd. Decebal 106
Craiova, Romania
Email: dburdescu@yahoo.com

***Abstract*—The paper presents a framework that implements our original algorithm of automatic mapping a MySQL relational database to a MongoDB NoSQL database. The algorithm uses the metadata stored in the MySQL system tables. It takes into consideration the concepts from Entity-Relationship (ER) model: entity type represented by a relation in the Relational Model (RM), 1:1 and 1:M relationship type represented with Foreign Keys (FK) in the RM and N:M relationship type represented in RM with a join table that contains the Primary Keys (PK) from the original tables, each representing a FK and two 1:M relationships between the original tables and the join table. The initial results of our algorithm that was tested on small size databases (10-15 tables with many relationships and 100 records/ table) are presented in this paper.**

I. Introduction

RELATIONAL Database Management Systems (RDBMS) have became the first choice for the storage of information in databases mostly used for financial records, manufacturing information, staff and salary data, and so on starting with 1980. RDBMSs are based on the relational model defined by a schema. This model uses two concepts: table and relationship. A relational table represents a well defined collection of rows and columns and the relationship is established between the rows of the tables. Relational data can be queried and manipulated using SQL query language [8].

In practice, there are situations in which storing data in the form of a table is inconvenient, or there are other kinds of relationships between records, or there is the necessity to quickly access the data. In order to solve such problems a new type of NoSQL has been created. A NoSQL or Not Only SQL database provides a mechanism for storage and retrieval of data that is different from the typical relational model.

Another issue that NoSQL solves is the mismatch between relational databases and object-oriented programming. It is known that SQL queries are not well suited for the object oriented data structures that are used in most applications now [8].

Another closely related issue is storing or retrieving an object along with all relevant data. Some operations require multiple and complex queries. In this case, data mapping and query generation complexity rise too much and becomes difficult to be maintained on the application side [8].

Some of these problems have found their answer both in Object-relational mapping (ORM) frameworks, even though it still requires a lot of development effort and also in Object-Oriented Database Management Systems (OODBMS). The down side in this last alternative is the fact that it did not gain much popularity in replacing relational databases. However, most object oriented databases may be considered NoSQL solutions as well [8].

Another problem that relational databases cannot handle is related to an exponentially increasing amount of data. The direct consequence is the so-called big data problem. This problem arises when standard SQL query operations do not have acceptable performances, especially when transactions are involved [8].

As a result, the subject of developing an automatic mapping instrument has been brought up. This instrument will be able to represent the existing relational databases, already populated with a large number of records, as NoSQL databases. A significant amount of time and human effort is spared this way, which would have otherwise been needed to create and populate the NoSQL database from scratch.

This paper proposes a framework which implements an algorithm for automatic mapping of MySQL relational databases to MongoDB.

The paper is organized in the following way: Section II presents the main concepts used by MongoDB, one of the most used NoSQL database, Section III presents general principles of mapping relational databases to MongoDB considering the main concepts from ER model and RM. Section IV presents in detail our proposed algorithm for automatic mapping of a MySQL database to MongoDB and also an example of application of this algorithm. Conclusions and future work are shown in Section V.

II. MongoDB

MongoDB is a cross-platform, document oriented database that provides high performance, high availability and easy scalability. The main concepts in MongoDB are collection and document. Database is a physical container for collections. Each database receives its own set of files on

the file system. A single MongoDB server typically manages multiple databases [2], [3], [4], [5].

The collection is a group of MongoDB documents. It has as correspondent a RDBMS table. A collection exists only within a single database.

A MongoDB document is a set of key-value pairs. The documents do not have to necessarily respect a schema. Typically, all documents in a collection are of similar or have a related purpose. Dynamic schema means that documents in the same collection do not need to have the same set of fields or structure, and common fields in a collection's documents may hold different types of data.

Table 1 shows the relationship of RDBMS terminology with MongoDB [3], [4], [5].

TABLE I.
THE RELATIONSHIP OF RDBMS TERMINOLOGY WITH MONGODB

RDBMS	MongoDB
Database	Database
Table	Collection
Tuple/Row	Document
Column	Field
Table Join	Embedded Documents
Primary Key	Primary Key (Default key _id provided by MongoDB itself)

Any relational database has a certain design schema that shows the tables and the relationships between them. In MongoDB there is no concept of relationship.

The advantages of MongoDB over RDBMS are [2], [3], [4], [5]: schema-less (MongoDB is a document database in which one collection holds different documents whose number of fields, content and size can be different from one document to another), structure of a single object is clear, no complex joins, deep query-ability (MongoDB supports dynamic queries on documents using a document-based query language that is almost as powerful as SQL), tuning, ease to scale-out, conversion / mapping of application objects to database objects is not needed, uses internal memory for storing the working set, enabling faster access to data.

III. THE GENERAL PRINCIPLES OF MAPPING RELATIONAL DATABASES TO MONGODB

In MongoDB the relational database remains a database. A relational table is mapping to a MongoDB collection. The tuples or rows become documents inside MongoDB collections [3], [4], [5].

The 1:1 relationship describes a relationship between two entities. For example a Student has a single Address relationship. A Student lives at a single Address and an Address only contains a single Student. The 1:1 relationship can be modeled in two ways using MongoDB. The first is to embed the relationship as a document and the second is as a link to a document in a separate collection [4], [5].

In the one to one relationship embedding is the preferred way to model the relationship as it's more efficient to retrieve the document.

The 1:M relationship describes a relationship where one side can have more than one relationship while the reverse relationship can only be single sided.

The 1:M relationship can be modeled in several different ways using MongoDB. The first model is embedding, the second is linking and the third is a bucketing strategy that is useful for cases like time series [4], [5].

A N:M relationship in the ER model is an example of a relationship between two entity types where they both might have many relationships between entities. An example might be a Book that was written by many Authors. At the same time an Author might have written many Books.

N:M relationships are modeled in the relational database by using a join table that contains the primary keys from the original ones, each representing a foreign key, and two 1:M relationships.

In MongoDB we can represent this situation in many ways. The first way is called Two Way Embedding [4], [5].

In Two Way Embedding we will include the Book foreign keys under the book field in the author document. Mirroring the Author document, for each Book we include the Author foreign keys under the Author field in the book document.

Another way of modeling N:M relationships is called One Way Embedding [4], [5].

The One Way Embedding strategy chooses to optimize the read performance of a N:M relationship by embedding the references in one side of the relationship. An example might be a N:M relationship between books and categories. The case is that several books belong to a few categories but a couple categories can have many books. Let's look at an example with the categories represented into a separate document.

An example of Category documents:

{id=1,
Cname="Multimedia"}
{id=2,
Cname="Databases"}

An example of a Book document with foreign keys for Categories:

{id:1,
title"Multimedia Databases",
categories:[1, 2],
authors:[1]}
id:2,
title"Multimedia",
categories:[1],
authors:[1,2]}

IV. FRAMEWORK DESCRIPTION

The framework implements an algorithm of automatic mapping of MySQL relational databases to MongoDB.

The algorithm uses the MySQL INFORMATION_SCHEMA that provides access to database metadata. Metadata is data about the data, such as the name of a database or table, the data type of a column, or access privileges. INFORMATION_SCHEMA is the information database, the place that stores information about

all the other databases that the MySQL server maintains. Inside INFORMATION_SCHEMA there are several read-only tables. They are actually views, not base tables [6], [7]. For examples we use a database db1 that contains 5 tables: Employee, Department, Project, Child and Works_on. The relationships are 1:M between Department and Employee, Employee and Child, Department and Project (figure 1). In the ER model there is a N:M relationship, implemented in the relational model by the join table Works_on and two 1:M relationship between Employee and Works_on and Project and Works_on.

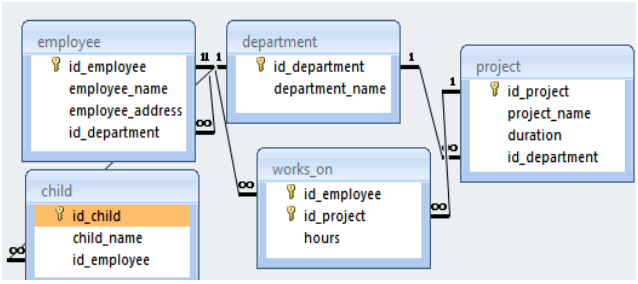


Fig. 1 A relational database

The steps of the algorithm implemented in our framework are presented next.

1. Creating the MongoDB database

The user must specify the MySQL database that will be represented in MongoDB. The database is created with the following MongoDB command: use DATABASE_NAME [5].

>use db1
switched to db db1

2. Creating tables in the new MongoDB database

The algorithm verifies for each table in what relationships is involved, if it has foreign keys and/or is referred by other tables.

2.1 If the table is not referred by other tables, it will be represented by a new MongoDB collection.

2.2 If the table has not foreign keys, but is referred by another table, it will be represented by a new MongoDB collection.

2.3 If the table has one foreign key and is referred by another table, it will be represented by a new MongoDB collection. In our framework, for this type of tables we use linking method, using the same concept of foreign key.

2.4 If the table has one foreign key but is not referred by another table, the proposed algorithm uses one way embedding model. So, the table is embedded in the collection that represents the table from the part 1 of the relationship.

2.5 If the table has two foreign keys and is not referred by another table, it will be represented using the two way embedding model, described in section IV.

2.6 If the table has 3 or more foreign keys, so it is the result of a N:M ternary, quaternary relationships, the algorithm uses the linking model, with foreign keys that refer all the tables initially implied in that relationship and already represented as MongoDB collections. The solution is good even the table is referred or not by other tables.

In order to find the name of the tables stored in MySQL database the next Select command is used:

Select table_name From information_schema.tables
Where table_schema='db1' Order By table_name;

The INFORMATION_SCHEMA.TABLES provides information about tables in databases [6], [7].

The table TABLES_CONSTRAINTS from INFORMATION_SCHEMA database describes which tables have constraints [6], [7]. It must be executed a Select command on each table in database, as in the next example:

Select constraint_type
From information_schema.tables_constraints
Where table_schema='db1'
And table_name='department';

The CONSTRAINT_TYPE value can be unique, primary key or foreign key. We are interested by primary key and foreign key constraints [6], [7].

The table REFERENTIAL_CONSTRAINTS from the same MySQL system database provides information about foreign keys. The attributes constraint_schema and constraint_name identify the foreign key. The attributes: unique_constraint_schema, unique_constraint_name and referenced_table_name identify the referenced key [6], [7].

With the data from these system tables: tables, tables_constraints and referential_constraints the framework can establish what step of the algorithm (2.1-2.6) must be applied.

Relational tables become collections in MongoDB. The collections are created using createCollection() method.

Basic syntax of **createCollection()** command is as follows [5]:

Db.createCollection(name, options)

Where name represents the collection name and options specify options about memory size and indexing.

For the relational database from figure the mapping according to the presented algorithm is presented next. The table Department has no foreign keys but is referred by other two tables, so it becomes a MongoDB collection (step 2.2).

The table Employee has one foreign key and is referred by the table Works_on, so it becomes a collection (step 2.3).

The table Project has one foreign key and is referred by Works_on, so it becomes also a collection (step 2.3).

The table Works_on has two foreign keys and is not referred by other tables, so it will be implemented using two way embedding model (step 2.5). The projects will be assigned to each employee and also, to each project will be assigned the employees that work on that project.

The table Child has foreign key but is not referred by another table, so it will be represented using one way embedding model (step 2.4). So it will be embedded in Employee collection.

The five relational tables will be represented by three MongoDB collections. Next, there are some samples of the MongoDB collections generated by the presented algorithm.

Department collection is presented in figure 2.

Fig. 2 MongoDB collection that represents the Department table

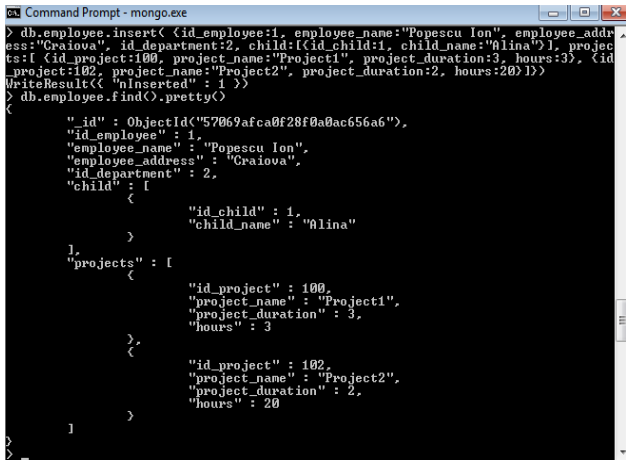


Fig. 3 MongoDB Employee collection

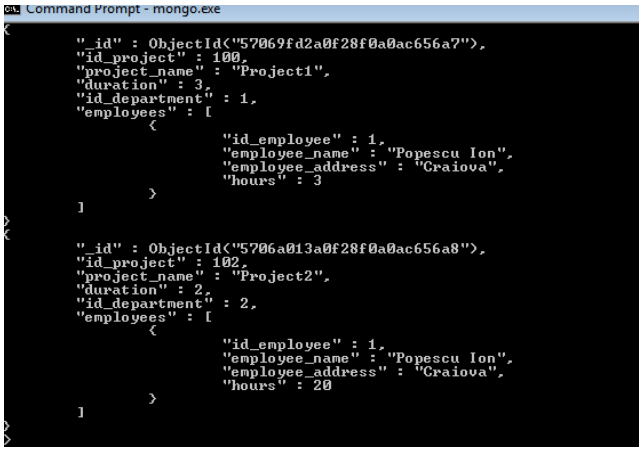


Fig. 4 MongoDB Project collection

Employee Collection that contains the document Child and Projects is presented in figure 3.

Project collection that embeds the documents employees that work on these projects is shown in figure 4.

V. Conclusion and Future Work

The paper presents a framework that implements our original algorithm of automatic mapping a MySQL relational database to a MongoDB NoSQL database. The algorithm uses the metadata stored in the MySQL system tables. It takes into consideration the concepts from Entity-Relationship (ER) model: entity type represented by a relation in the Relational Model (RM), 1:1 and 1:M relationship type represented with Foreign Keys (FK) in the RM and N:M relationship type represented in RM with a join table that contains the Primary Keys (PK) from the original tables, each representing a FK and two 1:M relationships between the original tables and the join table.

The algorithm was presented in detail, on steps. Also, the paper contains an example of automatic mapping of a MySQL database to MongoDB using our algorithm.

The paper also presents the initial results of our algorithm that was tested on small size databases (10-15 tables with many relationships and 100 records/table), the results being encouraging.

The future work will include the next steps:

- Experiments on complex databases (many tables and a large number of records/table)
- Taking into consideration the number of records in the tables and the operations on the database (insert, update, delete query) in order to implement the more appropriate model of mapping to MongoDB
- Modeling tree structures with parent references
- Extending the framework to execute mapping to MongoDB of other relational databases (Oracle, MS SQL Server and so on)

References

[1] http://www.datastax.com/nosql-databases
[2] https://www.thoughtworks.com/insights/blog/nosql-databases-overview
[3] https://leanpub.com/mongodbschemadesign/read
[4] http://code.tutsplus.com/articles/mapping-relational-databases-and-sql-to-mongodb--net-35650
[5] https://docs.mongodb.org/manual/core/data-modeling-introduction/
[6] https://dev.mysql.com/doc/refman/5.0/en/information-schema.html
[7] https://dev.mysql.com/doc/refman/5.5/en/introduction.html
[8] https://www.devbridge.com/articles/benefits-of-nosql/

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 841–848
DOI: 10.15439/2016F95
ACSIS, Vol. 8. ISSN 2300-5963

Real-Time Implementation of DC Servomotor Actuator with Unknown Uncertainty using a Sliding Mode Observer

Roxana-Elena Tudoroiu
University of Petrosani
20 Universităţii street, 332006, Petroşani, Hunedoara, Romania
Email: tudelena@mail.com

Wilhelm Kecs
University of Petrosani
20 Universităţii street, 332006, Petroşani, Hunedoara, Romania
Email: wwkecs@yahoo.com

Maria Dobritoiu
University of Petrosani
20 Universităţii street, 332006, Petroşani, Hunedoara, Romania
Email:mariadobritoiu@yahoo.com

Nicolae Ilias
University of Petrosani
20 Universităţii street, 332006, Petroşani, Hunedoara, Romania
Email: iliasnic@yahoo.com

Stelian-Valentin Casavela
University of Petrosani
20 Universităţii street, 332006, Petroşani, Hunedoara, Romania
Email: svcasavela@yahoo.com

Nicolae Tudoroiu
John Abbott College 2127
Lakeshore Road, Sainte-Anne-de-Bellevue, QC, H9X 3L9, Canada
Email: ntudoroiu@gmail.com

***Abstract*—The central idea of this paper is the modeling and implementation of a real-time dc servomotor angular speed control system with an unknown bounded uncertainty using a sliding mode observer (SMO) control strategy. We prefer to use a SMO in our approach due to its great potential in fault detection and isolation (FDI) of the actuators and sensor faults subjected frequently to several failures due to an abnormal change in their operating conditions or parameters. We use for this purpose the most suitable real time implementation tool MATLAB/SIMULINK software package. It provides special features for real time implementation by its extensions Real-Time Workshop (RTW) and the Real-Time Windows Target (RTWT). The novelty of our paper is to prove in an extensive simulation MATLAB/SIMULINK frame the real time implementation potential of a most recently sliding mode observer (SMO) control strategy applied to a particular case study, namely for a dc servomotor angular speed control system. The proposed real-time Sliding Mode Observer (SMO) consists of an embedded nonlinear Sliding Mode Observer (SMO) with the dc servomotor actuator in an integrated control system structure to estimate its angular speed and armature current and to implement the sliding mode control law.**

I. Introduction

THE most important common actuator integrated in a forward path of the control systems structure is the dc servomotor. It directly provides rotary motion and, coupled with wheels or drums and cables, can also provide transitional motion. The cause roots of the majority faults in control systems are the result of unexpected failures, interferences as well as the age of their crucial components. In our research defective measurement and control loops equipment, in particular a sensor and an actuator are under investigation. More precisely, we assume that the dc servomotor control system is subjected to an unknown but bounded uncertainty, and its speed is controlled in closed-loop feedback by using a state feedback control law with an embedded sliding mode observer (SMO) that is integrated in the closed-loop control structure. The dc servomotors are preferred in control area applications because they possess a high start torque characteristics, high response performance, and easier to be linearly controlled etc. Also the dc servomotor speed can be adjusted by varying its input voltage. Therefore, the dc motor control is better compared to all ac induction motors that need expensive frequency drivers. The real-time dc servomotor speed control can be easily understood and interfaced with MATLAB/SIMULINK or AnyLogic multi-paradigms hybrid simulator in the case of UML-RT implementations [8] with a fast response. Furthermore, the preliminary results of this research will be useful for us for future exploration of using a sliding mode observer to develop more attractive control strategies, for example the detection and isolation of the faults (FDI) [1], [5], [6], [7] that occur in the actuators and sensors, and also for trust modeling in Multi-agent Systems [3], will be an interesting future directions of our research. Whenever these critical situations come out the control systems could lose the control, require much more energy, and could operate harmfully. Therefore to operate in real-time at high energy efficiency and to guarantee the equipment safety and reliability it is important to develop suitable FDI strategies capable to detect and diagnose any time every faulty control system components and consequently corrective and reconfiguration actions should be initiated promptly [7]. Effectively the existing methods to identify and to adjust the equipment failures are mostly labor-intensive task, and consequently sustained, rhythmic and error-prone [7]. In the majority of these situations the windings currents are recorded to determine the health of the dc servomotors currents compared to statistical evaluation that necessitates considerable human knowledge, hence error-prone that could generate severely equipment operation. In these conditions the problem of control systems monitoring and fault diagnosis becomes a critical issue, of high complexity that need to be

978-83-60810-90-3/$25.00

implemented in mainframe environment by using more sophisticated control systems and artificial intelligence strategies. This research work is based on our previous results in implementation of different control systems strategies and now we are interested to prove the effectiveness of real-time implementation of our proposed SMO control strategy. Closing, the main objective of our research is to develop a more suitable, accurate and consistent real-time nonlinear SMO control strategy to be used in our future real-time FDI control strategies implementation development.

II. THE DC SERVOMOTOR DYNAMICS

The electric circuit of the dc servomotor armature and the free body diagram of the rotor are shown in figure 1. For simulation purpose we will assume the following experimental values for the physical parameters [9]:

- moment of inertia of the rotor: $J = 0.001\ [\frac{kgm^2}{s^2}]$
- damping ratio of the mechanical system: $b = 0.01[Nms]$
- counter electromotive force coefficient (*cemf*): $k_e = k_t = k = 0.0517\ [Nm/A]$
- motor electric resistance: R = 1 [Ω]
- motor electric inductance: $L = 0.5[H]$
- motor initial angular speed: $\omega = 1[\frac{\text{rad}}{\text{s}}]$,
- input dc power supply: $V = 12[V]$
- load torque: $T_L = 0.1\sin(t)\ [Nm]$

The dynamics of the dc servomotor actuator is described by the following input-state-output equations [8]-[9] :

$$J\frac{d^2\theta}{dt^2} + b\frac{d\theta}{dt} = k_t I_a - T_L$$
$$L\frac{dI_a}{dt} + RI_a = V - k_e\frac{d\theta}{dt}. \quad (1)$$

where $T_e = T = k_t I_a$ is the dc servomotor torque developed to the shaft, T_L is the load torque, and $e = k_e\frac{d\theta}{dt} = k_e\omega$ is the counter electromotive force (*cemf*) of the dc servomotor.

In a state-space representation the dc servomotor actuator dynamics is described by following equations:

(i) State Equation:

$$\begin{bmatrix}\frac{dx_1}{dt}\\ \frac{dx_2}{dt}\end{bmatrix} = \begin{bmatrix}-\frac{b}{J} & \frac{k_t}{J}\\ -\frac{k_e}{L} & -\frac{R}{L}\end{bmatrix}\begin{bmatrix}x_1\\ x_2\end{bmatrix} + \begin{bmatrix}-\frac{T_L}{J}\\ \frac{u}{L}\end{bmatrix} \quad (2)$$

$$x_1 = \omega, x_2 = I_a, u = V, x_1(0) = 1\left[\frac{rad}{s}\right], x_2(0) = 0[A]$$

(ii) Output equation:

$$y = [1 \quad 0]\begin{bmatrix}x_1\\ x_2\end{bmatrix} \quad (3)$$

where we assume that the only the state x_1 is measurable, and the input command $u = V$ is a function defined by the armature voltage (dc power supply) and is designed in closed-loop as a state feedback control law. Also the load torque T_L is considered unknown which is bounded and has a bounded derivative.

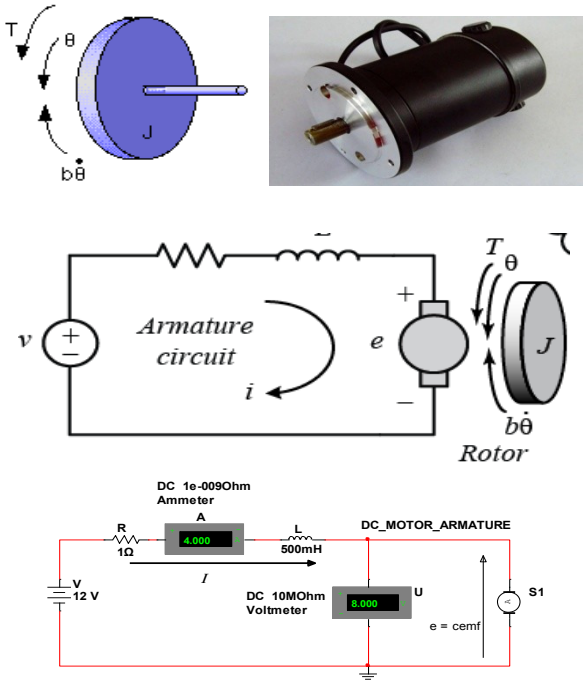


Fig. 1 The simplified equivalent electrical circuit of the dc servomotor (Reproduced from [8]-[9])

III. SLIDING MODE CONTROLLERS VERSUS SLIDING MODE OBSERVERS

To formulate any practical control application is not a straightforward task due to a more or less mismatch between the actual plant and its mathematical model used for the controller design. This mismatch could arise from unknown external disturbances, plant parameters, or modeling errors. To design a suitable control law in order to provide the desired performance to the negative feedback closed-loop control system in the presence of these disturbances or uncertainties is a very difficult task for any control engineer. This has led to an intensive research to develop some kind of robust control methods that are supposed to solve this problem. The sliding mode control (SMC) is one of particular attractive approach to design a robust controller due to its potential as a robust control method. A sliding mode control (SMC) is characterized by a suite of feedback control laws and a decision rule [1]. The decision rule is a sort of switching function that has as its input some measure of the current system behavior in order to generate as an output a particular feedback robust controller which should be used at that time instant [1]-[2]. The sliding mode control (SMC) strategy is designed such that the variable structure control systems have to be capable to drive the system states on the convergent phase trajectories and then to constrain these to lie within a neighborhood of the switching function

that can be represented by a switching line or switching surface, such is shown in figure 2 [8]. In this approach the dynamic behavior of the control system may be adapted to a particular choice of switching function, and also the feedback closed-loop response becomes totally insensitive to a particular class of uncertainty in the system, more precisely proving a high robustness feature to all kind of system uncertainties. The main drawback of using the sliding mode control (SMC) techniques to design a large variety of control industrial applications is the necessity to implement a basically discontinuous control signal which hypothetically must switch with infinite frequency to provide total rejection of the uncertainties [1]-[2].

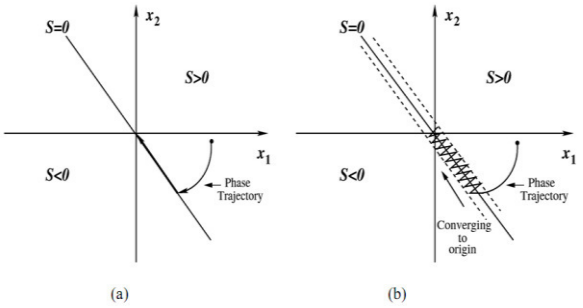


Fig. 2 Switching dynamics represented in phase plane for: (a) ideal sliding mode control and (b) practical sliding control mode (Reproduced from [8])

In contrast, the application of sliding mode methods in combination with observer control problems provide the ability to generate a sliding motion on the error between the measured plant output and the output of the observer such that to ensure that a sliding mode observer (SMO) produces a set of state estimates precisely matching with the actual output of the plant. Also the analysis of the average value of the applied observer injection signal, the so-called equivalent injection signal, contains useful information about the mismatch between the model used to define the observer and the actual plant. Furthermore, the discontinuous injection signals in this case don't remain anymore a drawback for software based observer frameworks in control applications [1], [2], [4], [5].

The development of SMO control strategy design in this research paper follows the same design guidelines suggested in [1],[2] and [4], [5] related to the design of sliding mode observer (SMO) control strategies for fault detection and isolation based on the equivalent injection signal principle.

IV. Sliding Mode Observer for Linear DC Servomotor Without Model or Disturbances Uncertainties

For simulation purpose we assume in this section that the experimental values for the physical parameters of the dc servomotor are the same with those introduced in section II. With these values, the dynamics of the dc servomotor is described in state-space representation by the following first order differential equation:

$$\frac{dx}{dt} = A_{n\times n}x + B_{n\times m}u + D_{n\times q}\Psi(x,u,t)$$
$$y = C_{p\times n}x = [1 \quad 0]x \tag{2}$$

$$A = \begin{bmatrix} -10 & 51.7 \\ -0.1034 & -2 \end{bmatrix}, B = \begin{bmatrix} 0 \\ 2 \end{bmatrix}, D = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$
$$\Psi(x,u,t) = (-1000T_{Load}) = -10\sin(t).$$

where $x \in R^n$ is a n - dimensional state vector $(n = 2)$, $y \in R^p$ is a p-dimensional output vector $(p = 1)$, and $u \in R^m$ is a m-dimensional input vector $(m = 1)$. For this development phase the load torque disturbance uncertainty is considered $T_{Load} = 0$, (dc Servomotor no-load speed) and so $\Psi(x,u,t) = 0$, in order to investigate in the first phase of our development only the linear case. By some manipulations of the matrices A, B, C we can easily find that B, C have a full rank and the pair (A, C) is observable, as main requirements assumed in [1].

To design a sliding mode observer (SMO) for this control system we follow the same procedure as in [1]. Firstly we attach to the original system an *Utkin observer* [1], [2] in a canonical form. For this task we need to find a nonsingular state transform $T_c \in R^{n\times n}$ that changes the state vector x in a state vector

$$z = T_c x = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}, z_1 \in R^{n-p}, z_2 \in R^p, T_c = \begin{bmatrix} N_c^T \\ C \end{bmatrix},$$
$$N_c \in R^{n\times(n-p)} \tag{3}$$

where the column of the matrix N_C spans the null space of $C, \exists z_1 \neq 0_{(n-p)\times 1} \xrightarrow{yields} N_c \times z_1 = 0_{(n-p)\times 1}$.

This nonsingular state transform T_c converts the nominal system (2) in the following canonical form:

$$\frac{dz_1}{dt} = A_{11}z_1(t) + A_{12}z_2(t) + B_1u(t)$$
$$\frac{dz_2}{dt} = A_{21}z_1(t) + A_{22}z_2(t) + B_2u(t) \tag{4}$$

Now the dynamics of the observer is described by the following similar equations:

$$\frac{d\hat{z}_1}{dt} = A_{11}\hat{z}_1(t) + A_{12}\hat{z}_2(t) + B_1u(t) + L\vartheta$$
$$\frac{d\hat{z}_2}{dt} = A_{21}\hat{z}_1(t) + A_{22}\hat{z}_2(t) + B_2u(t) - \vartheta \tag{5}$$

where the pair $(\hat{z}_1, \hat{z}_2)$ represent the estimated values of the transformed components state vector z, and $L \in R^{(n-p)\times p}$ is the observer gain matrix, given by

$$\vartheta_i = Msgn(\hat{z}_{2,i} - z_{2,i}), M \in R_+, i = 1, \dots, p \tag{6}$$

The dynamics of the system errors is described by the following first order differential equations:

$$\frac{de_1}{dt} = A_{11}e_1(t) + A_{12}e_2(t) + L\vartheta$$
$$\frac{de_2}{dt} = A_{21}e_1(t) + A_{22}e_2(t) - \vartheta \tag{7}$$
$$e_1(t) = \hat{z}_1(t) - z_1(t),\ e_2(t) = \hat{z}_2(t) - z_2(t)$$

The observer gain matrix $L \in R^{(n-p)\times p}$ is chosen in order to make the spectrum of the matrix ($A_{11} + LA_{21}$) to lie in $\mathbb{C}_-$, where the pair matrices (A_{11}, A_{21}) is observable due to the fact that the pair (A, C) is also observable.

Without to lose the generality we can choose the coordinates transform matrix such as:

$$T_c = \begin{bmatrix} N_c^T \\ C \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad (8)$$

that converts the triple (A, B, C) into $(\tilde{A}, \tilde{B}, \tilde{C})$, where the lines of the matrix N_c^T span the null space of the vector C, and also:

$$\tilde{A} = T_c A T_c^{-1} = \begin{bmatrix} -2 & -0.1034 \\ 51.70 & -10 \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$
$$\tilde{B} = T_c B = \begin{bmatrix} 2 \\ 0 \end{bmatrix} = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}, \tilde{C} = C T_c^{-1} = [0 \quad 1] = [C_1 \quad C_2] \quad (9)$$

From the matrix structure $\tilde{A}$ we get $A_{11} = -2$ (stable), and $A_{12} = -0.1034$, $A_{21} = 51.70, A_{22} = -10$. Also, $B_1 = 2, B_2 = 0, C_1 = 0, C_2 = 1$.

The value of the observer matrix gain L can be choose such as $A_{11} + LA_{21} < 0$, let take this $L = -1 < -\frac{A_{11}}{A_{21}} = 0.0387$. Setting the observer matrix gain L to 0.01 the dynamics of the linear observer and of its error are described by the following first order differential equations:

$$\begin{aligned} \frac{d\hat{z}_1}{dt} &= -2\hat{z}_1(t) - 0.1034\hat{z}_2(t) + 2u(t) - \vartheta \\ \frac{d\hat{z}_2}{dt} &= 51.70\hat{z}_1(t) - 10\,\hat{z}_2(t) - \vartheta \\ \frac{de_1}{dt} &= -2e_1(t) - 0.1034e_2(t) + 0.01\vartheta \\ \frac{de_2}{dt} &= 51.70e_1(t) - 10e_2(t) - \vartheta \\ \vartheta &= sgn(\hat{z}_2 - z_2) = sgn(e_2(t)), M = 1 \end{aligned} \quad (10)$$

4.1 SIMULATION RESULTS

In all the above equations we set $u(t) = 12\,[V]$ in order to solve the problem of designing a sliding mode observer (SMO) by using MATLAB/SIMULINK software package. The model of the original system in SIMULINK is shown in figure 3, and the evolution of the states, i.e. angular speed(x_1) and armature current (x_2) are shown in figures 4 and 5. The state-space representation model of the dc servomotor in canonical form and the *Utkin classical observer* dynamics model embedded in the same integrated control structure are represented in MATLAB/SIMULINK, as are shown in figure 6 and the evolution of the both model and estimated states, namely the angular speed (z_2) and the armature current (z_1) are shown in figures 7 and 8. The dc servomotor angular speed (e_2) and armature currents (e_1) SMO residuals are shown in figures 9 and 10. In figure 11 are shown the SMO errors dynamics modeled in SIMULINK.

An ideal sliding motion will take place on the sliding surface [1], [2], [4], [5]:

$$S_w = \{(e_1, e_2) | e_2 = 0)\} \quad (11)$$

After some finite time t_s, for all subsequent time, $e_2 = 0$, and $\frac{de_2}{dt} = 0$.

The corresponding sliding mode dynamics are given in [1]-[2]:

$$\frac{d\tilde{e}_1(t)}{dt} = \tilde{A}_{11}\tilde{e}_1(t) \quad (12)$$

where

$$\tilde{e}_1(t) = e_1(t) + Le_2(t) = e_1(t) - e_2(t)$$
$$\tilde{A}_{11} = A_{11} + LA_{21} = -2 - 51.70 = -53.70 < 0.$$

Since $\tilde{A}_{11} < 0$, the linear homogenous equation (12) has a stable solution, $\tilde{e}_1(t) = C_0 e^{-53.70t}$, where C_0 is a integration constant determined from the initial condition $\tilde{e}_1(0) = \tilde{e}_{10}$. By a suitable choice of the gain L, such as in our case study $L = -1 < 0.0387$,we can conclude that always the system is stable, therefore $\tilde{e}_1(t) \xrightarrow{yields} 0$, and also $\hat{z}_1(t) \xrightarrow{yields} z_1(t)$ as $t \xrightarrow{\Delta} \infty$.

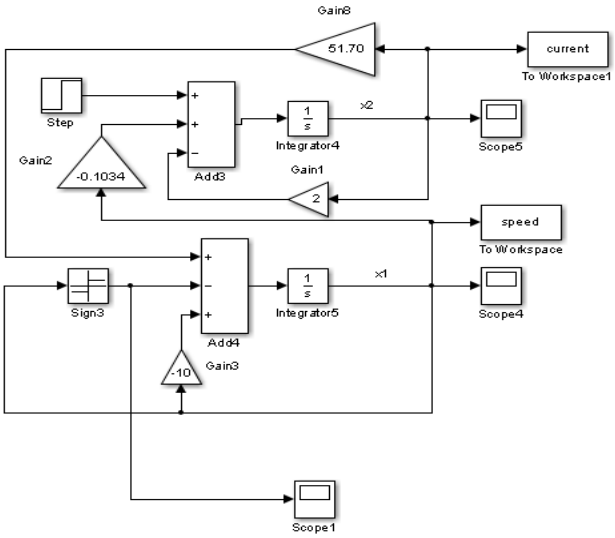


Fig. 3 DC servomotor state space-representation of nominal model in MATLAB/SIMULINK

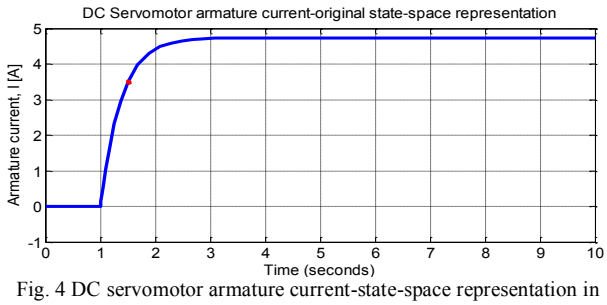


Fig. 4 DC servomotor armature current-state-space representation in MATLAB/SIMULINK

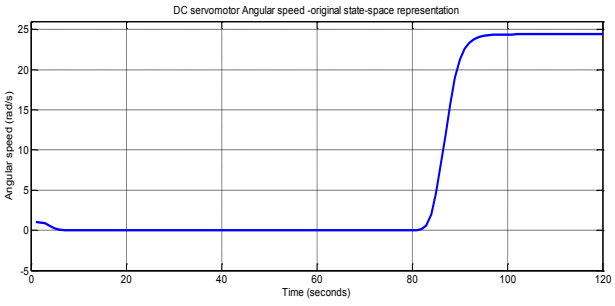


Fig. 5 DC servomotor angular speed- original state-space representation model in MATLAB/SIMULINK

In figure 12 is captured much more commutations around the sliding line:

$$S_w: \frac{de_2}{dt} = 0, e_2 = 0$$

and is shown the time switching control function

$$\vartheta = sgn(\hat{z}_2 - z_2) = sgn(e_2(t)), M = 1,\ L = -1.$$

Secondly, we build a classical Utkin observer with linear injection that is used in next section to introduce a constructive sliding mode observer design framework.

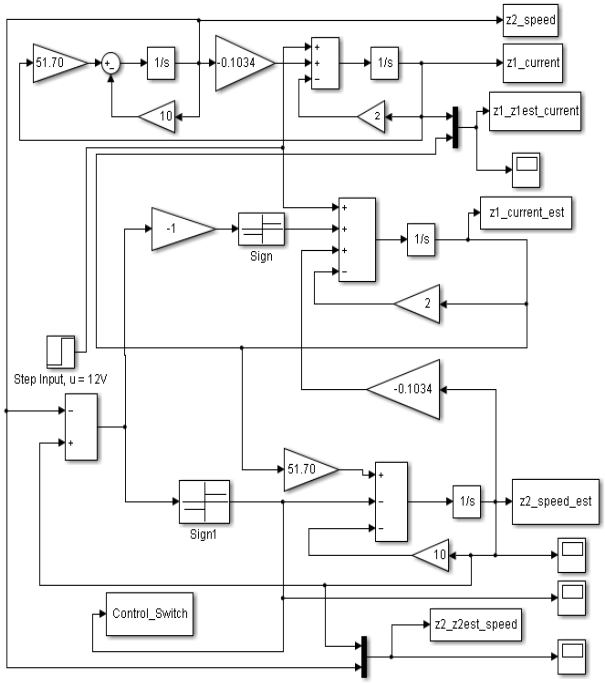


Fig. 6 Sliding Mode Observer - state space-representation in MATLAB/SIMULINK

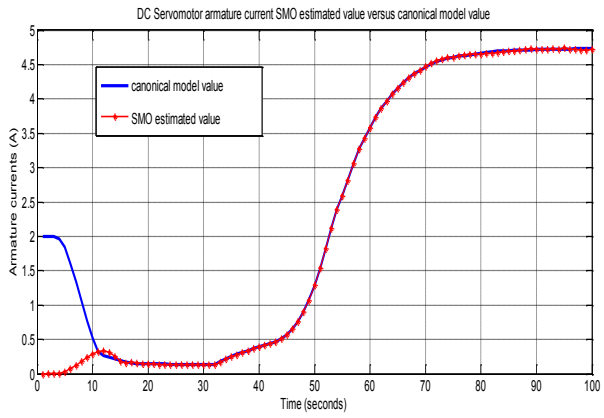


Fig. 7 DC servomotor armature current estimated versus canonical model current using SMO control strategy in MATLAB/SIMULINK

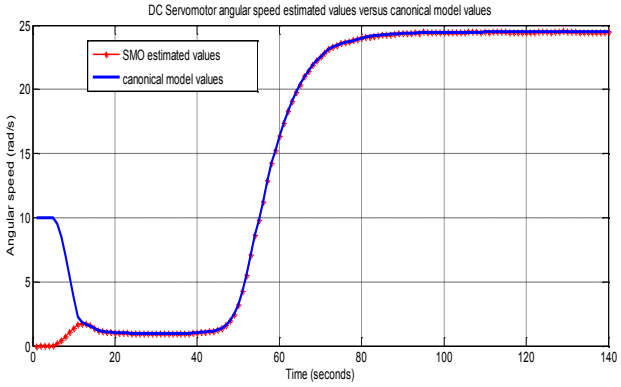


Fig. 8 DC servomotor angular speed estimated versus model angular speed using SMO control strategy in MATLAB/SIMULINK

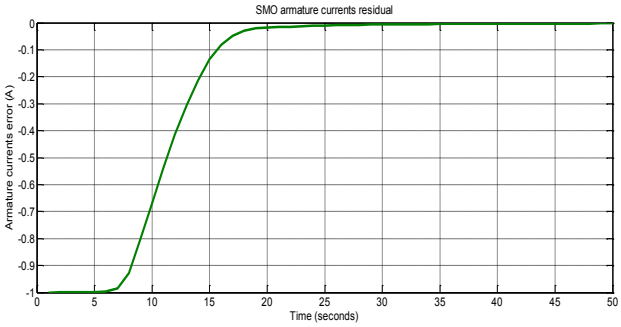


Fig. 9 DC servomotor SMO armature currents residual

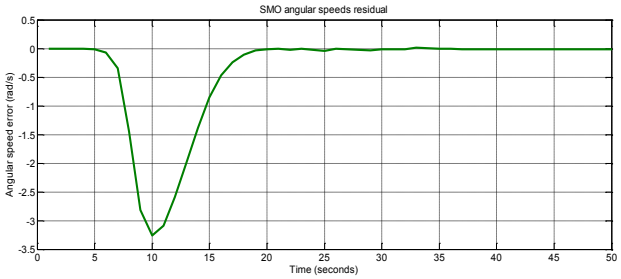


Fig. 10 DC servomotor angular residual speed using SMO control strategy

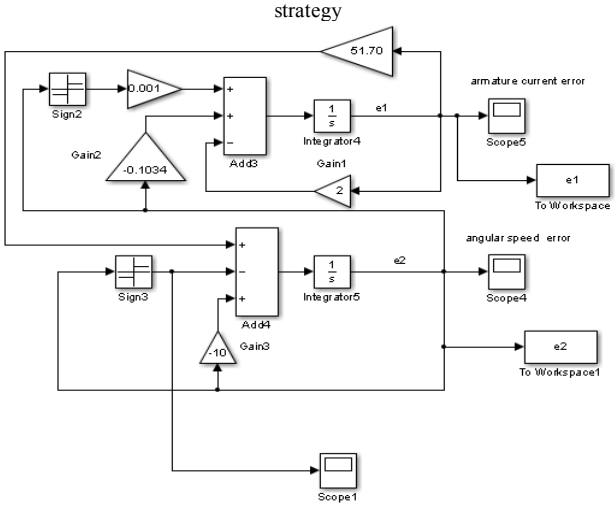


Fig. 11 DC servomotor SMO error dynamics - SIMULINK model

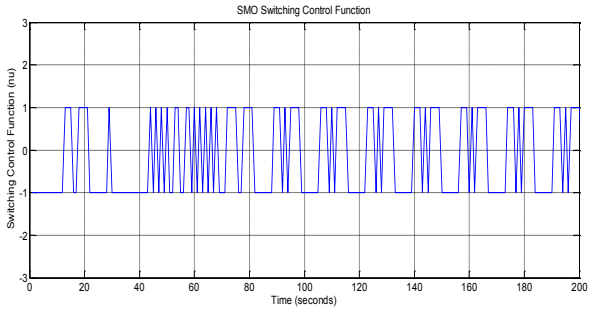


Fig. 12 SMO control switching function around sliding line

V. SLIDING MODE OBSERVER FOR LINEAR DC SERVOMOTOR WITH MODEL OR DISTURBANCE UNCERTAINTY

For the uncertain system (2) we define the following observer described in state-space representation [1]-[2], [4]-[5] by the following equations:

$$\frac{d\hat{x}}{dt} = A_{n\times n}\hat{x} + B_{n\times m}u - G_{L,n\times p}C_{p\times n}e(t) + G_{N,n\times p}\vartheta \quad (13)$$

where the error $e(t)$ is

$$e(t) = \hat{x}(t) - x(t) \quad (14)$$

ϑ is a discontinuous switching function about the sliding hyperplane:

$$S_w = \{e \in R^n | Ce = 0)\} \quad (15)$$

and the precise structure of injection signal matrices gains $G_{L,n\times p}, G_{N,n\times p}$ is to be determined.

The existence conditions for a SMO of the form (13) that rejects the uncertainty class $\Psi(x,u,t)$: $R_+ \times R^n \times R^m \to R^n$ of dc servomotor load torque $T_{Load} = 10sint(t)$ described in (2) are given in [1]-[2], [4]-[5]:

- $rank\,(C \times D) = q = 1,\ 1 = q \le p < n = 2$ (16)
- any invariant zeros of the original nominal system matrices (A, D, C) given in (2) must lies in the half left complex plane C_- (stable).

For a square nominal linear system these two conditions require that the triplet (A, D, C) to be relative degree one and minimum phase. Furthermore, the existence conditions depend upon a specific selection of the uncertainty channel and the observer design will be directly determined by the uncertainty distribution matrix D. The necessary and sufficient conditions for existence of a sliding mode observer (SMO) together with the canonical form provide a pathway to a constructive method for observer design [2], [4], [5].

The original nominal system (2) is converted in the canonical form (3) using a possible change in coordinates $z_c = T_c x$ through a canonical transformation matrix T_c given in (8):

$$A_c = T_c A T_c^{-1} = \begin{bmatrix} -2 & -0.1034 \\ 51.70 & -10 \end{bmatrix} = \begin{bmatrix} A_{11c} & A_{12c} \\ A_{21c} & A_{22c} \end{bmatrix}$$

$$B_c = T_c B = \begin{bmatrix} 2 \\ 0 \end{bmatrix} = \begin{bmatrix} B_{1c} \\ B_{2c} \end{bmatrix} \quad (17)$$

$$C_c = CT_c^{-1} = [0 \quad 1] = [C_{1c} \ \ C_{2c}]$$

$$D_c = T_c D = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} D_{1c} \\ D_{2c} \end{bmatrix}$$

In this structure all the required existence conditions mentioned in [2] to design a SMO are satisfied. The matrix gain $\bar{L}$ defined in [Sarah] for our case study becomes $\tilde{L} = \bar{L} = 0$.

Define now a new nonsingular matrix T_L of the following structure [2], [4], [5]:

$$T_L = \begin{bmatrix} I_{n-p} & \bar{L} \\ 0 & D_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad (18)$$

that converts the system from canonical form (17) in the following state-space representation:

$$z_L = T_L x_c$$

$$A_L = T_L A_c T_L^{-1} = A_c = \begin{bmatrix} -2 & -0.1034 \\ 51.70 & -10 \end{bmatrix}$$

$$A_L = \begin{bmatrix} A_{11,L} & A_{12,L} \\ A_{21,L} & A_{22,L} \end{bmatrix}$$

$$B_L = T_L B_c = \begin{bmatrix} 2 \\ 0 \end{bmatrix} = \begin{bmatrix} B_{1,L} \\ B_{2,L} \end{bmatrix}, \quad (19)$$

$$C_L = T_L^{-1} C_c = C_c = [0 \quad 1] = [C_{1,L} \ \ C_{2,L}]$$

$$D_L = T_L D_c = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} D_{1,L} \\ D_{2,L} \end{bmatrix},$$

and so, by chance, the same structure as in the canonical form.

The system triplet (A_L, C_L, D_L) can be put in the following form [1], [2]:

$$\frac{dz_{1,L}}{dt} = A_{11,L}z_{1,L}(t) + A_{12,L}z_{2,L}(t) + B_{1,L}u(t)$$

$$\frac{dz_{2,L}}{dt} = A_{21,L}z_{1,L}(t) + A_{22,L}z_{2,L}(t) + B_{2,L}u(t) + \cdots$$

$$D_{2,L}\Psi(x,u,t) \quad (20)$$

$$y(t) = C_{2,L}z_{2,L}(t) = z_{2,L}(t)\ ,$$

and so we can write also the following equivalent equation:

$$\frac{dy(t)}{dt} = A_{21,L}z_{1,L}(t) + A_{22,L}y(t) + B_{2,L}u(t) + \cdots$$

$$D_{2,L}\Psi(x,u,t) \quad (21)$$

The corresponding observer is described by the following equations:

$$\frac{d\hat{z}_{1,L}}{dt} = A_{11,L}\hat{z}_{1,L}(t) + A_{12,L}\hat{y}(t) + B_{1,L}u(t) - A_{12,L}e_y(t)$$

$$\frac{d\hat{y}(t)}{dt} = A_{21,L}z_{1,L}(t) + A_{22,L}\hat{y}(t) + B_{2,L}u(t) - \cdots$$

$$\left(A_{22,L} - A_{22,L}^s\right)e_y(t) + \vartheta \quad (22)$$

$$e_{2,L}(t) = \hat{z}_{2,L}(t) - z_{2,L}(t) = \hat{y}\,(t) - y(t) = e_y(t)$$

where $A_{22,L}^s$ is a stable design matrix, let us to take it $A_{22,L}^s = -2$. Let us to consider also a symmetric positive definite matrix for $A_{22,L}^s$, $P_2 \in R^{p\times p}$, that is a unique solution of the Lyapunov equation:

$$(A_{22,L}^s)^T P_2 + P_2(A_{22,L}^s) = -Q_2 \quad (23)$$

with $Q_2 \in R^{p\times p}$ a symmetric positive definite design matrix. For a particular selection of matrix $Q_2 = 4$ we get:

$$-4P_2 = -4 \to P_2 = 1 \quad (24)$$

Now a robust observer design is well defined and can be described by the following equations:

$$\frac{d\hat{x}}{dt} = A_{L_{n\times n}}\hat{x} + B_{L_{n\times m}}u - G_{L\_L,n\times p}C_{L_{p\times n}}e(t) + G_{N\_L,n\times p}\vartheta$$

$$\hat{y}(t) = C_{L_{p\times n}}\hat{x}(t) \quad (26)$$

$$e(t) = \hat{x}(t) - x(t)$$

$$e_y(t) = \hat{y}(t) - y(t)$$

where the gain matrices have the particular form [1], [2],[4], [5]:

$$G_{L\_L,n\times p} = T_c^{-1}\begin{bmatrix} A_{12,L} \\ A_{22,L} - A_{22,L}^{s} \end{bmatrix} = \begin{bmatrix} -8 \\ -0.10348 \end{bmatrix}$$

$$G_{N\_L,n\times p} = ||D_2||\, T_c^{-1}\begin{bmatrix} 0 \\ I_p \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad (27)$$

Corresponding to a particular selection of canonical coordinates transform, $T_c = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$.

Also, the discontinuous switching function ϑ about the sliding hyperplane S_w (15) is given by:

$$\vartheta = \begin{Bmatrix} -\rho(t,y,u)||D_{2,L}|| \dfrac{P_2 e_y(t)}{||P_2 e_y(t)||} \Big| e_y(t) \neq 0 \\ 0 \Big| e_y(t) = 0 \end{Bmatrix} \quad (28)$$

Remark: A big advantage of this development in this formulation of the sliding mode observer (SMO) design framework is that there is no requirement for (A, C) to be observable [2].

On the state components the developed robust observer given in (26) can be written in the following form:

$$\frac{d\hat{x}_1}{dt} = -2\hat{x}_1(t) - 0.1034\hat{y}(t) + 2u(t) + 8e_y(t) - \cdots$$
$$\rho(t,y,u)sign(e_y(t)) \quad (29)$$
$$\frac{d\hat{y}(t)}{dt} = 51.70\hat{x}_1(t) - 10\hat{y}(t) + 0.1034e_y(t)$$

with the dynamical errors described by following equations:

$$\frac{de_1(t)}{dt} = A_{11,L}e_1(t) = -2e_1(t)$$
$$\frac{de_y(t)}{dt} = A_{21,L}e_1(t) + A_{22,L}^{s}e_y(t) + \vartheta - D_{2,L}\Psi(x,u,t)$$
$$= 51.70e_1(t) - 2e_y(t) + \vartheta - \Psi(x,u,t)$$
$$e_1(t) = \hat{x}_1(t) - \hat{x}_1(t) \quad (30)$$
$$e_y(t) = \hat{y}(t) - y(t)$$

where the uncertainty function $\rho(t,y,u)$ is bounded by:

$$\rho(t,y,u) \geq r||u(t)|| + \alpha(y,t) + \delta \geq 12r + 10sin(t) + \delta$$

For a particular selection according to [Sarah]:

$$r = 0.7 > 0, \delta = 1.6 > 0, \alpha(y,t) = -10\sin(t), \quad (31)$$

the equations become:

$$\frac{d\hat{x}_1}{dt} = -2\hat{x}_1(t) - 0.1034\hat{x}_2(t) + 2u(t) + 8e_y(t) -$$
$$\ldots - 20sign(e_y(t)) \quad (32)$$
$$\frac{d\hat{y}(t)}{dt} = 51.70\hat{x}_1(t) - 10\hat{y}(t) + 0.1034e_y(t)$$

where the scalar function $\rho(t,y,u)$ is bounded by:

$$\rho(t,y,u) \geq 20 \quad (33)$$

5.1 SIMULATION RESULTS

The dynamics errors of the Sliding mode Observer model (30) attached to the dc Servomotor actuator with disturbance uncertainty (the load torque, $T_{Load} = 10sint(t)$) are modeled in SIMULINK and shown in figure 13. Their dynamic evolution, i.e. armature current residual (e_1) and angular speed residual(e_y), is shown in figures 14 and 15.

The SMO control switching function around sliding line ϑ is calculated according to (28) and (31) and is shown in figure 16. The state-space representation of the robust Sliding Mode Observer model (26) in canonical form is represented also in SIMULINK, and the evolution of the both model and estimated states, namely the angular output speed $(y(t), \hat{y}(t))$ and the armature current $(x_1(t), \hat{x}_1(t))$ are shown in figures 17 and 18, for open-loop system for the same setting $u(t) = 12\,[V]$. From the last two figures we observe that after approximately 2 seconds, visually perfect replication of the true and estimated states is taking place. To demonstrate the robustness of the nonlinear observer tracking the output from the dc Servomotor when the initial conditions of the true states and observer states are deliberately set to different values. If the nonlinear component is removed by setting $\rho(t,y,u)$ to zero, the resulting Luenberger Observer [2] behaves as simple observer without sliding motion.

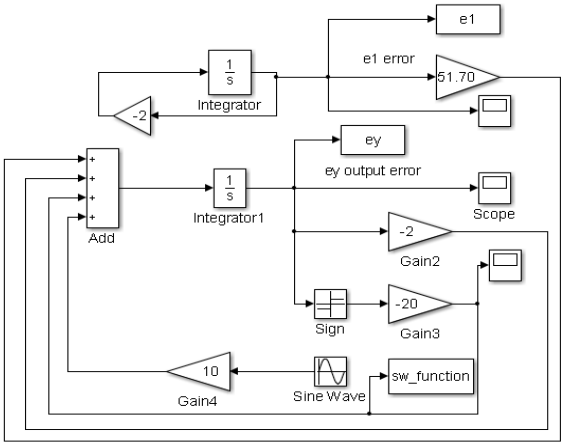


Fig. 13 DC servomotor SMO error dynamics with modeling uncertainty SIMULINK model

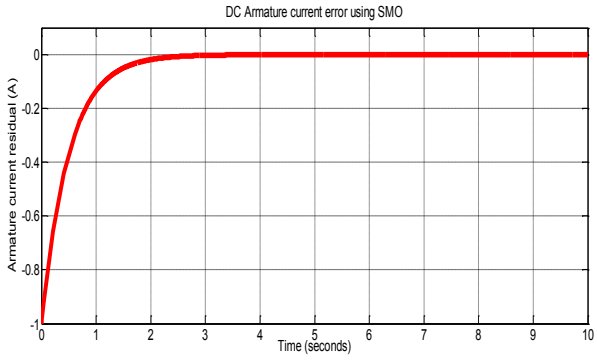


Fig. 14 DC servomotor armature residual current using SMO control strategy with modeling uncertainty

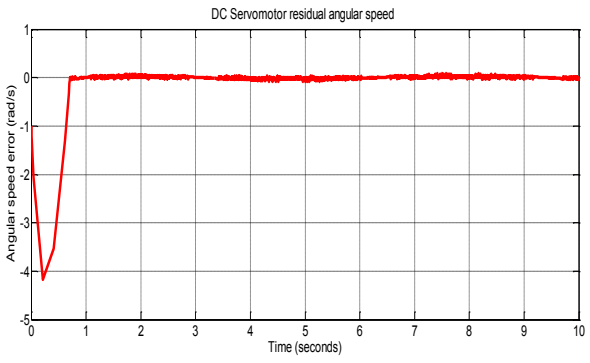


Fig. 15 DC servomotor angular residual speed using SMO control strategy with modeling uncertainty

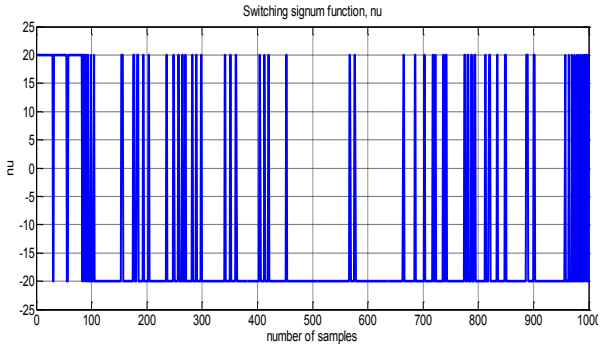


Fig. 16 SMO control switching function around sliding line

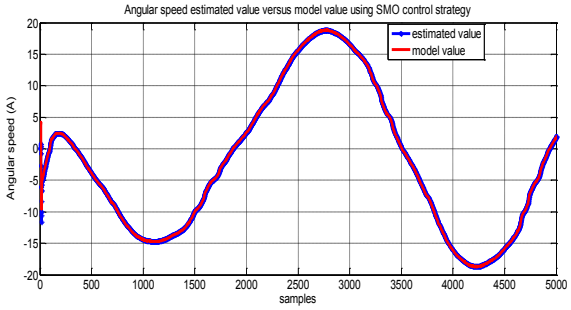


Fig. 17 DC servomotor angular speed estimated versus model angular speed using SMO control strategy with modeling uncertainty - MATLAB/SIMULINK simulations

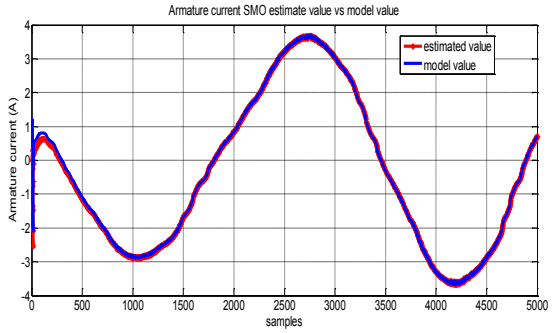


Fig. 18 DC servomotor armature current estimated versus model value using SMO control strategy with modeling uncertainty - MATLAB/SIMULINK simulations

VI. SLIDING MODE OBSERVER REAL-TIME IMPLEMENTATION

In control systems literature rarely we find details about the real-time software and hardware implementation aspects, and no sufficient attention is given about the algorithms and the sampling time selection. Usually the implementation aspect and real-time control systems design are connected together but in the most cases this connection is always ignored. Furthermore the real-time control systems design is treated from control perspective ignoring the implementation aspects of the control algorithms. Fortunately, recently the real-time implementation and design aspects get a considerable attention from part of control engineering community due to the introduction of new software tools like MATLAB/SIMULINK with its RTW (Real-Time Workshop) and the RTWT (Real-Time Windows Target) Toolboxes. The real-time platform used to perform these real-time simulations is a MATLAB R2013a with SIMULINK running on two processors WINDOWS OS machine. Certainly these new real-time platforms do the implementation of real-time experiments easier and save much time but on the other hand they have some drawbacks regarding a good perception of the real-life problems that could appear during the real-time implementation of the control systems.

VII. CONCLUSIONS

In this paper, we have studied the possibility of using a Sliding Mode Observer strategy design to a dc Servomotor actuator with disturbance uncertainty that is integrated in the same control system structure. The implementation in real time of SMO proposed control strategy will be very useful for our future developments in fault detection and isolation (FDI) control applications based on the equivalent signal injection principle [1]-[4]. This new FDI control strategy will be design in the future work based on the injection signal principle [1]-[2], [4]-[5]. The main contributions in our research are summarized briefly as follows:

(a) Comparison of performance capabilities and advantages of real-time implementation of a Sliding Mode Observer (SMO) versus Sliding Mode Control (SMC),
(b) Implementation in real time a Sliding Mode Observer of a linear dc Servomotor actuator without uncertainty,
(c) Implementation in real time a Sliding Mode Observer for a linear dc Servomotor actuator with bounded disturbance uncertainty.

REFERENCES

[1]. Sarah K. Spurgeon, “Sliding Mode Observers - historical background and basic introduction”, Spring School, Aussois, June 2015.
[2]. Sarah K. Spurgeon, “Sliding Mode Observers – toward a constructive design framework”, Spring School, Aussois, June 2015.
[3]. A. Aref, T. Tran, “Using Fuzzy Logic and Q-Learning for Trust Modeling in Multi-agent Systems”, Proceedings of the 2014 Federated Conference on Computer Science and Information Systems, Warsaw, Poland, ACSIS, Vol. 2, pp. 59–66, 2014, DOI: 10.15439/2014F482.
[4]. H. K. Khalil, L. Praly, “High-gain observers in nonlinear feedback control”, *Int. J. Robust. Nonlinear Control*, vol.24, issue 6, pp.993-1015, John Wiley & Sons, Ltd, 2013, DOI: 10.1002/rnc.3051.
[5]. X.G. Yan, C. Eduards, “Nonlinear robust fault reconstruction and estimation using a sliding mode observer”, Elsevier, ScienceDirect, Automatica, vol. 43, pp.1605 - 1614, 2007, DOI:10.1016/j.automatica.2007.02.008.
[6]. Amira Sayed A. Aziz, Ahmad Taher Azar , Mostafa A. Salama, Aboul Ella Hassanien, Sanaa El-Ola Hanafy, “Genetic Algorithm with Different Feature Selection Techniques for Anomaly Detectors Generation”, Proceedings of the 2013 Federated Conference on Computer Science and Information Systems, Krakow, Poland, pp. 769–774, 2013.
[7]. N.Tudoroiu, K. Khorasani, “Satellite Fault Diagnosis using a Bank of Interacting Kalman Filters”, *IEEE Transactions on Aerospace and Electronic Systems*, vol. 43, no.4, 2007, pp. 1334-1350, IEEE*Xplore*, DOI: 10.1109/TAES.2007.4441743.
[8]. E-R. Tudoroiu, “Conceiving and Implementing Applications using Real-Time UML”, PhD Thesis, Cluj-Napoca Technical University, Romania, 2012.
[9]. Control Tutorials for MATLAB, Carnegie Mellon Lab, University of Michigan, http://ctms.engin.umich.edu.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 849–852
DOI: 10.15439/2016F143
ACSIS, Vol. 8. ISSN 2300-5963

Toward adaptive heuristic video frames capturing and correction in real-time

Marcin Woźniak*, Dawid Połap*, Giacomo Capizzi‡, Grazia Lo Sciuto‡§
*Institute of Mathematics, Silesian University of Technology, Kaszubska 23, 44-100 Gliwice, Poland
Email: marcin.wozniak@polsl.pl, dawid.polap@gmail.com
‡Department of Electrical and Informatics Engineering, University of Catania, Viale A. Doria 6, 95125 Catania, Italy
Email: capizzi@dieei.unict.it
§Department of Electronic Engineering, University of Roma Tre, Via della Vasca Navale 84, 00146 Roma, Italy
Email: glosciuto@dii.unict.it

***Abstract*—Multimedia devices are widely used in professional applications as well as personal purposes. The use of computer vision systems enables detection and extraction of important features exposed in images. However constantly increasing demand for this type of video with high quality requires simple however reliable methods. The objective of presented research is to investigate applicability of heuristic method for real-time video frames capturing and correction.**

***Index Terms*—Video Stream Correction, Real-time Processing, Heuristic Method.**

I. Introduction

MULTIMEDIA devices are widely used in professional applications as well as personal purposes. We can find cameras applied in security systems present in CCTV (Closed Circuit TeleVision) applications where several cameras are detecting motion to supervise it against criminal actions and unlawful activities. Various types of these are used in financial institutions like banks, airports and any other railway/bus stations, sport stadiums and culture institutions like cinemas, theaters and concert places. Similarly to these, video recording is very useful for houses and car-parks where we use vision systems to prevent robbery. Recent years have also shown that modern video technologies with their various multimedia applications are also supported by international and national organizations and authorities which are exploring possibilities of implementations of computer aided analysis in order to assist i.e. law officers on duty. For these reasons it is paramount to develop automated systems that can actively improve precision of capturing in real time in various conditions, lightening and other factors that can actively influence CCTV.

Therefore in this article we want to discuss a new tracking model for image capturing, where dedicated version of heuristic attempt is applied to assist in detecting proper camera orientation in real time. Research presented in this work tend to move toward development of such a model.

A. Related Works

Image capturing and feature extraction efficiency in multimedia applications have been investigated in various research projects. Pope and Lowe proposed probabilistic approach to the problem of various 3-D object recognition [1], while Grycuk et al. proposed approach to use SURF for video key features detection for following frames [2]. With development in technology it became possible to evaluate more features with higher precision. Drozda et al. presented proposition of different orderings for visual sequences alignment implemented as algorithms for image classification [3]. Knop et al. discussed improvements in application of neural methods into video compression based on dedicated scene change detection algorithm [4], while Capizzi et al. proposed novel attempt to process images of oranges to be classified by proposed neural network architecture [5], Stateczny et al. discussed application of intelligent methods for image processing in batymetric systems [6]. Intelligent methods are also widely adapted into detection processes: Starzyk developed visual saccades for object recognition [7], Pabiasz et al. proposed three–dimensional facial landmarks recognition [8] and novel approach to 3D face images processing [9]. Similarly heuristic methods, as newly developed algorithms inspired by nature, gave new possibilities to multimedia streaming aspects. Panda et al. developed edge magnitude solution, where classic heuristic approach based on Cuckoo Search Algorithm was implemented to search over multilevel thresholding [10]. Mishra et al. proposed heuristic attempt to watermarking on gray-scale images by application of Firefly Algorithm [11]. Heuristic methods are extensively examined in recent years, and various new or hybrid methods are developed for multimedia applications in detection and image capturing systems. Woźniak and Połap presented extensive comparison of efficiency in key-points extraction between developed Cuckoo Search Algorithm and classic methods like SIFT and SURF [2], [12]. Similarly Firefly Algorithm application was proposed by Woźniak and Marszałek [13]. Walendzik et al. reported development in gaming technologies for automatically generated evaluation [14] and Swiechowski et al. discussed self adapting strategies to gaming reality [15]. Decision making systems widely use adaptive strategies to simulate intelligent data streaming processing as reported by Rutkowski et al. [16].

Multimedia processing by possible applications of various methods of Computational Intelligence started important trend in nowadays technology. Multimedia storing systems are applied to manage visual information [17]. Korytkowski et al.

proposed boosted fuzzy classifiers for captured images [18].

In this article we propose novel approach to implement heuristic method to work as real-time detector for cameras and vision systems.

II. Proposed Frame Processing Technique

Computational Intelligence (CI) methods widely use various soft computing techniques for detection and extraction of features. One of these processes is frame capturing. It is non-trivial to preform this operation in real-time along with correction of quality. Proposed in this article technique is based on application of CI method, in particular dedicated versions of bio-inspired heuristic approach, as a dedicated solution to improve vision tracking in CCTV systems or any image recognition systems that use this or similar type of vision capturing.

In this type of bio-inspired simulation approach we can adapt birds, fish or any other species which together as a cluster behave in a very specific way. The population adapts to given initial conditions of the environment following to the destination. This type type of behavior is very useful in various applications, i.e. where we want to search object space for specific features.

A. Ad-hoc Filtering Method

Before application of heuristic detection we need to extract features from video frames. These features will serve as objects, which will be traced along rotation of the camera. To extract features we have applied simplified filtering method, which is run ad-hoc to filter video frames and extract objects to be forwarded to heuristic tracking method. In proposed filtering we introduce evaluation of the luminosity to extract edges of the traced objects. Extraction leaves pixels of high luminosity which extract shapes as bright pixels over dark background. We simply approximate both dimensions of the luminance gradient along exes of the video frame.

1) Applied Operator: Extraction of objects is based on idea to use directional differential operator $\vec{\nabla}$ on brightness ϕ as:

$$\vec{\nabla}\phi \cdot d\mathbf{x_i} = [\partial_1\phi, \partial_2\phi] \cdot [dx_{i,1}, dx_{i,2}]\,, \tag{1}$$

where partial derivatives $\partial_1\phi$ and $\partial_2\phi$ are computed for video frame points $\mathbf{x_i}$ according to each coordinate.

Proposed ad-hoc frame filtering is using luminance intensity matrix L to compute function

$$\begin{cases} \tilde{\phi}(\mathbf{x_i}) = \sqrt{\sum_{k=1}^{2} \phi_k^2(\mathbf{x_i})} \\ \phi_k(\mathbf{x_i}) = \sum_{m,n=1}^{3} \max_{k=1,2}(M_{mn}^k \cdot L(x_{i,1}+m-2, x_{i,2}+n-2)) \end{cases} \tag{2}$$

as convolution of matrices

$$\begin{cases} M^1 = \begin{pmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{pmatrix} \\ M^2 = \begin{pmatrix} -1 & 0 & -1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix} \end{cases}. \tag{3}$$

Matrix M^1 is applied for vertical extraction and matrix M^2 is applied for horizontal extraction, however application of $\max_{k=1,2}(\cdot)$ returns only highest convolution value to the output bit of video frame points $\mathbf{x_i}$ to enable faster shape features extraction. Composed in this way filtered frame is forwarded to applied heuristic method for tacking.

B. Proposed Bio-Inspired Heuristic Approach

The main idea of bio-inspired heuristic approach is to simulate entire population of mapped organisms into implemented algorithm. During iterations we assume that individuals can exchange information to find destination. This make the implemented population act similarly to swarms of fish, birds or other species. This assumptions are composed into mathematical model, where destination of the swarm in each iteration is the object traced by the CCTV camera. The algorithm is implemented to search the following video frames for destination objects by matching trajectories of individuals (particles) and therefore trace the object in real-time. Each of individuals is a vector of coordinates that move along the rotation of the camera tracing the object.

Movement of tracing individuals is based on stochastic and deterministic approach, where we combine random walk toward optimum with deterministic distance between particles. The knowledge about traced object is updated in each iteration according to positions of particles that correctly detected traced object. This information serves as a staring location for further iteration, where particles compare new situation to previous frame and follow the best situated individual to the destination.

1) Applied Model: To keep the randomness of movements along with so called "communications between particles" we introduce deterministic and random factors along with the following assumptions:

- Tracing points are moving along the captured video frames in search of the object,
- Each individual is referring to it previous position while tracing the object,
- At the end of each iteration, all the individuals exchange information,
- Number of tracing individuals is constant.

Each tracing individual position is denoted as $\mathbf{x}_i^t$ whose i components correspond to dimensions of the video frame and t is iteration in the algorithm. Move is denoted as $\mathbf{m}_i^t$ with appropriate symbols for each iteration t according to the formula:

$$\mathbf{m}_i^{t+1} = \mathbf{m}_i^t + \alpha \cdot \epsilon_1 \cdot [g_*^{t-1} - f(\mathbf{x}_i^t)] + \beta \cdot \epsilon_2 \cdot [\mathbf{x}_*^t - \mathbf{x}_i^t], \tag{4}$$

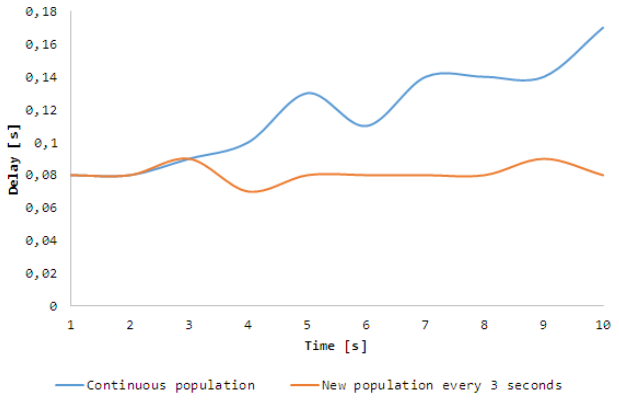


Fig. 1. Image capturing process in real-time improvements in two examined solutions: blue line for continuous usage of the same population of $tracing\_individuals$, orange line for new population of $tracing\_individuals$ introduce in each 3 sec.

where the symbols are: $\mathbf{m}_i^t$ – tracing move of i individual in t iteration, α – optimum value memory factor, β – optimum individual position memory factor, $\epsilon_1, \epsilon_2 \in [0, 1]$ – random values, g_*^t – previous position of the tracing individual at the frame in $t-1$ iteration, $f(\mathbf{x}_i^t)$ – present position of the tracing individual at the frame in t iteration, $\mathbf{x}_*^t$ – position of best situated individual in t iteration, $\mathbf{x}_i^t$ – position of tracing individual i in t iteration.

For that modeled trace move we perform movements of all tracing individuals using formula:

$$\mathbf{x}_i^{t+1} = \mathbf{x}_i^t + (-1)^K \cdot \mathbf{m}_i^t, \tag{5}$$

where the symbols are: $\mathbf{x}_i^t$ – position of tracing individual i in t iteration, $\mathbf{m}_i^t$ – trace move i particle in t iteration according to (4), K–random factor applied to randomize direction of movements.

Equations (4) and (5) allow trafing of objects in real-time using implementation of the proposed Algorithm 1. To start tracing we place initial population of individuals at random over first video frame. While camera is rotating along the axis implemented However, to improve tracing abilities we can also apply some boundary criteria to enable additional movements control.

III. Experimental Results

In the experimental tests we have applied two sample video streams. First was captured at one of polish parks. Second was captured in egyptian pyramid. The task for proposed system was to follow rotation of the camera to trace the object in real-time and therefore improve quality of video recording. Based on tests, the maximum displacement of particles between two frames has been appointed as 5 pixels. With this value, the amount of calculations in the proposed algorithm is significantly minimized - in the last stages of the algorithm only circles of radius equal to 5 are analyzed. Results of proposed real-time heuristic tracking are presented in Fig. 2. Chart of the delay for both solutions is presented in Fig. 1.

Algorithm 1 Heuristic Approach to Video Frames Processing in Real-Time

Define coefficients: α – memory factor, β – position memory factor, $generation$ – number of iterations, $tracing\_individuals$ – number of individuals in swarm,
while video frames are captured from rotating CCTV camera **do**
  Capture 2 following video frames with a delay of 1 sec,
  Perform Ad-hoc Filtering Algorithm on each of them,
  Start tracing using first video frame,
  Create at random initial population,
  t:=0,
  while $t \leq generations$ **do**
    Move $tracing\_individuals$ according to (5) and (4),
    Sort $tracing\_individuals$ according to brightness,
    Evaluate $tracing\_individuals$ and take $best\_ratio$ of them to next $generation$,
    Rest of $tracing\_individuals$ take at random,
    Next $generation$: $t++$,
  end while
  Place $tracing\_individuals$ from last $generation$ over second filtered video frame,
  Divide this frame into 4 parts,
  Take this part where we have highest concentration of $tracing\_individuals$,
  for $-5 \leq \alpha \leq 5$ **do**
    for $-5 \leq \beta \leq 5$ **do**
      Move $tracing\_individuals$ using correction (α, β),
      Calculate percentage of all points whose adaptation is the same as for the first frame,
      Save point for which the percentage is highest,
      $\beta++$,
    end for
    $\alpha++$,
  end for
  Determine the direction on the basis of selected point.
end while

A. Conclusions

In the experimental tests we have compared two attempts for proposed solution: to continue heuristic processing using only one population of $tracing\_individuals$ and to change population in each 3 seconds. Comparing results of benchmark tests we can see that introduction of new population in regular intervals can increase efficiency of the real-time processing. Chart presented in Fig. 1 show relation of delay between two compared solutions. With increasing time of video processing newly introduced population tends to reduce delay what influence efficiency of tracking, therefore proposed processing becomes faster and even more adapted for CCTV real-time video systems.

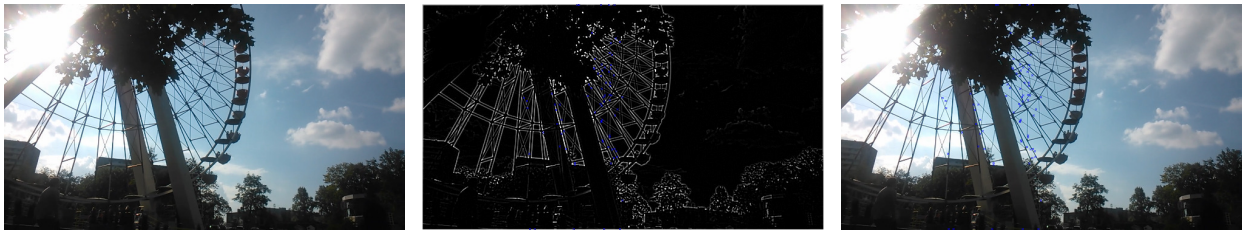


Fig. 2. Image capturing process of wheel in in real-time improvements. From left to right: original frame, filtered with heuristic points, original with heuristic points.

IV. Final Remarks

Proposed solution enabled us to obtain the direction of camera movement at the time of recording. The proposed method uses heuristic processing and ad-hoc filtering. Due to simple implementation and low number of operations it is possible to perform it all in real-time during video recording, when CCTV system is already loaded. The results, i.e. the average delay is constant over time, allow for practical use in various applications such as sport equipment or systems of the virtual view. Moreover, such a solution in conjunction with the navigation system GPS can create real guidance system.

In future work, it is planned to reduce the calculations and the inclusion of additional factors such as navigation system in order to create an easy-to-use solutions for a variety of purposes.

Acknowledgment

Authors acknowledge contribution to this project of Operational Programme: Knowledge, Education, Development financed by the European Social Fund under grant application POWR.03.03.00-00-P001/15.

References

[1] A. Pope and D. Lowe, "Probabilistic models of appearance for 3-d object recognition," *International Journal of Computer Vision*, vol. 40, no. 2, pp. 149–167, 1998.

[2] R. Grycuk, M. Knop, and S. Mandal, "Video key frame detection based on SURF algorithm," *Lecture Notes in Artificial Intelligence - ICAISC'2015*, vol. 9119, pp. 566–576, 2015, DOI: 10.1007/978-3-319-19324-3.

[3] P. Drozda, K. Sopyla, and P. Górecki, "Different orderings and visual sequence alignment algorithms for image classification," *Lecture Notes in Artificial Intelligence - ICAISC'2014*, vol. 8467, pp. 693–702, 2014, DOI: 10.1007/978-3-319-07173-2.

[4] M. Knop, T. Kapuscinski, W. K. Mleczko, and R. A. Angryk, "Neural video compression based on RBM scene change detection algorithm," *Lecture Notes in Artificial Intelligence - ICAISC'2016*, vol. 9693, pp. 660–669, 2016, DOI: 10.1007/978-3-319-39384-1_58.

[5] G. Capizzi, G. Lo Sciuto, C. Napoli, E. Tramontana, and M. Woźniak, "Automatic classification of the fruit defects based on co-occurrence matrix and neural networks," in *Proceedings of the Federated Conference on Computer Science and Information Systems - FedCSIS'2015*. 13-16 September, Lodz, Poland: IEEE, 2015, pp. 861–867, DOI: 10.15439/2015F258.

[6] A. Stateczny, M. Wlodarczyk-Sielicka, and G. Zaniewicz, "Different orderings and visual sequence alignment algorithms for image classification," *Annual of Navigation*, vol. 19, no. 2, pp. 99–108, 2012, DOI: 10.2478/v10367-012-0020-x.

[7] J. A. Starzyk, "Visual saccades for object recognition," *Lecture Notes in Artificial Intelligence - ICAISC'2015*, vol. 9119, pp. 778–788, 2015, dOI: 10.1007/978-3-319-19324-3_70.

[8] S. Pabiasz, J. T. Starczewski, and A. Marvuglia, "A new three-dimensional facial landmarks in recognition," *Lecture Notes in Artificial Intelligence - ICAISC'2014*, vol. 8468, pp. 179–186, 2014, DOI: 10.1007/978-3-319-07176-3_16.

[9] S. Pabiasz, J. T. Starczewski, and A. Marvuglia, "SOM vs FCM vs PCA in 3d face recognition," *Lecture Notes in Artificial Intelligence - ICAISC'2015*, vol. 9120, pp. 120–129, 2015, DOI: 10.1007/978-3-319-19369-4_12.

[10] R. Panda, S. Agrawal, and S. Bhuyan, "Edge magnitude based multilevel thresholding using cuckoo search technique," *Expert Systems with Applications*, vol. 40, no. 18, pp. 7617–7628, 2013.

[11] A. Mishra, C. Agarwal, A. Sharma, and P. Bedi, "Optimized gray-scale image watermarking using dwt svd and firefly algorithm," *Expert Systems with Applications*, vol. 41, no. 17, pp. 7858–7867, 2014.

[12] M. Woźniak and D. Połap, "Basic concept of cuckoo search algorithm for 2D images processing with some research results : An idea to apply cuckoo search algorithm in 2d images key-points search," in *SIGMAP 2014 - Proceedings of the 11th International Conference on Signal Processing and Multimedia Applications, Part of ICETE 2014 - 11th International Joint Conference on e-Business and Telecommunications*. 28-30 August, Vienna, Austria: SciTePress, 2014, pp. 157–164, DOI: 10.5220/0005015801570164.

[13] M. Woźniak and Z. Marszałek, "An idea to apply firefly algorithm in 2D images key-points search," *Communications in Computer and Information Science - ICIST'2014*, vol. 465, pp. 312–323, 2014, DOI: 10.1007/978-3-319-11958-8_25.

[14] K. Waledzik and J. Mandziuk, "An automatically generated evaluation function in general game playing," *IEEE Trans. Comput. Intellig. and AI in Games*, vol. 6, no. 3, pp. 258–270, 2014, DOI: 10.1109/TCIAIG.2013.2286825.

[15] M. Swiechowski and J. Mandziuk, "Self-adaptation of playing strategies in general game playing," *IEEE Trans. Comput. Intellig. and AI in Games*, vol. 6, no. 4, pp. 367–381, 2014, DOI: 10.1109/TCIAIG.2013.2275163.

[16] L. Rutkowski, M. Jaworski, L. Pietruczuk, and P. Duda, "A new method for data stream mining based on the misclassification error," *IEEE Trans. Neural Netw. Learning Syst.*, vol. 26, no. 5, pp. 1048–1059, 2015, DOI: 10.1109/TNNLS.2014.2333557.

[17] R. Grycuk, M. Gabryel, R. Scherer, and S. Voloshynovskiy, "Multi-layer architecture for storing visual data based on WCF and microsoft SQL server database," 2015, pp. 715–726, DOI: 10.1007/978-3-319-19324-3_64.

[18] M. Korytkowski, L. Rutkowski, and R. Scherer, "Fast image classification by boosting fuzzy classifiers," *Information Sciences*, vol. 327, pp. 175–182, 2016, DOI: 10.1016/j.ins.2015.08.030.

8th Workshop on Scalable Computing

THE Workshop on Scale Computing (WSC) is a result of evolution in the world of computing. It originated (as Workshop on Large Scale Computing in Grids; LaSCoG) in 2005. Next, cloud computing became popular and, in response to this new trend, Workshop on Scalable Computing in Distributed Systems (SCoDiS) emerged. The two workshops (under a joint name LaSCoG-SCoDiS) have been organized till 2014 (information about past events can be found here). However, the world of large-scale computing continuously evolves. In particular, data-intensive computations (known as “Big Data") brough a completely new set of issues that have to be solved (in addition to those that exist since late 1990th and that still deserve our attention). Therefore we have decided to refresh the name of the event (to better represent the scope of interest). This is how the Workshop on Scalable Computing (WSC) came to being.

TOPICS

- General issues in scalable computing
  - Algorithms and programming models for large-scale applications, simulations and systems
  - Large-scale symbolic, numeric, data-intensive, graph, distributed computations
  - Architectures for large-scale computations (GPUs, accelerators, quantum systems, federated systems, etc.)
  - Data models for large-scale applications, simulations and systems
  - Large-scale distributed databases
  - Security issues for large-scale applications and systems
  - Load-balancing / intelligent resource management in large-scale applications, simulations and systems
  - Performance analysis, evaluation and prediction
  - Portals, workflows, services and collaborative research
  - Data visualization
  - On-demand computing
  - Virtualization supporting computations
  - Self-adaptive computational / storage systems
  - Volunteer computing
  - Scaling applications from small-scale to exa-scale (and back)
  - Computing for Big Data
  - Business applications
- Grid / Cloud computing
  - Cloud / Grid computing architectures, models, algorithms and applications
  - Cloud / Grid security, privacy, confidentiality and compliance
  - Mobile Cloud computing
  - High performance Cloud computing
  - Green Cloud computing
  - Performance, capacity management and monitoring of Cloud / Grid configuration
  - Cloud / Grid interoperability and portability
  - Cloud / Grid application scalability and availability
  - Economic, business and ROI models for Cloud / Grid computing
  - Big Data cloud services

EVENT CHAIRS

- **Ganzha, Maria,** University of Gdańsk and Systems Research Institute Polish Academy of Sciences, Poland
- **Gusev, Marjan,** University Sts Cyril and Methodius, Macedonia
- **Paprzycki, Marcin,** Systems Research Institute Polish Academy of Sciences, Poland
- **Petcu, Dana,** West University of Timisoara, Romania
- **Ristov, Sashko,** University Sts Cyril and Methodius, Macedonia

PROGRAM COMMITTEE

- **Anderson, David,** University of California, Berkeley, United States
- **Bass, Len,** NICTA, Australia
- **Brodnik, Andrej,** University of Ljubljana, Faculty of Computer and Information Science, Slovenia
- **Camacho, David,** Universidad Autonoma de Madrid, Spain
- **D’Ambra, Pasqua,** ICAR-CNR, Italy
- **Filippone, Salvatore,** University Rome Tor Vergata, Italy
- **Gepner, Paweł,** Intel Corporation, United Kingdom
- **Gordon, Minor,** Software development consultant, United States
- **Goscinski, Andrzej,** Deakin University, Australia
- **Gravvanis, George,** Democritus University of Thrace, Greece
- **Grosu, Daniel,** Wayne State University, United States
- **Holmes, Violeta,** The University of Huddersfield, United Kingdom
- **Hsu, Ching-Hsien (Robert),** Chung Hua University, Taiwan
- **Kalinov, Alexey,** Cadence Design Systems, Russia
- **Karaivanova, Aneta,** IICT-BAS, Bulgaria
- **Kitowski, Jacek,** AGH University of Science and Technology, Department of Computer Science, Poland

- **Knepper, Richard,** Indiana University, United States
- **Kranzlmüller, Dieter,** Ludwig-Maximilians-Universität München (LMU), Germany
- **Kwiatkowski, Jan,** Wrocław University of Technology, Poland
- **Lang, Tran Van,** Vietnam Academy of Science and Technology, Vietnam
- **Lastovetsky, Alexey,** University College Dublin, Ireland
- **Legalov, Alexander,** Siberian Federal University, Russia
- **Luo, Mon-Yen,** National Kaohsiung University of Applied Sciences, Taiwan
- **Margaritis, Konstantinos G.,** University of Macedonia, Greece
- **Milentijevic, Ivan,** University of Nis, Serbia
- **Morrison, John,** University College Cork, Ireland
- **Nosovic, Novica,** Faculty of Electrical Engineering, University of Sarajevo, Bosnia and Herzegovina
- **Olejnik, Richard,** CNRS - University of Lille I, France
- **Ouedraogo, Moussa,** Public Research Centre Henri Tudor, Luxembourg
- **Rak, Massimiliano,** Seconda Università di Napoli, Italy
- **Schikuta, Erich,** University of Vienna, Austria
- **Schreiner, Wolfgang,** Johannes Kepler University Linz, Austria
- **Shen, Hong,** University of Adelaide, Australia
- **Song, Ha Yoon,** Hongik University, South Korea
- **Stankovski, Vlado,** University of Ljubljana, Slovenia
- **Talia, Domenico,** University of Calabria, Italy
- **Telegin, Pavel,** JSCC RAS, Russia
- **Trystram, Denis,** Grenoble Technical University, France
- **Tudruj, Marek,** Inst. of Comp. Science Polish Academy of Sciences/Polish-Japanese Institute of Information Technology, Poland
- **Tvrdik, Pavel,** Faculty of Information Technology, Czech Technical University in Prague, Czech Republic
- **Vazhenin, Alexander,** University of Aizu, Japan
- **Wei, Wei,** School of Computer science and engineering, Xi'an University of Technology, China
- **Wyrzykowski, Roman,** Czestochowa University of Technology, Poland
- **Xu, Baomin,** Beijing JiaoTong University, China
- **Zavoral, Filip,** Charles University in Prague, Czech Republic

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 855–864
DOI: 10.15439/2016F308
ACSIS, Vol. 8. ISSN 2300-5963

Modeling energy consumption of parallel applications

Paweł Czarnul, Jarosław Kuchta, Paweł Rościszewski
Faculty of Electronics, Telecommunications and Informatics
Gdansk University of Technology
Narutowicza 11/12, 80-233 Gdansk, Poland
Email: {pczarnul,qhta}@eti.pg.gda.pl, pawel.rosciszewski@pg.gda.pl

Jerzy Proficz
Academic Computer Center
Gdansk University of Technology
Narutowicza 11/12, 80-233 Gdansk, Poland
Email: jerp@task.gda.pl

***Abstract*—The paper presents modeling and simulation of energy consumption of two types of parallel applications: geometric Single Program Multiple Data (SPMD) and divide-and-conquer (DAC). Simulation is performed in a new MERPSYS (Modeling Efficiency, Reliability and Power consumption of multilevel parallel HPC SYStems using CPUs and GPUs) environment. Model of an application uses the Java language with extensions representing message exchange between processes working in parallel. Simulation is performed by running threads representing distinct process codes of an application, with consideration of process counts. Instead of running time consuming calculations, their times are simulated using functions representing computational time dependent on input data sizes. The simulator considers performance and power consumption values for compute devices stored in its database. We performed verification of running the two applications on up to 512 and 1024 processes respectively on a large cluster from Academic Computer Center in Gdansk demonstrating a high degree of accuracy between simulated and measured results.**

I. Introduction

IN TODAY'S High Performance Computing (HPC) landscape performance and power consumption are key factors, both of which are of key concerns in design of future systems. As of today, Tianhe-2 is the most powerful computing cluster on the TOP500 list with performance of over 33 PFlop/s at 17.8 MWs of power consumption. Tianhe-2 uses the hybrid architecture that couples multicore CPUs and accelerators within a single node. Examples of accelerators used today are GPUs or coprocessors such as Intel Xeon Phi. These are used in the top high performance clusters listed on the TOP500 list. The recently announced Tesla P100 offers 5.3 TFlop/s double-precision performance with Thermal Design Power (TDP) 300 Watts[1]. Intel® Xeon Phi™ Coprocessor 7120P offers 1.2 TFlop/s theoretical peak double-precision performance[2] with TDP 300 Watts[3].

As computational power of HPC systems comes from engaging more and more processing cores and consequently increasing the sizes of compute devices and the number of compute devices within a system, there is often a need for assessment of not only performance but also power consumption of such systems. A typical use case is when the user or system owner already know several applications that are run in their contexts or environments and need to assess performance and power consumption of an HPC system after an upgrade or after a new HPC system is to be purchased.

This paper focuses on a model and methodology for assessment of power and energy consumption of parallel applications adopted in the MERPSYS simulation environment[4]. This work follows modeling execution time of parallel applications in MERPSYS that is presented in [1].

II. Related Work

In terms of applications, energy consumption and its reduction is very important. Proper techniques involving load shifting and machine management may result in energy bill savings [2]. Paper [3] analyzes optimization of energy consumption for large virtualized service centers.

In work [4], authors present a workflow that allows prediction of energy and power consumption of HPC applications using available data for a given application regarding power and energy consumption for specific values of nodes used. Then, based on a predictor, that uses the available data and proper interpolation, predicted values can be found. The paper shows a high degree of accuracy of the predictor for Hydro (computational fluid dynamics) and EPOCH (plasma physics simulation) benchmarks executed on the SuperMUC (near Munich, hence MUC) HPC platform.

In paper [5], experiments with Co-Design Molecular Dynamics (CoMD) and Livermore Unstructured Lagrangian Explicit Shock Hydrodynamics (LULESH) codes were performed on a system with host Xeon E5 CPUs and Xeon Phi 5110P coprocessors with measurements of energy and power for the whole system, CPUs and Xeon Phis. Results were used to obtain parameters of theoretical model coefficients with high confidence (R^2 coefficient). Results were presented for host frequency scaling as well as problem size scaling.

In paper [6] authors used neural networks to train models that predicted power and energy consumption when running high performance computing codes. It has been shown that

[1]http://wccftech.com/nvidia-pascal-gpu-gtc-2016/

[2]http://www.intel.com/content/www/us/en/benchmarks/server/xeon-phi/xeon-phi-theoretical-maximums.html

[3]http://ark.intel.com/products/75799/Intel-Xeon-Phi-Coprocessor-7120P-16GB-1_238-GHz-61-core

[4]http://merpsys.eti.pg.gda.pl/

after training, using various versions of codes, it is possible to predict power consumption and energy usage of CPUs and DIMMs with less that 5.5% error for LU factorization, Jacobi and matrix multiplication.

In work [7] authors have presented a detailed energy usage model for parallel master-slave applications, including modeling energy consumption of communication operations, based on execution times. Furthermore, the model was verified in a real environment with a master and 4 or 6 slaves for single or dual core configurations with error rate lower than 4% across the tested configurations.

In paper [8] authors investigated execution times and energy used when running MPI-only and hybrid MPI with OpenMP codes such as Parallel Multiblock Lattice Boltzmann or Gyrokinetic Toroidal Code. In particular, on the largest configurations tested with 8 nodes and a total of 32 cores, hybrid versions showed better execution times and energy consumption than MPI-only codes. Energy used was broken into CPU, memory, disk and motherboard energy.

In paper [9] authors, following analysis performed for Amdahl's law, present a general energy speed-up model in a parallel environment, for multicore systems. Furthermore, authors present specific results for three various power consumption models for a multicore CPU, based on the number of cores used: in the first one all cores are always on, the second with consideration of active cores only and the third with base power, active and idle core power values.

Modeling power consumption of cluster nodes depending on the number of threads active with verification against real measurements were presented in [10]. This showed an idle system power consumption and a non-linear increase until a saturation point. Such a model has been incorporated into the MERPSYS simulator. In work [1], modeling and verification of performance of parallel applications in MERPSYS was presented for computation of vector similarities along with verification in a real parallel environment.

For some types of applications, such as embarrassingly parallel ones, volunteer computing may be an alternative to clusters. Clusters and volunteer systems are different in terms of locality (centralized vs distributed), payer of infrastructure and electrical bill cost, involvement (or lack thereof) of society, security. Comparison of performance and power consumption as well as computational efficiency of cluster based systems and volunteer based systems which use distributed volunteers' computers is presented in [11]. For the latter, sets of volunteer hardware configurations were taken from BOINC projects and `http://cpubenchmark.net/` benchmarks for relevant CPUs and TDPs were used. On average performance/power consumption ratio for modern CPU based clusters turned out to be 2-3 times better than for machines in volunteer based systems.

In [12] authors statistically analyze average CPU utilization and draw a conclusion that in the typical operating region of between 10 and 50% of utilization, energy efficiency is low and aim at designing energy proportional machines that would consume energy proportional to the executed work.

Modelling energy consumption of distributed systems can be useful for exploring the time-energy trade-off, defined in [13]. The authors consider the relationships between execution time, energy consumption and power draw for a set of chosen applications, both on shared memory devices such as Intel Xeon Phi coprocessor and Intel Xeon processor, as well as the Vesta IBM Blue Gene/Q cluster. Formal formulation of the multi-objective code optimization problem is presented, as well as evidence that the energy-performance trade-offs exist in practice.

Paper [14] analyzes energy and makespan trade-offs as a Pareto front in heterogeneous computing systems. Pareto fronts for the multi-objective optimization problem can be determined using mathematical modeling and linear programming [15]. However, such model has to closely match the characteristics of the real executions, which can significantly vary depending on the application model (i.e. synchronization scheme, communication overlapping). Additionally, the model may require defining the execution times of each type of task on each type of hardware beforehand. Thus, for more accurate modeling of various application executions on various hardware, it is important to develop a more flexible method which can give an approximate result with a possibility to quickly modify the application and hardware models.

Paper [16] considers tuning of application execution by proper tiling in the code (cache usage) and CPU frequency. It considers impact on the execution time and energy usage using an example of Poisson's equation with stencil computations.

In work [17] a methodology and experiments were presented for a distributed KernelHive [18] system that is aimed at parallelization of computations in a heterogeneous environment that consists of potentially several clusters each with multicore CPUs and accelerators such as GPUs. Based on an imposed power consumption limit, an optimizer is able to select compute devices such that the total power consumption does not exceed the threshold and execution time is minimized taking into consideration application configuration (including OpenCL's kernel NDRange configurations for GPUs and CPUs), network parameters etc. Dependence of execution times against power consumption limits were shown for a real environment.

As demonstrated in paper [4], a model for prediction of power and energy usage in an HPC system can potentially be very desirable e.g. for budget estimation and prediction of peak requirements in terms of power consumption.

In work [19] authors presented Energy Efficient Task Duplication Schedule (EETDS) algorithm with a grouping and energy efficient group allocation schedule phases of a DAGs (Directed Acyclic Graphs) onto a parallel environment. The algorithm was compared, in terms of energy consumption, to Task Duplication Scheduling algorithm (TDS), Non-Duplication Scheduling algorithm (NDS) and Energy-Efficient Non-Duplication Scheduling (EENDS) strategies for Gaussian Elimination and FFT for various values of communication to computation ratio (CCR) demonstrating benefits of EETDS for larger CCR values.

In paper [20] optimization of hybrid MPI/OpenMP parallel application execution is considered in terms of execution efficiency. Algorithms used consider Dynamic Concurrency Throttling (DCT) and Dynamic Voltage Frequency Scaling (DVFS), also in a combined setting. It is demonstrated that the proposed approach results in savings in energy usage with little loss in performance or even gains.

III. Motivations and Problem Statement

Motivations for simulation of execution of parallel applications on large systems stated for the MERPSYS environment, involving execution time and energy consumption, include:

1) Finding good configurations for running parallel applications i.e. specific compute devices in the available environment, numbers of nodes as well as application parameters such as data packet sizes etc.
2) Testing various potential (e.g. from a database of available components) hardware configurations for running a set of applications. MERPSYS allows instant substitution of one component by another e.g. exchanging a CPU or a GPU for another CPU or GPU model, similarly for interconnects.
3) Simulations of a set of applications in a distributed multi-level system composed of clusters and volunteer based systems in order to find approximately optimal hardware allocation, task mapping and scheduling.

In view of the aforementioned works and challenges, *the goal of this paper is to define and verify a model of energy consumption of a parallel application run on a parallel system that would return acceptably accurate results from fast running simulations of parallel runs*. Specifically, this requires finding the following function

$$\text{energy consumption}(\text{parallel application}, \text{parallel system}, \text{input data})$$

It should be noted that there are two possible ways of how energy consumption is calculated. In one, within the makespan of the application only energy used for duration of computations on particular nodes, only when these are used by the application, is accounted for. In the latter, energy of all nodes is integrated over the makespan irrespective of how many application processes/threads run there, considering idle energy consumption if none processes/threads are active. MERPSYS adopted the second method.

In essence, the function mentioned above can be expressed in terms hardware count, thread count, time of effective application execution (stress time) and time of ineffective processor work (idle time) as follows:

$$\begin{aligned}\text{energy consumption}(\text{parallel application}, \text{parallel system}, \\ \text{input data}) = \\ \sum_{i=1}^{Hardware_{count}} (t_{application} PW[i]_{idle} + \\ \sum_k t_{exec}[i,k](PW[i]_{stress}(threadcount[i,k]) - \\ PW[i]_{idle}))\end{aligned}$$

which considers hardware used and power consumption in idle state multiplied by execution time as well as additional power consumption under stress when running a given number of threads on particular hardware multiplied by activity period.

IV. Modeling Energy Consumption

We modeled energy consumption in a supercomputer Galera Plus located in Academic Computer Centre in Gdansk (CI TASK). This supercomputer consists of 192 computational nodes each containing two Intel Xeon Six-core processors. We used two models of parallel applications: a Single-Program-Multiple-Data application model and a Divide-And-Conquer application model.

Before energy modeling, we modeled the time of a application execution dependency on the number of processors used for calculation. We proved that our timing model is valid using MERPSYS simulation environment (described in the next section). In our simulation, we assumed usage of 1, 8, 27, 64, 125 ... 512 processes for the SPMD application and 1, 2, 4, 8, ... 1024 processes for the DAC application. We achieved results of modeling in a high accordance to the real execution times (see Figure 1) [21].

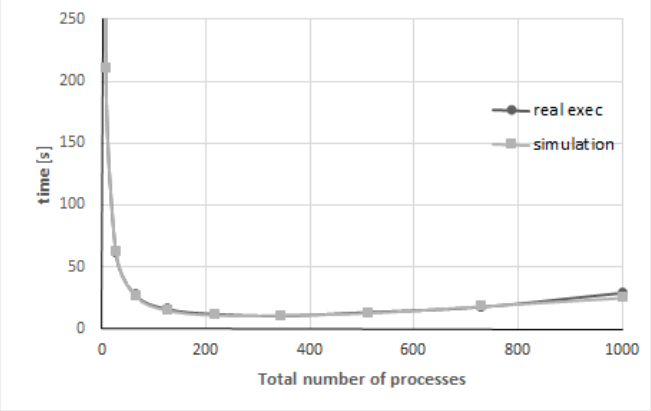
In the first application, all used computational nodes are almost equally loaded during the whole time of application execution. So energy consumption should be a simple multiplication of execution time and the power used by computational nodes involved in calculations. However in our testbed environment only 32 nodes were assigned to experiments. We assumed in our model that when the modeled number of processes was smaller than 32, each process runs on a separate node, and only a part of computational nodes are used. If the modeled number of processes is equal or greater than 32, the processes are distributed among all the available computational nodes, and all the computational nodes are used. So energy consumption in the SPMD application can be expressed as a sum of energy consumed by active nodes (E_{an}) and inactive nodes (E_{in}):

$$E = E_{an} + E_{in}$$

where the energy consumed by active and inactive node are evaluated as:

$$E_{an} = N_{an} \cdot P_{an} \cdot t_{exec}$$
$$E_{in} = N_{in} \cdot P_{in} \cdot t_{exec}$$

(a)



(b)

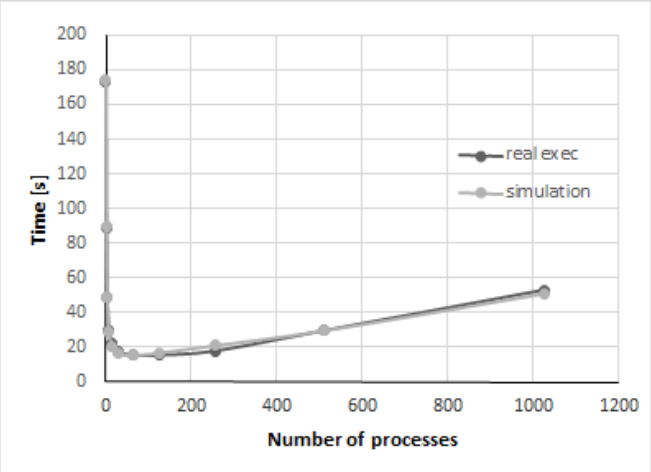


Fig. 1. Execution time modeling of (a) SPMD application and (b) DAC application

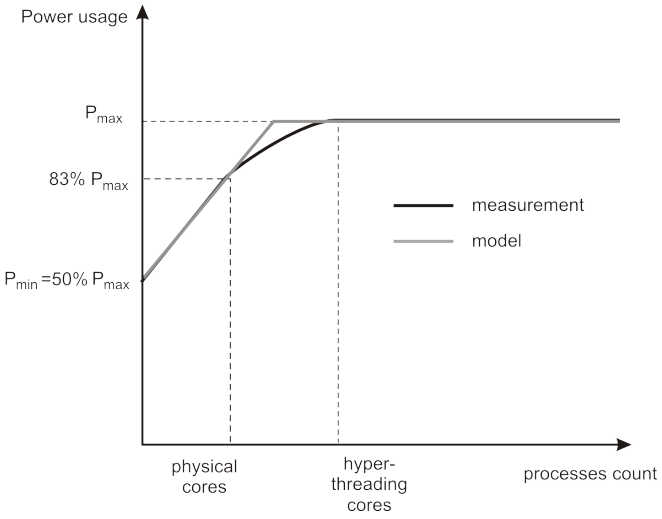


Fig. 2. Power usage on a single node depending on processes count

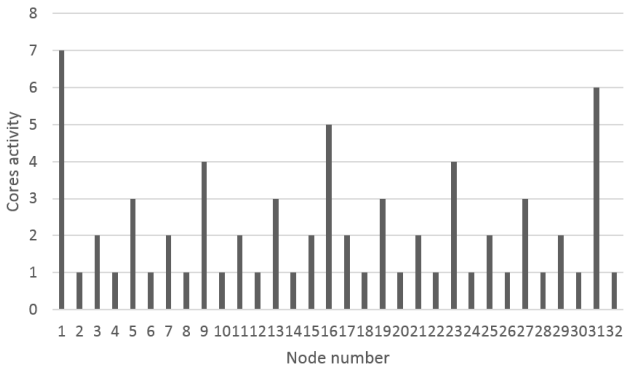


Fig. 3. Model of cores activity in DAC application consecutive steps

The number of active nodes (N_{an}) and the number of inactive nodes (Nin) are simply:

$$N_{an} = min(N_{proc}, N_{total})$$
$$N_{in} = N_{total} - Nan$$

where:

N_{proc} is the number of processes in application,
N_{total} is the total number of computational nodes (here 32),
P_{an} is modeled power usage at one active node,
P_{in} is modeled power usage at inactive node.

We measured that power usage at Intel Xeon processors in inactive node (P_{in}) equals approximately 50% of maximum power usage (P_{max}) which is consumed when all the cores are active [10]. We simplified the function of power usage due to number of active cores as a linear broken function (see Figure 2).

The model of energy consumption in the DAC application is much more complex. Computational nodes are unevenly loaded in consecutive steps of application. At the beginning all needed cores are active. In following steps every second core goes to an idle state (see Figure 3). Thus energy consumption must be evaluated in each step separately. It means that not only the number of active/inactive cores (and active/inactive nodes), but also the time of execution in each application step must be evaluated.

In the DAC application active and inactive energy is expressed by the following formulas:

$$E_{an} = \sum_{k=1..N} E_{an}(k)$$
$$E_{in} = \sum_{k=1..N} E_{in}(k)$$
$$E_{an}(k) = N_{an}(k) \cdot P_{an}(k) \cdot t(k)$$
$$E_{in}(k) = N_{in}(k) \cdot P_{in}(k) \cdot t(k)$$

where:

N is the number of steps in the application process,
k is the index of step,
$t(k)$ is time of execution of the k^{th} step,

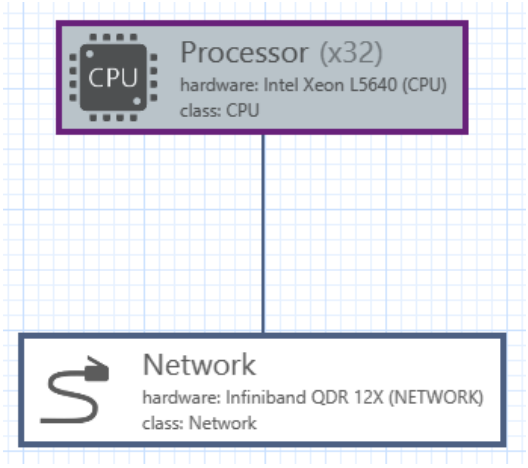


Fig. 4. Sample model of supercomputer architecture in MERPSYS

$P_{an}(k)$ is power used on an active node in the k^{th} step,
$P_{in}(k)$ is power used on an inactive node in the k^{th} step,
$N_{an}(k)$ is number of active nodes in the k^{th} step,
$N_{in}(k)$ is number of inactive nodes in the k^{th} step,
$E_{an}(k)$ is energy used on all active nodes in the k^{th} step,
$E_{in}(k)$ is energy used on all inactive nodes in the k^{th} step,

We are aware that the above model, where the total energy used for computations depends on the number of computational nodes and their usage, ignores the energy used by the whole infrastructure (e.g. cooling system), but this payload was beyond our research at this time.

V. Simulation Environment

We modeled time of execution and energy consumption in the MERPSYS simulation environment. MERPSYS enables modeling of a calculation environment (by drawing an architecture model diagram) and simulation of application execution in this environment (by writing an application model as a simulation program).

The architecture model is a graph diagram in which nodes model key architecture components, and edges model connections between components. As we modeled the Galera Plus supercomputer with homogeneous nodes our diagram consisted of two nodes: one single node modeling all computational components (all processors) and the second node modeling the internal Infiniband network connecting the computational components (see Figure 4). Next we specified component instances count (i.e. the number of computational nodes in the modeled supercomputer). Afterwards MERPSYS looked up to its component database and assigned timing parameters to components.

The application modeling program is written in the Java language, with the use of a special simulator interface, accessible by the `sim` object. We can see a sample fragment of simulation program in Figure 5. The simulation program is not the application itself. To create the simulation program we had to translate the application written in C

```
if (tag.equals("center"))      {

   sim.computation(boundary_cells_size,// computing boundary cells
         ComputationType.CPU,
         SoftwareStack.Undefined,
         1,
         linearComplexity,
         OperationType.Calculations,
         OptimizationType.None);
   sim.p2pCommunicationSend(ysize, "right");
   sim.p2pCommunicationSend(xsize, "bottom");
   sim.p2pCommunicationSend(ysize, "left");
   sim.p2pCommunicationSend(xsize, "top");
```

Fig. 5. Sample fragment of simulation program in MERPSYS

programming language to the simulation Java language. However, the simulation program is much simpler than the corresponding application program. All computational routines are modeled as sim.Computation method invocations. Interprocess communication is modeled as sim.p2pCommunicationSend/sim.p2pCommunicationReceive or one2oneCommunicationSend/one2oneCommunicationReceive invocations. All the researcher has to do is to determine the data count and the computational routines complexity.

VI. Experiments and Results

As we have mentioned above we modeled and simulated time of execution and energy consumption of two parallel applications (SPMD and DAC) in the Galera Plus supercomputer. The applications were written in C using MPI library. We compared the results of simulation with real application execution measured in this supercomputer.

A. Testbed Environment

The testbed consists of a number of identical computation nodes provided by the Academic Computer Center in Gdansk University of Technology in Poland. Each node is based on two Intel(R) Xeon(TM) CPU 2.27GHz processors (EM64T) with 6 physical processing cores with HyperThreading, 12MB cache, running Linux kernel version 2.6.32. Each node has 16 Gigabytes of RAM and they are composed in the cluster architecture, with fast (QDR, 40Gbps) Infiniband interconnection. The power meters of the cluster are served by the specialized, autonomous management subsystems: HP Integrated Lights-Out 3 (iLO 3)[5].

B. Testbed Applications

The first of the tested application is a geometric SPMD application. This kind of application can be used to solve such problems as weather prediction, heat distribution or other physical phenomena. The evaluated 3D geometric space is divided to many cuboidal regions, each region is evaluated by a separate node. The evaluation process is repeated in many iterations, between each iteration the calculation nodes interchanged data corresponding to regions borders. Data is

[5]http://h20565.www2.hpe.com/hpsc/doc/public/display?docId=emr_na-c02714903&lang=en-us&cc=us

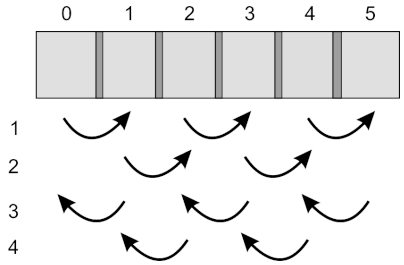


Fig. 6. The schematic inter-process data exchange in the SPMD application

interchanged in 4 steps for each dimension. Considering X dimension in the first step even nodes send data to their right neigbours, next odd nodes send data to their right neighbours, next odd nodes send data to their left neighbours, and at last event neighbours send data to their left neighbours (see Figure 6 for illustration).

Sample fragments of the SPMD application are shown below. In the beginning some common data variables are defined following by four auxiliary routines (getdata, setdata, compute_cell, and cell_to_rank). The main simulation logic is iterated in four nested *for* instructions. The most external loop iterates for an arbitrary number of steps, the internal loops iterate in the three dimensions of the geometric space. After a process has computed an associated cuboidal region in the three dimensions, the process sends data to its neighbors using the scheme shown in Figure 6.

```
double *data;
int X,Y,Z;
int procx,procy,procz;
int proccount;
...
// single data cell get method
double getdata(int x,int y,int z)
{
 return data[...];
}

// single data cell set method
void setdata(int x,int y,int z,double val)
{
 data[...]=val;
}

double compute_cell(int x,int y,int z)
{
 // computes the value of the cell
}

int cell_to_rank(int x,int y,int z)
{
// returns the rank of a process
// that owns cell (x,y,z)
}
```

```
main(int argc,char **argv)
{
 ...
 // the main simulation loop
 for(t=0;t<steps;t++)
 {
  for(i=myminx;i<=mymaxx;i++)
  for(j=myminy;j<=mymaxy;j++)
  for(k=myminz;k<=mymaxz;k++)
  {
   setdata(i,j,k,
           compute_cell(i,j,k));
  }
  // exchange data in X direction
  if (myblockx%2)
  {  // receive from left
    MPI_Recv(..., YZ_wall,
    cell_to_rank(myminx-1,myminy,myminz),
    ...);
  }
  else
  {
   // send to right
   if (myblockx+1<procx)
     MPI_Send(..., YZ_wall,
     cell_to_rank(mymaxx+1,myminy,myminz),
     ...);
  }
  ...
  // do the same for Y direction
  // and Z direction
 }// end of the iteration loop
 MPI_Finalize();
 exit(0);
}
```

The second test application is a Divide-and-Conquer merge-sort algorithm implementation. The first node, which gets the large data set, divides the data into two parts and sends one part to its free neighbor node. This process is repeated in parallel until the size of each partition reaches its limit. Then each node sorts its part of data and "odd" nodes return the sorted fragments to their "parent" nodes. The "parent" nodes merge two sorted data fragments and the process repeats until all data flow to the first node when they are merged to one sorted set (see Figure 7).

The DAC application code is shown below. For simplicity it is assumed that the size of the vector to be sorted is a power of 2. The same applies to the number of processes.

```
...
int *mergelocal(...)
{
// a function for local merging
// (i.e. one process, one thread)
```

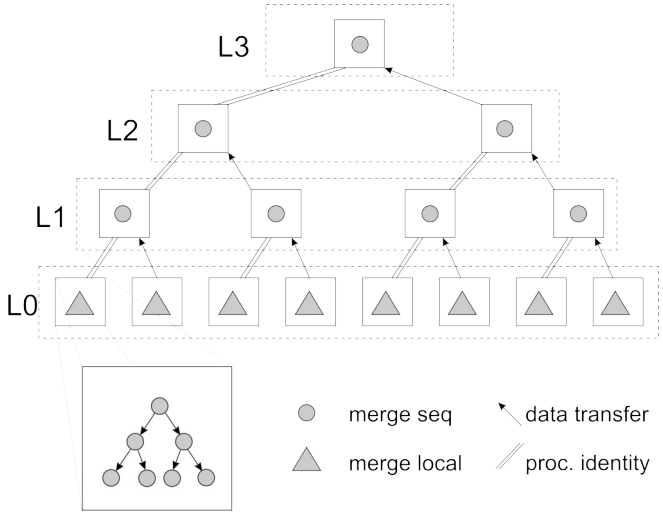


Fig. 7. The schematic algorithm of the SPMD application

```
}

int *mergeseq(int *arrayin,long length)
{
 // the main function for a sequential
 // merge (iterative)
 for(;currentlength*2<=length;
             currentlength*=2)
 {
  for(i=0;i<length;
              i+=2*currentlength)
   mergelocal(...);
   ...
 }
}

int main(int argc,char **argv)
{
 ...
 // in the first step each process needs
 // to sort its part of the array
 arrayout=mergeseq(arrayin, ...);

 // now send the data to an upper process
 if (myrank%2)
 { // then send the data
   // to process with rank myrank-1
   MPI_Send(...);
 }

 // now each process needs to check its
 // role in the divide-and-conquer tree
 // whether it should quit or receive data
 // from another process, merge and send
 // to another process
 int currentskip=2;
 // the current skip between process
 // ranks as in the above scheme
 for(;currentskip<=proccount;
       currentskip*=2)
 {
  if (!(myrank%currentskip))
  {
   // then I am involved in the given step
   // this means that I need
   // to receive the data
    MPI_Recv(...);

   // now merge data
    mergelocal(...);
    ...
   // and send to an upper process
   // if it is not the last iteration
    if (currentskip*2<=proccount)
    { // if not the last iteration
      // now check if I should
      // send the data or not
      if (myrank%(currentskip*2))
      { // then I should send the data
        // to process with rank
        // myrank-currentskip
        MPI_Send(...);
     ...
     break;
    }
   }
  }
 }
 MPI_Finalize();
}
```

C. Simulation Programs

Real calculations are not performed in the simulation program. Instead we invoke sim.computation method passing a string that descibes computational complexity. This string is composed in the simulation program as a JavaScript function that returns the number of operations. However we had to calibrate the result to the real application execution time measured in the testbed environment. It is represented by the last factor (60.94) in the computationComplexity function expression.

```
String computationalComplexity="function "
+ ConstVar.complexityFunctionName + "("
+ ConstVar.parameters + "){"
+ "return " + ConstVar.getDataSize
+"*60.94;"
+ "}";

for (t=0; t<steps; t++)
```

```
{
 sim.computation(compDataSize,
  ComputationType.CPU,
  SoftwareStack.Undefined,
  1,
  computationalComplexity,
  OperationType.Calculations,
  OptimizationType.None);
  ...
```

In the real MPI application each process is identified by a "rank" integer number. In the MERPSYS simulator we can not identify a single process. Instead we can identify a "role" of a process (with a "tag" string). So we mapped application process algorithm based on individual process number to simulation program algorithm based on process group role. We defined 7 roles for the SPMD application using relative processes position: "Center", "Left", "Right", "Top", "Bottom", "Front", and "Back". We replaced the four-step inter-process data exchange with send-receive simulation to the neighbor processes groups (see below):

```
// first send data to all neighbors
neighborTag = "center";
if (!tag.equals(neighborTag)
    && centerCount>0)
  sim.p2pCommunicationSend(wallSize,
                    neighborTag);

  ...

// then receive data from all neighbors
neighborTag = "center";
if (!tag.equals(neighborTag)
    && centerCount>0)
  sim.p2pCommunicationReceive
                    (neighborTag);
```

For the second application, we defined program roles as "levels" (from L0 to L10). We also had to define three computational complexity functions: InitComplexity, MergeSeqComplexity, and MergeLocalComplexity. Instead a peer-to-peer communication send function we used one-to-one communication send. In peer to peer communication, it is assumed that *all* pairs of processes communicate. In the DAC application *one* process sends data to *one* other process. At the same time a *half* of all the processes in the lower level send data to their corresponding processes in the upper level, so the time of one pair communication must be multiplied by sendersCount/2.

```
for (level=0;level<levelCount;level++)
{
 thisTag = "L"+Integer.toString(level);
 if (tag.equals(thisTag))
 {
  nextTag = "L"
          +Integer.toString(level+1);
  if (level==0)
  {
   sim.computation(...,
   InitComplexity,...);

   sim.computation(...,
   MergeSeqComplexity,...);
  }
  else
  {
   priorTag="L"
       +Integer.toString(level-1);
   sim.one2oneCommunicationReceive(
   priorTag);
   sim.computation(...,
   MergeLocalComplexity,...);
  }
  if (level<levelCount-1)
  {
    int sendersCount=
      sim.getNumberOfProcessesForTag
        (thisTag);
    dataSize = dataCount*4;
    sim.one2oneCommunicationSend
      (dataSize, nextTag, sendersCount/2);
  }
 }
}
```

D. Power Measurement

We measured power usage at a single node of a real execution cluster. Power usage was measured in Watts every 10 seconds of the application execution. We repeated the execution three times for each assumed processes count and then we averaged the measured values. The results are shown in Table I and Table II.

TABLE I
POWER USAGE ON A ONE NODE DURING SPMD APPLICATION EXECUTION

Processes count	Measured probes	Power usage on one node [W]	Standard deviation
8	17	84	0.68
27	5	84	0.00
64	2	93	1.49
125	1	106	0.00
216	1	126	0.00
343	1	147	4.71
512	1	159	4.71
729	1	48	0.00
1000	2	99	43.14

E. Simulation Results and Comparison

Having accurate time simulaton results (see Figure 1), we could base energy simulation on solid foundations. We evaluated energy consumption in the MERPSYS simulator and compared the results to the measured energy consumption.

TABLE II
POWER USAGE ON ONE NODE DURING DAC APPLICATION EXECUTION

Processes count	Measured probes	Power usage on one node [W]	Standard deviation
1	16	86.7	2.14
2	8	87.0	1.15
4	4	87.0	1.00
8	2	87.3	0.94
16	1	87.3	0.94
32	1	86.7	0.94
64	1	98.0	0.00
128	1	113	1.89
256	1	137	1.89
512	2	123	43.97
1024	4	134	41,20

TABLE III
REAL AND SIMULATED ENERGY CONSUMPTION IN SPMD APPLICATION

Proc. count	Active nodes count	Inact. nodes count	Real energy cons. [Wh]	Simul. energy cons. [Wh]
8	8	24	151.20	152.1
27	27	5	45.67	46.6
64	32	0	22.71	22.7
125	32	0	15.13	15.0
216	32	0	13.46	13.2
343	32	0	13.86	14.2
512	32	0	18.97	19.0

However as we measured energy consumption at a one node only, we had to recalculate the measured results according to the model of application. In the SPMD applications, as all nodes assigned to the application are active all the time, it was easy to calculate the energy consumption in the whole experimental environment. We show the compared results in Table III. However in the DAC application nodes are unevenly active. We applied the model of activity shown in Figure 3 and recalculated active and inactive nodes real energy consumption in all the applications steps separately, and next summarized them. As the results depend not only on the really measured energy consumption but on the model of activity as well, we call the results the "modeled" energy. Comparison between modeled energy consumption and simulated energy consumption is shown in Table IV.

TABLE IV
MODELED AND SIMULATED ENERGY USAGE IN DAC APPLICATION (COMPARISON)

Proc. count	Modeled energy usage at active nodes	Modeled energy usage at inactive nodes	Total modeled energy usage	Sim. energy usage
1	4.15	118.85	123.0	123.2
2	4.26	59.30	63.6	63.6
4	4.54	30.10	34.6	34.7
8	5.17	15.61	20.8	20.8
16	6.52	7.47	14.0	14.4
32	9.54	2.41	12.0	11.9
64	8.83	3.10	11.9	11.7
128	10.27	4.01	14.3	13.4
256	15.42	5.40	20.8	16.9
512	25.95	7.76	33.7	27.2
1024	47.11	12.83	59.9	55.9

VII. CONCLUSIONS AND FUTURE WORK

In the paper we presented a way to model parallel SPMD and divide-and-conquer applications within the MERPSYS environment including application and system models. Next, we presented verification of results obtained from the fast MERPSYS simulator against energy consumption that stemmed from consideration of power usage of real cluster nodes. We performed tests and calculations for up 512 and 1024 processes for SPMD and divide-and-conquer applications respectively reaching a high degree of accuracy. This allows to obtain results for these applications for other configurations such as input data sizes with ease without the need for rerunning the real application and much faster than the latter.

The future works should cover wider variety of the software and hardware systems. Different vendors and configurations should be tested as well as new simulated programs.

REFERENCES

[1] P. Czarnul, P. Rosciszewski, M. R. Matuszek, and J. Szymanski, "Simulation of parallel similarity measure computations for large data sets," in *2nd IEEE International Conference on Cybernetics, CYBCONF 2015, Gdynia, Poland, June 24-26, 2015*. IEEE, 2015. doi: 10.1109/CYBConf.2015.7175980. ISBN 978-1-4799-8322-3 pp. 472–477. [Online]. Available: http://dx.doi.org/10.1109/CYBConf.2015.7175980

[2] W. McFadden, A. Nikolich, R. Parpart, and B. Runesha, "Saving on data center energy bills with edeals: Electricity demand-response easy adjusted load shifting," in *USENIX Workshop on Cool Topics on Sustainable Data Centers (CoolDC 16)*. Santa Clara, CA: USENIX Association, 2016. [Online]. Available: https://www.usenix.org/conference/cooldc16/workshop-program/presentation/mcfadden

[3] T. Cioara, I. Anghel, I. Salomie, D. Moldovan, G. Copil, and P. Plebani, "Dynamic consolidation methodology for optimizing the energy consumption in large virtualized service centers," in *Federated Conference on Computer Science and Information Systems - FedCSIS 2011, Szczecin, Poland, 18-21 September 2011, Proceedings*, M. Ganzha, L. A. Maciaszek, and M. Paprzycki, Eds., 2011. ISBN 978-83-60810-22-4 pp. 1005–1011. [Online]. Available: http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6078295

[4] H. Shoukourian, T. Wilde, A. Auweter, and A. Bode, "Predicting the energy and power consumption of strong and weak scaling hpc applications," *Supercomputing frontiers and innovations*, vol. 1, no. 2, 2014. doi: 10.14529/jsfi140202. [Online]. Available: http://dx.doi.org/10.14529/jsfi140202

[5] G. Lawson, M. Sosonkina, and Y. Shen, "Towards modeling energy consumption of xeon phi." *CoRR*, vol. abs/1505.06539, 2015. [Online]. Available: http://dblp.uni-trier.de/db/journals/corr/corr1505.html#LawsonSS15

[6] A. Tiwari, M. A. Laurenzano, L. Carrington, and A. Snavely, "Modeling power and energy usage of hpc kernels," in *Parallel and Distributed Processing Symposium Workshops PhD Forum (IPDPSW), 2012 IEEE 26th International*, May 2012. doi: 10.1109/IPDPSW.2012.121 pp. 990–998.

[7] F. Almeida, V. B. Pérez, A. C. Pérez, and J. Ruiz, "Modeling energy consumption for master-slave applications," *The Journal of Supercomputing*, vol. 65, no. 3, pp. 1137–1149, 2013. doi: 10.1007/s11227-013-0914-y. [Online]. Available: http://dx.doi.org/10.1007/s11227-013-0914-y

[8] C. Lively, X. Wu, V. Taylor, S. Moore, H.-C. Chang, and K. Cameron, "Energy and performance characteristics of different parallel implementations of scientific applications on multicore systems," *International Journal of High Performance Computing Applications*, vol. 25, no. 3, pp. 342–350, 2011. doi: 10.1177/1094342011414749 Energy. [Online]. Available: http://dx.doi.org/10.1177/1094342011414749

[9] R. Isidro-Ramirez, A. M. Viveros, and E. H. Rubio, "Energy consumption model over parallel programs implemented on multicore architectures," *International Journal of Advanced Computer Science and Applications(IJACSA)*, vol. 6, no. 6, 2015. doi: 10.14569/IJACSA.2015.060635. [Online]. Available: http://dx.doi.org/10.14569/IJACSA.2015.060635

[10] J. Proficz and P. Czarnul, *Parallel Processing and Applied Mathematics: 11th International Conference, PPAM 2015, Krakow, Poland, September 6-9, 2015. Revised Selected Papers, Part II*. Cham: Springer International Publishing, 2016, ch. Performance and Power-Aware Modeling of MPI Applications for Cluster Computing, pp. 199–209. ISBN 978-3-319-32152-3. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-32152-3_19

[11] P. Czarnul and M. Matuszek, *Parallel Processing and Applied Mathematics: 11th International Conference, PPAM 2015, Krakow, Poland, September 6-9, 2015. Revised Selected Papers, Part II*. Cham: Springer International Publishing, 2016, ch. Considerations of Computational Efficiency in Volunteer and Cluster Computing, pp. 66–74. ISBN 978-3-319-32152-3. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-32152-3_7

[12] L. A. Barroso and U. Hölzle, "The case for energy-proportional computing," *Computer*, vol. 40, no. 12, pp. 33–37, Dec. 2007. doi: 10.1109/MC.2007.443. [Online]. Available: http://dx.doi.org/10.1109/MC.2007.443

[13] P. Balaprakash, A. Tiwari, and S. M. Wild, *High Performance Computing Systems. Performance Modeling, Benchmarking and Simulation: 4th International Workshop, PMBS 2013, Denver, CO, USA, November 18, 2013. Revised Selected Papers*. Cham: Springer International Publishing, 2014, ch. Multi Objective Optimization of HPC Kernels for Performance, Power, and Energy, pp. 239–260. ISBN 978-3-319-10214-6. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-10214-6_12

[14] K. M. Tarplee, R. Friese, A. A. Maciejewski, and H. J. Siegel, "Efficient and scalable computation of the energy and makespan pareto front for heterogeneous computing systems," in *Computer Science and Information Systems (FedCSIS), 2013 Federated Conference on*, Sept 2013, pp. 401–408.

[15] ——, "Efficient and Scalable Pareto Front Generation for Energy and Makespan in Heterogeneous Computing Systems," in *Recent Advances in Computational Optimization*, S. Fidanova, Ed. Cham: Springer International Publishing, 2015, vol. 580, pp. 161–180. ISBN 978-3-319-12630-2 978-3-319-12631-9. [Online]. Available: http://link.springer.com/10.1007/978-3-319-12631-9_10

[16] A. Tiwari, M. A. Laurenzano, L. Carrington, and A. Snavely, "Auto-tuning for energy usage in scientific applications," in *Proceedings of the 2011 International Conference on Parallel Processing - Volume 2*, ser. Euro-Par'11. Berlin, Heidelberg: Springer-Verlag, 2012. doi: 10.1007/978-3-642-29740-3_21. ISBN 978-3-642-29739-7 pp. 178–187. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-29740-3_21

[17] P. Czarnul and P. Rosciszewski, "Optimization of execution time under power consumption constraints in a heterogeneous parallel system with gpus and cpus," in *Distributed Computing and Networking - 15th International Conference, ICDCN 2014, Coimbatore, India, January 4-7, 2014. Proceedings*, ser. Lecture Notes in Computer Science, M. Chatterjee, J. Cao, K. Kothapalli, and S. Rajsbaum, Eds., vol. 8314. Springer, 2014. doi: 10.1007/978-3-642-45249-9_5. ISBN 978-3-642-45248-2 pp. 66–80. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-45249-9_5

[18] P. Rościszewski, P. Czarnul, R. Lewandowski, and M. Schally-Kacprzak, "KernelHive: a new workflow-based framework for multilevel high performance computing using clusters and workstations with CPUs and GPUs," *Concurrency and Computation: Practice and Experience*, vol. 28, no. 9, pp. 2586–2607, Jun. 2016. doi: 10.1002/cpe.3719. [Online]. Available: http://doi.wiley.com/10.1002/cpe.3719

[19] Z. Zong, X. Qin, X. Ruan, K. Bellam, M. Nijim, and M. I. Alghamdi, "Energy-efficient scheduling for parallel applications running on heterogeneous clusters," in *2007 International Conference on Parallel Processing (ICPP 2007), September 10-14, 2007, Xi-An, China*. IEEE Computer Society, 2007. doi: 10.1109/ICPP.2007.39 p. 19. [Online]. Available: http://dx.doi.org/10.1109/ICPP.2007.39

[20] D. Li, B. R. de Supinski, M. Schulz, D. S. Nikolopoulos, and K. W. Cameron, "Strategies for energy-efficient resource management of hybrid programming models." *IEEE Trans. Parallel Distrib. Syst.*, vol. 24, no. 1, pp. 144–157, 2013. doi: 10.1109/TPDS.2012.95. [Online]. Available: http://dx.doi.org/10.1109/TPDS.2012.95

[21] P. Czarnul, Ed., *Modeling Large-Scale Computing Systems. Practical Approaches in MERPSYS*. Gdansk University of Technology, 2015. ISBN 978-83-938367-2-7

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 865–872
DOI: 10.15439/2016F255
ACSIS, Vol. 8. ISSN 2300-5963

Efficient parallel execution of genetic algorithms on Epiphany manycore processor

Łukasz Faber, Krzysztof Boryczko
AGH University of Science and Technology
al. Mickiewicza 30, 30-059 Kraków, Poland
E-mail: {faber,boryczko}@agh.edu.pl

***Abstract*—Recent years have seen a growing trend towards the introduction of more advanced manycore processors. On the other hand, there is also a growing popularity for cheap, credit-card-sized, devices offering more and more advanced features and computational power.**

In this paper we evaluate Parallella – a small board with the Epiphany manycore coprocessor consisting of sixteen MIMD cores connected by a mesh network-on-a-chip. Our tests are based on classical genetic algorithms. We discuss some possible optimizations and issues that arise from the architecture of the board. Although we achieve significant speed improvements, there are issues, such us the limited local memory size and slow memory access, that make the implementation of efficient code for Parallella difficult.

I. Introduction

FOLLOWING the Manycore Revolution [1] and the popularity of small integrated, power consumption- and cost-oriented computing boards (for example, Raspberry Pi), it was expected that these two "directions" would merge at some point. One of the results of this "merge" is the Parallella board[1] [2], created by Adapteva. It is a small (credit card-sized) board, comprising of a 16-core Epiphany coprocessor and the main dual-core ARM processor.

Such boards are interesting for researchers due to their costs, simplicity, and the low level requirements for beginning work with them. Parallella has the benefits of being a standalone, plug-and-play coprocessor similarly to Intel Phi [3]. In this case, "standalone" means that it is a completely separate computer unit that can run independently of any other nodes. On the other hand, it is plug-and-play, because it only requires an Ethernet cable to connect to it.

In this paper we want to review the Parallella board as a simple and effective tool for implementing strictly computational systems. We do not expect the platform to provide higher performance than well established manycore architectures like GPGPU. However, it seems that the programming model and achievable efficiency are *good enough* for creating quick, low-cost hardware-accelerated parallel platforms for simulations and computations. We want to show that it is feasible to implement various execution strategies on the platform and demonstrate its possible weaknesses.

The work reported in this paper concentrates on the realization of genetic and evolutionary algorithms on the Parallella board. It is related to and extends our previous publications regarding the implementation of effective tools for running population-based computational intelligence systems [4], especially using the agent paradigm [5], [6] in both parallel and distributed [7], as well as heterogeneous environments [8].

In the following sections, we briefly introduce the Parallella platform and Epiphany manycore (Section II) and its memory and programming models. Then, in Section III, we discuss our benchmarks and introduced optimizations. In Section IV we present the results. Finally, these results are discussed in Section V, alongside introducing the next steps we are taking with Parallella.

II. Parallella

Parallella is a hardware platform built on top of the manycore Epiphany [2] coprocessor, created by Adapteva in 2011. It was funded by a Kickstarter campaign[2] The board was first presented in June 2014.

The Epiphany processor consists of a 2D array of nodes (known as "eNodes") connected by a mesh network-on-a-chip. Each node consists of a single RISC core (called "eCore"), a DMA engine, 32 kB of memory and a network interface. Each core includes a 32-bit floating point RISC CPU, local memory, a DMA engine, an event monitor and a network interface. The mesh connections on a 16-core processor are shown in Figure 1.

Each "eCore" contains a floating-point unit (FPU), an arithmetic logic unit (ALU) and a 64-word register file, as shown in Figure 2.

The address space in Epiphany is flat and consists of 2^{32} bytes. Each node has 32 kB of its own local range of memory aliased in addresses `0x0000–0x7FFF`. However, memory of each node can be accessed by prefixing the address with a globally addressable ID consisting of 6 bits for a row ID and 6 bits for a column ID (counted from 0) – thus giving a theoretical maximum of $64 \times 64 - 1 = 4095$ cores. For example, if a core wanted to access the memory of the core located in the second row and the third column (1,2) it would access addresses `0x04200000–0x04207FFF`.

Some specifics of the Epiphany architecture related to eMesh include:

[1] https://www.parallella.org/

[2] https://www.kickstarter.com/projects/adapteva/parallella-a-supercomputer-for-everyone

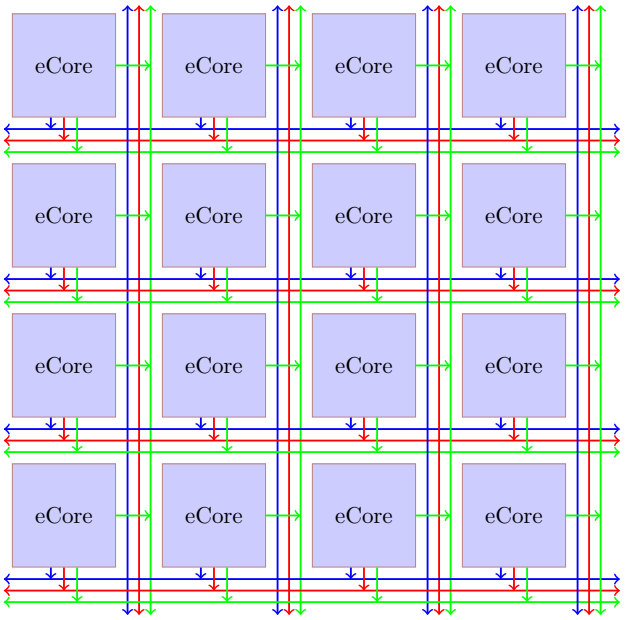


Fig. 1. eMesh Network-on-a-Chip. The blue lines indicate cMesh (used for on-chip writes), green — xMesh (off-chip writes), red — rMesh (read requests).

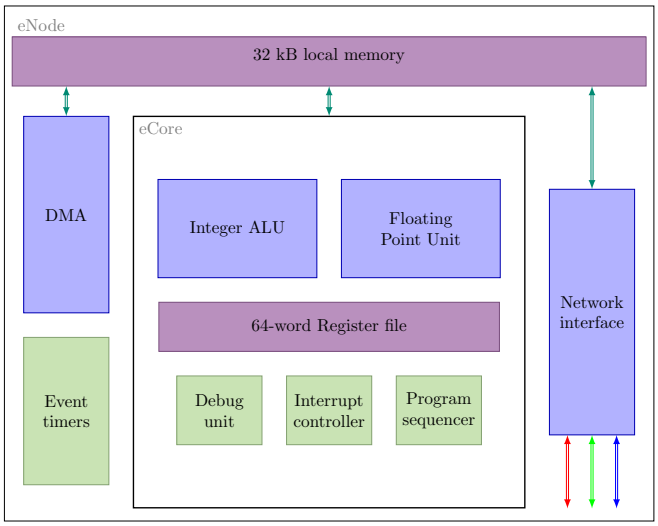


Fig. 2. eNode components. Each eNode has an eCore and 32 kB of local memory, a network router, a DMA engine and two event timers. Each eCore has a 64-word register file, FPU, ALU, interrupt controller, sequencer and debug unit.

- writes are preferred over reads – for a single read there need to be two transactions: one for a read request and one for an answer,
- non-local memory accesses are weakly ordered.

The main goals of the Epiphany architecture are: power efficiency (a single 16-core Epiphany processor consumes a maximum of 2 W, the whole Parallella board requires around 5 W), scalability, an easy programming model, and high performance (2 GFLOPS per single core).

Currently, there are 16- and 64-cores Epiphany processors available. However, 64-core versions have only been produced in limited numbers and are only available directly from Adapteva.

The Parallella board uses a 16-core Epiphany processor (E16G301), a Xillinx Zynq (models 7010 or 7020 with two ARM cores) and 1 GB of RAM. Additional components include: 1 Gbit Ethernet interface, USB and HDMI ports, and a MicroSD slot. The standard operating system (in this case – Linux) boots from the MicroSD card onto an ARM processor and can communicate with the Epiphany using an e-Link interface. 32 MB of memory is shared between ARM processors (host) and Epiphany. It is mapped in eNodes to `0x8E000000–0x8FFFFFFF` address space. eCores can use it in the same way as internal memory, however the performance will be lower as shown in Section IV-A1.

The memory size usable by the programmer is dependent on many factors. In the most common linker configuration the internal memory (32 kB) is used, for example, to store:

- program code,
- global variables,
- stack.

And a fragment of the external memory (32 MB) is used for the C standard library code, data and stack [9].

Programming on the Epiphany side is done in the usual way. The SDK supports the standard C library with mathematics functions. Additionally, it provides some specific utilities for managing hardware resources: registers operations, interrupts handling, timers, mutexes, barriers and DMA functions. The "workgroup" concept is supported and each created workgroup can have a different code loaded. The important thing to remember is that Epiphany is an MIMD processor and each core can execute a completely different code. There is no synchronization between cores (besides library functions and experimental `SYNC` instruction).

Alternative approaches to using the basic SDK are MPI [10] and OpenMP [11].

Epiphany does not provide double precision float operations in hardware. As such, these are emulated by a compiler and thus carry performance loss when used.

III. Programs

Our benchmark application was a simple genetic algorithm [12] with the fitness-proportional selection, mutation

enabled and using the two-dimensional Beale's function as a fitness function:

$$f(x,y) = (1.5 - x + xy)^2 + \left(2.25 - x + xy^2\right)^2 + \left(2.625 - x + xy^3\right)^2$$

The initial population was generated within $[-5.0, 5.0]$ boundary.

With this basic skeleton we prepared several versions of the implementation for the Parallella using different solutions and optimizations.

As we wanted to implement the whole algorithm on the Epiphany processor, we needed a separate pseudo-random number generator. We used a "tiny" version of the Mersenne Twister [13], [14] that requires only 127 bits of memory, and could be easily used with Epiphany.

A. Fitness computation offloading

The first and the simplest way to use the Epiphany processor is to offload what is usually the heaviest computation in genetic algorithms – the evaluation of the fitness function.

The main loop of the computation performs the following operations:

1) Generate a new population (on the host side).
2) Compute the fitness (on the Epiphany processor).
3) Find the best organism (on the host side).

The simplified version of the function initializing and starting Epiphany cores is shown in Listing 1. We perform the following steps:

1) reset workgroup,
2) load the device code,
3) write population and its size,
4) start the workgroup,
5) wait for all cores to finish working,
6) read the computed fitness values.

```
void epi_fitness_fill(
    simulation_t * simulation,
    e_platform_t * platform) {

  // Reset device and load code
  e_reset_group(&dev);
  e_load_group("e_main.srec", &dev, 0, 0,
    platform->rows, platform->cols,
    E_FALSE);

  // Write required data - size and population
  e_write(&e_size_mem, 0, 0, 0x0,
    &(simulation->size), sizeof(float));
  e_write(&e_population_mem, 0, 0, 0x0,
    simulation->population,
    simulation->population_size);

  // Send start interrupt
  e_start_group(&dev);

  // Wait for all cores to finish work
  epi_wait_for_status(STATUS_EXITED);

  // Read computed fitness
  e_read(&e_fitness_mem, 0, 0, 0x0,
    simulation->fitness.values,
    simulation->fitness_size);
}
```

Listing 1. Host part of fitness computation

For transferring the population we use an array of pairs of floats (with a size twice that of the population). The output (fitness values) is stored in a separate array of floats. In the basic version, both are located in the external memory.

B. Full population evaluation

In the second version we offloaded both the fitness computation and finding the best individual to the Epiphany cores. In this case, each core, finds the best individual in its fragment of the population. After computing all the fitness values, each core sends (using a 16-element `size_t` array) the index of the best individual that was found. The host collects these 16 elements and chooses the best among them. It also generates a new population.

C. Whole algorithm on Epiphany

In the final stage, we also implemented the whole algorithm on the Epiphany processor. Most of the implementation was straightforward, as the code flow could be similar. The main issue was handling large data with the small amount of local memory.

In order to handle a large number of organisms efficiently, for each iteration, we split the population into chunks each consisting of 16×1024 organisms (except the last one). Each core copies 1024 organisms from the external memory into its local memory. Then, it computes the fitness and generates a new population for this fragment and puts the new organisms to the external memory. After all cores perform these operations on all chunks, they load new organisms from the external memory and execute the next iteration.

A loss in the exchange of organisms between cores can be observed and we basically operate on subpopulations. However, to resolve (at least partially) this problem, we shuffle the population when copying organisms from the external memory. We do this efficiently using the DMA engine (see Section III-D3).

There is no explicit communication between cores and the host, although the host can "preview" the data at any time. In addition, cores do not exchange data in any other way than using the external memory, as previously mentioned.

The role of the host is only limited to initializing the population, copying data to the external memory and, after all iterations, copying the final population and fitness values back.

D. Implementation solutions used in test programs

In order to improve the efficiency of the implementation, we have implemented and tested technical optimizations such as removing the need for reloading the Epiphany code and using better communication facilities.

1) Removing need for reset–load before each computation: By default Epiphany does not provide a simple way to "restart" the same computation as before. The programmer needs to reset a device group and load the device code again when needed. Although such an approach generally works, it could have a noticeable impact on performance. We have measured the reload time and presented the results in Section IV.

The solution to this issue is to make the computation execute in an infinite loop and signal new data available for processing by using the same interrupt that is used for starting the computation (`E_SYNC`). At the beginning of the computation we register and enable an empty handler for this interrupt. Then, when waiting, we execute `idle` opcode in a loop effectively putting the core to sleep. When it receives the interrupt, it wakes up and retests the loop condition which can be, for example, checking for new data. The waiting loop is shown in the Listing 2.

```
while (1) {
  // Perform computation
  while (/* Test status/flag */){
    // Go into idle mode
    __asm__ __volatile__ ("idle");
  }
}
```

Listing 2. Waiting loop

On the host side we simply execute `e_start_group` on the whole workgroup each time we want to signal new data. Internally, it sends the correct interrupt.

2) Communication: In most cases there is no need for communication between cores, as data can be cleanly split among them. On the other hand, communication with the host is performed at least two times for every step of the computation (for versions executing on Epiphany only partially). Epiphany does not provide any way to synchronize its cores with the external host.

We solved this issue by using a simple status variable that eCores use to share their statuses with the host and the host signals that there is new work to do. We store statuses in a 16-element array of `uint8_t` (8-bit unsigned integer) elements. This gives us enough flexibility to use multiple different statuses and is still small enough not to waste the memory.

To wait for a specific status on the host side we make a tight loop for reading the whole array and check all values after every read.

On the Epiphany side if the core needs to wait for a particular status, it is done by a simple `while` loop checking a dedicated array index.

3) DMA: We have also decided to test the difference the DMA engine makes to memory operations. Epiphany offers two DMA channels per core. They can be used in a linear (copying continuous blocks of memory) or non-linear (copying regularly spread fragments of memory) fashion. In most cases we used the former method, but for the porting of the whole algorithm, we used the latter. This allowed us to efficiently shuffle populations without involving cores.

TABLE I
MEASURED MEMORY BANDWIDTH ON PARALLELLA.

Initiator	Target	Type	Bandwidth (MB/s)
ARM Host	eCore (0,0)	write	45.82
ARM Host	eCore (0,0)	read	5.20
ARM Host	DRAM	write	88.25
ARM Host	DRAM	read	131.96
ARM Host	DRAM	memcpy	353.01
eCore (0,0)	eCore (1,0)	write (DMA)	1242.38
eCore (0,0)	eCore (1,0)	read (DMA)	401.46
eCore (0,0)	DRAM	write (DMA)	233.94
eCore (0,0)	DRAM	read (DMA)	87.45
eCore (0,0)	eCore (1,0)	write	534.37
eCore (0,0)	eCore (1,0)	read	115.70
eCore (0,0)	DRAM	write	71.61
eCore (0,0)	DRAM	read	4.29

IV. RESULTS

Our tests focused mainly on execution time and memory performance. We performed two micro-benchmarks in order to determine:

- memory bandwidths,
- Epiphany initialization penalty.

Then, we measured times related to the execution of the test case presented in the previous section.

A. Microbenchmarks

1) Memory bandwidth: The memory bandwidth results shown in Table I were obtained using a micro-benchmark provided with the Parallella SDK. It is easily observable that writes involving the Epiphany processor perform significantly (several times) better than corresponding reads. This is related to the way the reads are executed (see Section II) – they consist of two transactions. The DMA engine also provides a large speed boost.

Moreover, these operations are significantly slower than the theoretical limits which would be 1.6 GB/s [15] for the bidirectional off-chip traffic and 8 GB/s for on-chip DMA [16]. However, for the former the cMesh implementation limits it to around 4.8 GB/s. There are also two errata items reporting issues with DMA engine limiting its bandwidth to around 25% of this value.

These speeds and significant differences to the theoretical values are similar to other results [17].

2) Loading the code on Epiphany: To load and start the code on the Epiphany processor we need to call the following functions:

1) `e_reset_group`
2) `e_load_group`
3) `e_start_group` (this can be implicitly called by previous function).

There is no means provided to "restart" the same code on Epiphany and it is necessary to handle it manually using interrupts (see Section III). Thus, we decided to perform a microbenchmark measuring the time it takes to execute a full reload operation on all 16 cores.

The results are as follows: for the code and data filling the whole 32 kB of core memory: 314.613 ms. For the

minimal example (writing a single integer to the well-known memory buffer): 117.620 ms. These values are not high, but considering possible multiple iterations of the algorithm they can accumulate to significant delays.

B. Genetic algorithm

We implemented several different optimization scenarios and measured execution times for each of them. These scenarios were:

- naive — no Epiphany-specific optimizations: we put the data in the external (DRAM) memory and restarted Epiphany in each iteration, cores use data directly from DRAM,
- no reload — as above but without code reloading and Epiphany restart (see Section III-D1),
- local — as "naive" but each core copied fragments of data to local memory before processing,
- no reload, local — as "no reload" but each core copies fragments of data to the local memory before processing (using standard `memcpy`),
- no reload, local, dma — as "no reload, local" but cores used DMA engine to perform copying of data,
- no reload, push — as "no reload" but the host (ARM) pushed data directly to the local memory of each core; it is the host's responsibility to split the data into sizes fitting the local Epiphany memory.

Additionally, we executed the basic one-thread computation on the Parallella's ARM CPU as a reference for other measurements.

For every scenario we executed 100 (one hundred) iterations over various selected population sizes: 16, 32, 128, 256, 1024, 2048, 8192, 10240, 51200. These numbers needed to be divisible by 16, as each core should have the same population to work on. The results for 51200 are presented in Table II.

Our time measurements included three, increasing in size, portions of the code:

- code execution on the Epiphany coprocessor in a single iteration and the same code on the host (in the CPU version),
- a single iteration — both ARM host and Epiphany code including code upload in some of the above scenarios but without the generation of a new population,
- a whole algorithm — from the Epiphany initialization to closing the device.

For measuring the execution time we used the `clock_gettime()` function with the `CLOCK_MONOTONIC` clock.

We tested the three scenarios described in Section III: fitness offloading, population evaluation and a whole algorithm. For the whole algorithm we measured only the full computation time, as the measurement of a single iteration would not be efficient or useful.

It is worth noting that in the largest measured population (51200 organisms) each Epiphany core has only 3200 organisms to evaluate.

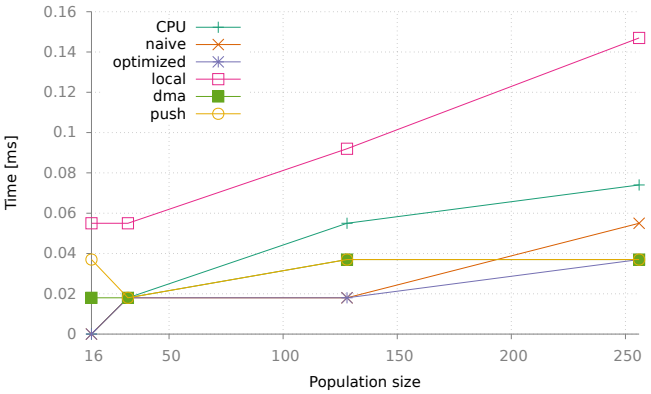


Fig. 3. Time of a single iteration execution of the "fitness offloading" version on the Epiphany processor for population sizes 16, 32, 128, 256.

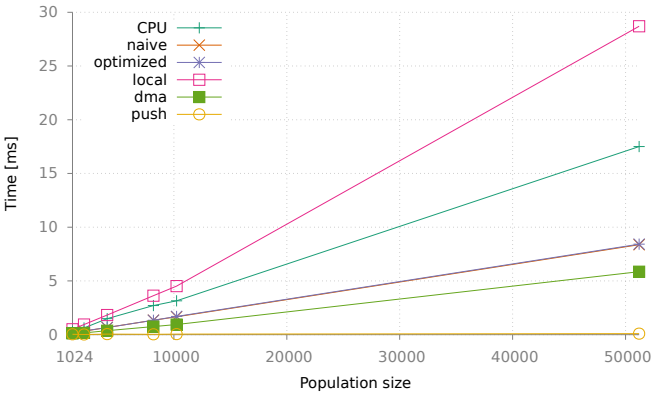


Fig. 4. Time of a single iteration execution of the "fitness offloading" version on the Epiphany processor for population sizes 1024, 2048, 8192, 10240, 51200.

Figures 3 and 4 show times for the Epiphany-offloaded code and the equivalent code on the CPU. Firstly, we can observe, that the "push" version performs the best. This is a direct result of the fact that there are no external memory operations on the Epiphany side in this version. All memory handling is done on the host. Secondly, the "local" version that uses `memcpy` performs more than two times worse than the CPU version. This is caused by very slow (around 4.3 MB/s, as shown in Table I) read rates for non-DMA copies between cores and DRAM. After changing the copy method to DMA the measured time was significantly reduced, performing even better than no-copy versions ("naive" and "no reload"). Finally, there are no significant differences between versions with and without code reloading so more complicated iteration logic on the Epiphany side has no penalties (as expected).

For smaller population sizes some of these observations differ (DMA and "push" versions are slower), but this is due to the initialization of these memory access paths [16].

Figures 5 and 6 show times for the whole iteration – host and device sides – without generation of the new population. They do not include lines for the "naive" version for readability.

TABLE II
MEASURED EXECUTION TIMES FOR 51200 ORGANISMS.

Version	Version	Epiphany [ms]	Iteration [ms]	Full [s]
CPU	—	13.769	16.533	545.53
fitness offloading	none	8.368	223.764	546.83
	no reload	8.423	22.303	546.24
	no reload, local	28.699	42.559	546.03
	no reload, local, dma	5.806	19.593	545.89
	local, dma	5.861	232.556	546.93
	no reload, push	0.276	51.702	546.11
full	none	8.645	215.802	546.68
	no reload	8.737	18.248	545.90
	no reload, local, dma	5.714	15.280	545.87
whole	local memory	—	—	2.42
	external memory	—	—	38.73

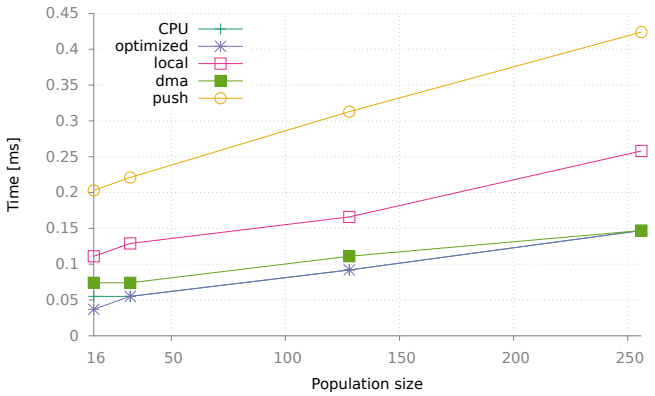


Fig. 5. Time of a single iteration execution of the "fitness offloading" version (Epiphany + host) for population sizes 16, 32, 128, 256. The "naive" version is not included due to the scale.

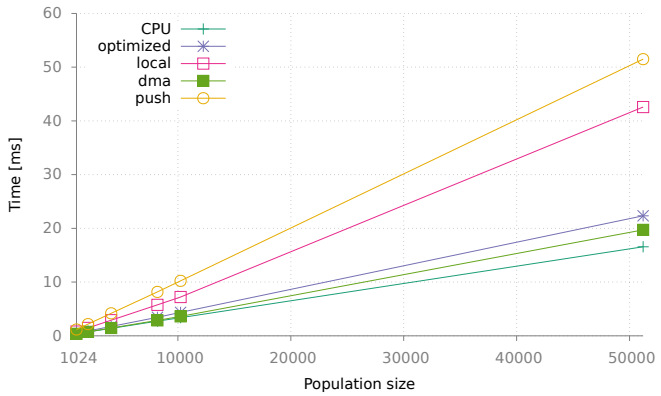


Fig. 6. Time of a single iteration execution of the "fitness offloading" version (Epiphany + host) for population sizes 1024, 2048, 8192, 10240, 51200. The "naive" version is not included due to the scale.

The first thing to observe is the very poor performance (as much as ten times worse) of the versions that reset the Epiphany during the iteration. As we noted in Section IV-A2, such an operation takes at least 110 ms (for the smallest possible binary).

In these results, we see that the "push" version performs poorly. This is due to the fact, that all of the memory operations are executed on the host side and, firstly, host-to-core transfers are slower (as seen in Table I), secondly, the host needs to perform sixteen separate copies instead of one to the external memory. The CPU version has the best performance, but it is important to note, that there are no additional memory copies in this version. In all Epiphany implementations we need to perform at least one additional copy of the population per iteration.

We also measured the execution time of the whole program (for a smaller number of iterations), however, as the most time-consuming task for large populations is the generation of the new population, the results are similar in each implementation.

The "full evaluation" version, which also included finding the best organism on Epiphany, did not make any significant difference to the previous one. We can observe, that iteration times are several milliseconds shorter than the corresponding "offload" version. This follows from the fact, that we split the lookup on all cores.

The final version we tested — the whole algorithm ported to Epiphany — was implemented in two versions: one which copied the data to the local memory of cores (using DMA), and another which operated fully on the external memory.

The latter performed nearly 15 times better than the single CPU version, which is very good considering the memory bandwidths shown in the Table I. However, the former, using the local memory, presented the best computation time: below 3 s.

C. Memory limits of Parallella

As mentioned in Section II, the Epiphany cores in Parallella board can address only 32 kB of the internal and 32 MB of the external memory. These limits have significant impact on the size of program and data size.

This 32 kB of internal memory is divided in four banks, of which only two are usable for the user data without limitations, because the first one contains the code and the last one – the stack (starting at the end of the block). It is safe to assume that, for user data, 16 kB is fully available on each core, plus the remaining memory from the fourth bank. We have limited the local memory used in test scenarios to these sizes. As we use 24 bytes for a single organism, we can store a maximum

of 1024 organisms (if we ignore the stack). However, in cases where we do not need to remember the population between iterations we can safely reuse the same memory for delivering results to the host, giving us 1536 organisms for a single core.

As previously mentioned, one has to consider the stack, so the final numbers have to be smaller and the real code size must be taken into account.

D. Observed issues

During development we observed several issues with the board and SDK.

First, the board has a tendency to overheat. When the temperature measured by sensors reached 65°C we observed unstable behavior: missing writes (the host did not see updates from Epiphany) and hangups. To counteract these issues we kept the board cooled to around 50°C.

Second, the default linker configuration places standard C library and math functions in the external memory (DRAM). For example, we tested the code with the two-dimensional Ackley test function:

$$f(x,y) = -20\exp\left(-0.2\sqrt{0.5\left(x^2+y^2\right)}\right) - \exp\left(0.5\left(\cos\left(2\pi x\right)+\cos\left(2\pi y\right)\right)\right) + e + 20$$

In this case, the Epiphany version was several times slower than the plain CPU implementation. This is due to the execution of the math functions from the external memory. One of the solutions for this issue is to change linker configuration to place them in the local memory of the Epiphany.

V. Conclusions and future work

Our main concern in this paper was to review the Parallella board in order to follow later with more advanced plans touching on simulations and multi-agent systems. We have reviewed the board internals and its programming model. We then performed some microbenchmarks and tested a genetic algorithm in various versions from a naive implementation to more advanced optimizations.

Our benchmark results show that the Parallella can be fast and offers quite large benefits, but these are destroyed by slow memory transfers and the large number of manual optimizations required to get to them. Comparing only the results for a single-core CPU version and Epiphany "push" versions (as described in Section IV) we can notice that actual computation is nearly 50 times faster on Parallella. However, the work required for memory copying between the device and the host causes the Epiphany version to perform worse in a single iteration than the single-core ARM one. To gain really better results, we need to port the whole algorithm to the Epiphany in order to avoid Epiphany–ARM memory copying. However, porting the code, especially with a very small local memory, requires significant work.

Nevertheless, in the case of offloading only the most costly part of algorithms (for example, fitness computation), the local copying by Epiphany cores using the DMA engine has the best performance among all implementations and would be recommended for simple scenarios.

It is clear from our and others' results that Parallella is not mature enough to replace more advanced and well-established manycore platforms. However, it is an interesting board that can be used to accelerate parallel workloads in a very simple and cheap way. There is nearly no effort required to port programs written in C to the Epiphany compiler in a naive way. Optimization of such programs requires more work, but it is still simple considering the relatively poor feature set of the Software Development Kit and the processor.

Our next steps regarding Parallella will focus on the usage of multiple boards in a single cluster and Java–Epiphany interaction. We see possibilities for deploying island and agent-based models of evolutionary algorithms [4] on such clusters, as they usually can be organized in a way that requires little communication between islands. Separate populations could be computed on different boards with rare synchronization events between them.

The last model, namely Evolutionary Multi-Agent Systems (EMAS) [18], [19], which uses the agent paradigm for decentralizing the process of evolution, is of special interest to us, since it allows achieving a fine-grained parallelism with its implementation of agents [7]. This opens the way for another possible approach, even without using multiple boards – the implementation of multi-agent systems with lightweight agents (or groups of agents) executing on separate cores [20]. The MIMD nature of the Epiphany processor should be a matching architecture for these systems.

Acknowledgment

The research reported in the paper was supported by the grant "Hybrid model of the early detection of internal diseases based on the paradigm of interacting particles and multi-agent system" (No. DEC-2013/09/N/ST6/01011) from the Polish National Science Centre.

References

[1] J. Shalf, J. Bashor, D. Patterson, K. Asanovic, K. Yelick, K. Keutzer, and T. Mattson, "The MANYCORE revolution: will HPC lead or follow," *SciDAC Review*, vol. 14, pp. 40–49, 2009.

[2] A. Olofsson, T. Nordström, and Z. Ul-Abdin, "Kickstarting High-performance Energy-efficient Manycore Architectures with Epiphany," Nov. 2-5, 2014. doi: 10.1109/ACSSC.2014.7094761

[3] G. Chrysos, "Intel® Xeon Phi™ Coprocessor-the Architecture," *Intel Whitepaper*, 2014.

[4] M. Kisiel-Dorohinicki, G. Dobrowolski, and E. Nawarecki, "Agent populations as computational intelligence," in *Neural Networks and Soft Computing: Proceedings of the Sixth International Conference on Neural Networks and Soft Computing, Zakopane, Poland, June 11–15, 2002*, L. Rutkowski and J. Kacprzyk, Eds. Heidelberg: Physica-Verlag HD, 2003, pp. 608–613. ISBN 978-3-7908-1902-1

[5] M. Kisiel-Dorohinicki, "Agent-based models and platforms for parallel evolutionary algorithms," in *Computational Science - ICCS 2004*, M. Bubak, G. D. van Albada, P. M. A. Sloot, and J. Dongarra, Eds. Springer Berlin Heidelberg, 2004, pp. 646–653.

[6] Ł. Faber, K. Piętak, A. Byrski, and M. Kisiel-Dorohinicki, *Agent-Based Simulation in AgE Framework*. Springer Berlin Heidelberg, 2012, pp. 55–83. ISBN 978-3-642-28888-3

[7] D. Krzywicki, W. Turek, A. Byrski, and M. Kisiel-Dorohinicki, "Massively concurrent agent-based evolutionary computing," *Journal of Computational Science*, vol. 11, pp. 153–162, nov 2015. doi: 10.1016/j.jocs.2015.07.003

[8] M. Pietroń, A. Byrski, and M. Kisiel-Dorohinicki, "GPGPU for difficult black-box problems," *Procedia Computer Science*, vol. 51, pp. 1023–1032, 2015. doi: 10.1016/j.procs.2015.05.249

[9] *Epihany SDK Reference, rev. 5.13.09.10*. [Online]. Available: http://adapteva.com/docs/epiphany_sdk_ref.pdf

[10] J. A. Ross, D. A. Richie, S. J. Park, and D. R. Shires, "Parallel Programming Model for the Epiphany Many-Core Coprocessor Using Threaded MPI," in *Proceedings of the 3rd International Workshop on Many-core Embedded Systems*. ACM, 2015. doi: 10.1145/2768177.2768183 pp. 41–47.

[11] A. Papadogiannakis, S. N. Agathos, and V. V. Dimakopoulos, *OpenMP 4.0 Device Support in the OMPi Compiler*. Cham: Springer International Publishing, 2015, ch. OpenMP 4.0 Device Support in the OMPi Compiler, pp. 202–216. ISBN 978-3-319-24595-9

[12] T. Back, D. B. Fogel, and Z. Michalewicz, Eds., *Handbook of Evolutionary Computation*, 1st ed. Bristol, UK, UK: IOP Publishing Ltd., 1997. ISBN 0750303921

[13] M. Matsumoto and T. Nishimura, "Mersenne twister: A 623-dimensionally equidistributed uniform pseudo-random number generator," *ACM Trans. Model. Comput. Simul.*, vol. 8, no. 1, pp. 3–30, Jan. 1998. doi: 10.1145/272991.272995

[14] (2011) Tiny Mersenne Twister (TinyMT): A small-sized variant of Mersenne Twister. [Online]. Available: http://www.math.sci.hiroshima-u.ac.jp/~m-mat/MT/TINYMT/

[15] Andreas Olofsson - Public forum communication. [Online]. Available: https://parallella.org/forums/viewtopic.php?f=9&t=2391&p=13653

[16] *Epihany Architecture Reference, rev. 14.03.11*. [Online]. Available: http://adapteva.com/docs/epiphany_sdk_ref.pdf

[17] A. Varghese, B. Edwards, G. Mitra, and A. P. Rendell, "Programming the Adapteva Epiphany 64-core network-on-chip coprocessor," *International Journal of High Performance Computing Applications*, 2015. doi: 10.1109/IPDPSW.2014.112

[18] A. Byrski and M. Kisiel-Dorohinicki, *Man-Machine Interactions 3*. Cham: Springer International Publishing, 2014, ch. Agent-Based Approach to Continuous Optimisation, pp. 487–494. ISBN 978-3-319-02309-0

[19] A. Byrski, R. Dreżewski, L. Siwik, and M. Kisiel-Dorohinicki, "Evolutionary Multi-Agent Systems," *The Knowledge Engineering Review*, vol. 30, no. 02, pp. 171–186, mar 2015. doi: 10.1017/s0269888914000289

[20] D. Krzywicki, Ł. Faber, A. Byrski, and M. Kisiel-Dorohinicki, "Computing agents for decision support systems," *Future Generation Computer Systems*, vol. 37, pp. 390–400, jul 2014. doi: 10.1016/j.future.2014.02.002

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 873–876
DOI: 10.15439/2016F463
ACSIS, Vol. 8. ISSN 2300-5963

An Overview of Cloud Interoperability

Magdalena Kostoska*, Marjan Gusev*, and Sasko Ristov*†
*University Ss Cyril and Methodius, FCSE, Skopje, Macedonia
Email: {magdalena.kostoska, marjan.gushev}@finki.ukim.mk
†University of Innsbruck, Innsbruck, Austria
Email: sashko@dps.uibk.ac.at

***Abstract*—Unlike the network TCP/IP's and OSI's layered structure of protocols, which allows the independence of protocols of different layers, as well as defining the upper layer protocols through the protocols of the lower layers, the cloud service layers are tightly dependent on each other. For example, an application of the SaaS layer can neither communicate nor exchange data with another application found on the same layer. The goal of this paper is to overview the cloud interoperability and to analyze it as a service model perspective. Several aspects and categories of cloud interoperability are analyzed in this paper.**

***Index Terms*—Migration, interoperability, portability.**

I. Introduction

ADOPTION of cloud rests largely on interoperability and standardization as they define the new age IT industry [1].

The main stakeholders (cloud providers and clients) have opposite motivations for cloud interoperability. The providers prefer vendor lock-in situations to keep the clients and ensure higher profits enabling more and more cloud features. On contrary, the clients would like freedom, and the ability to choose the provider that offers the highest quality of services they want. Therefore, the need for cloud interoperability is more initiated by the clients than the providers.

Cloud interoperability is closely linked with cloud portability [2]. Whenever one analyzes the process of porting data and applications from one cloud provider to another, then the cloud interoperability is the essential problem to be solved. An easy way to port data, application and platform is through transferring an image of the virtual machine (VM) between the providers that use the same cloud environment.

There are a lot of surveys about ongoing initiatives that address the cloud computing interoperability. Petcu [3] discusses the classical problem of too many different approaches by various vendors in the way they realize the interface to the cloud and offer cloud features.

The interoperability problems in cloud computing arise when the clients are trying to exchange data, applications and services between different cloud providers. The identified problems can be classified into the following categories: a) system, initiated by incompatible implementations of cloud virtualization; b) applications, including incompatible application programs and code, c) service, defined by the ability to use various services hosted on different clouds; d) data, initiated by different standards of data presentation. Data interoperability is mainly addressed by other computing areas, whereas, the cloud interoperability addresses mainly the system, application and service interoperability.

This paper observes the cloud interoperability on IaaS, PaaS and SaaS levels analyzing its context from the management, platform or application levels. An approach is introduced to analyze the cloud interoperability *as a service model*.

II. Background

The concept of *interoperability* is not a new one. In the fields of information technology or systems engineering, it has been defined as the ability of two or more heterogeneous elements to not only exchange, but also use the exchanged information (interoperate). However, in the field of Cloud computing, the concept of interoperability is rather new and has recently been an active field of research. In this section we will define and explain all related concepts.

The interoperability can not be uniformly defined - there are very many different definitions which vary in technological aspects, and development frameworks, which can be more general or address only some standard details. Generally, the definition of *interoperability* depends on the context of its application.

IEEE describes the interoperability as a system or product feature to work with other systems or products without additional intervention of the client [4]. According to NIST [5], the cloud interoperability allows seamless exchange and use of data and services among various cloud infrastructure offerings and to use the data and services exchanged to enable them to operate effectively together.

Interoperability can be regarded as the ongoing process of ensuring that the systems, procedures and culture of an organization are managed in such a way as to maximize opportunities for exchange and reuse of information. It includes many areas with its characteristics [6]:

- *Technical interoperability* - development of standards of communication, transport and representation
- *Semantic interoperability* - the use of various different terms to describe similar concepts may cause problems in communication, execution of programmes and data transfers.
- *Political/Human interoperability* - the decision to make resources widely available has implications for organizations, their employees and end-users

978-83-60810-90-3/$25.00

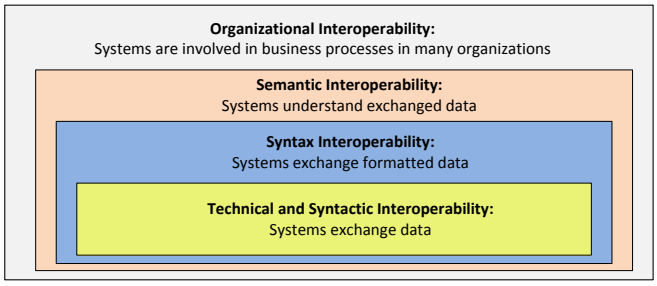


Fig. 1. Interoperability types and their correlation

- *Interoperability of communities or societies* - there is an increasing need to require access to information from a wide range of sources and communities.
- *International interoperability* - in international matters, there are variations in standard, communication problems, language barriers, differences in communication styles, and a lack of common basis.

In most known interoperability frameworks, this term is considered in three levels [7], as follows:

- *Technical interoperability* which includes standards and protocols. This aspect of interoperability covers the technical issues of linking computer systems and services.It includes key aspects such as open interfaces, interconnection services, data integration and middleware, data presentation and exchange, accessibility and security services [8]. Technical interoperability is usually associated with hardware/software components, systems and platforms that enable machine-to-machine communication to take place. This kind of interoperability is often centred on (communication) protocols and the infrastructure needed for those protocols to operate [9].
- *Syntax interoperability* is usually associated with data formats when they are exchanged among systems. Certainly, the messages transferred by communication protocols need to have a well-defined syntax and encoding, even if it is only in the form of bit-tables [9].
- *Semantic interoperability* is concerned with ensuring that the precise meaning of exchanged information is understandable by any other application that was not initially developed for this purpose. Semantic interoperability enables systems to combine received information with other information resources and to process it in a meaningful manner. Semantic interoperability is therefore a prerequisite for the front-end multilingual delivery of services to the client [8].

Additionally, *organizational interoperability* is also discussed, which allows systems to be involved in business processes of multiple organizations [8].

An overview of all interoperability types is given in Fig. 1. The application of interoperability in any domain is usually realized by defining and applying standards. Generally, the goal of interoperability and the standards is the same - to allow exchange and cooperation of computer services. Standards define protocol by which all service suppliers that implement the standards offer structured data and information exchange no matter the inner architectural design or implementation is used for the service.

III. Cloud interoperability

Interoperability in cloud can be considered and defined as a service model, and therefore, we will discuss interoperability of applications, platforms, and management.

Cloud application interoperability addresses the application components, whether they are deployed as IaaS, PaaS, or SaaS. An application component may be a complete monolithic application, or a service as a part of a distributed application. These components invoke respective platforms that implement various communications protocols and data presentation standards; and therefore, can not be used without cloud platform interoperability.

Cloud platform interoperability concerns the platform components, usually deployed as PaaS or IaaS. Information exchange and service discovery requires standard protocols to realize interoperable platforms.

Cloud management interoperability targets the management aspects between various cloud services deployed on SaaS, PaaS, or IaaS levels. Each provider realizes different cloud features and interfaces to manage them, so the clients would prefer to have a unique approach and generic off-the-shelf system management, offered via standard interfaces.

A. IaaS level interoperability

Interoperability on the IaaS level of cloud management implies simple and standardized management of infrastructures of different cloud systems. The management includes instantiating and control of virtual machines, enabling and discovering network characteristics, setting and editing security rules, etc.

This type of interoperability is the best defined when compared to other types. Fig. 2 shows a taxonomy of its basic concepts [10], [11], [12]:

- *Access Mechanism* - defines how a service in cloud may be accessed by users and/or software developers,
- *Virtual Resources* - service delivery as a complete software stack of installing a virtual machine,
- *Network* - addressing and API,
- *Storage* - management and organization of storage,
- *Security* - authentication, authorization, user accounts and encryption,
- *Service-Level Agreement* - architecture format, monitoring, and
- *Other*.

B. PaaS level interoperability

Interoperability on the PaaS level implies simple exchange of data and services among different platforms hosted on different infrastructures on cloud, and their effective reuse without extra effort on part of the user.

When analyzing the data exchange, one can consider data compatibility among different platforms, such as if numbers

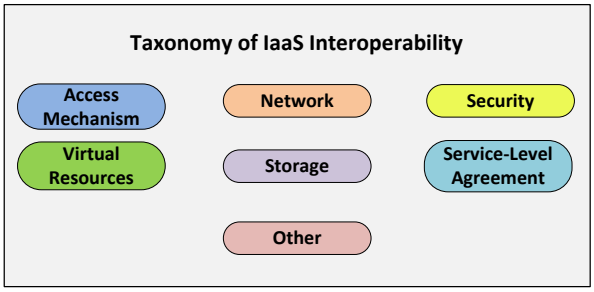


Fig. 2. Taxonomy of IaaS interoperability [10]

are to be transferred, then Little or Big Endian mode should be preserved, or a special function enabled to realize an easy transformation between the formats prior to transfer.

Analysis of the interoperability of services to be hosted in different cloud platforms rises the question of portability. For example, to transfer a service from one cloud to another that uses a different platform, initiates a lot of portability problems. If the origin and target clouds use the same environment, then a simple packing and copying procedure can be used to realize the porting process.

In case of different platforms on the origin and target clouds, one has to start a different transfer procedure that consists of packing, copying, instantiating, installing, deployment and customization to enable an interoperability. Still, there are open issues that address the way services interact with others, if there are additional cloud services invoked by the origin cloud that are not supported by the target cloud, or any dependence on a specific operating system hosted in the origin cloud.

C. SaaS level interoperability

Interoperability on the SaaS level of cloud applications implies simple exchange of data and services among different applications hosted on different platforms and infrastructures on cloud, and their effective reuse without extra effort on part of the user. Additionally, this type of interoperability can be considered from different application domains.

Interoperability on the level of applications is first defined by Kumar et al. [13] in 2010. According to their definition, interoperability can be considered in four categories (Fig. 3):

- Interoperability among applications in the same cloud,
- Data exchange and operation calls in applications on different cloud-computing environments
- Software programs that are distributed in different cloud environments and integrate data and applications in cloud in a unified way, and
- Migration of applications from one cloud environment to another.

When a client switches between two cloud providers on the SaaS level does not involve porting the applications and services, rather it involves exchange of structured data.

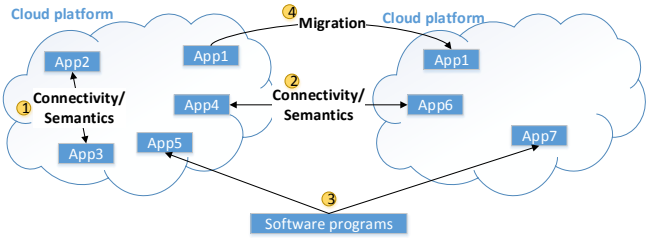


Fig. 3. Interoperability types in cloud applications [13]

TABLE I
DEVELOPMENT STATUS OF EACH ASPECT OF INTEROPERABILITY OF CLOUD COMPUTING

Context	Layer	Developing standards
Management	IaaS	OCCI, CIMI, UCI
Platform	PaaS	Stub
Application	SaaS	mOSAIC

The cloud interoperability is probably the most significant to address the compatibility of exchanged data on the functional level, it means not just to transfer the structured data, but also all relations between them. So far, no interface has been developed to allow such an interoperability. The problem is mainly manifested on the definition on functional level of realization the interface to the application.

Most of the research in the area of SaaS level, and even on the PaaS level is limited by the support of the vendors. Usually, the vendors prefer to lock-in the customer to its cloud and do not cooperate in the efforts to support the interoperability on this level.

IV. DISCUSSION

Table I presents the current development stage for each of the perspectives (i.e. categories), as well as, mapping of the use cases to the perspectives.

One can notice in Table I that certain aspects are more developed than other (i.e. interoperable management of virtual machines and application portability). We can also notice that the Platform context is least developed.

Large number of developing standards has arisen during the past few years:

- OCCI - The Open Cloud Computing Interface standard represents protocol and API for all kinds of IaaS management tasks [14]
- CIMI - Cloud Infrastructure Management Interface standard represents an interface for management of cloud services and the operations and attributes [15]
- UCI - Unified Cloud Interface concept aim to provide a unified interface for entire infrastructure stack using semantic technology [16]
- mOSAIC - The mOSAIC platform and engine enables deployment, configuration and management of applications using semantic technology [17]

- OVF - Open Virtualization Format standard provides open and platform-independent packaging format for software solutions based on virtual systems [18]
- CAMP - Cloud Application Management for Platforms aims to standardizing cloud PaaS management API [19]
- TOSCA - The Topology and Orchestration Specification for Cloud Applications aims to standardize application description in order to provide portability and management [20]. We introduced the extension - P-TOSCA, which handles several TOSCA weaknesses and ambiguities [21]. The demo applications for automated portability with P-TOSCA are developed for porting a SOA application [22] and an N-tier application [23].
- OData - The Open Data Protocol enables service creation to publish, share and edit resources via HTTP [24]
- CDMI - The Cloud Data Management Interface standard defines interface for creation, retrieval, update and deletion of data elements from the Cloud [25]

V. Conclusion

This paper gives an overview of the cloud interoperability on different service layers analyzed from cloud management, platform and application aspects *as a service* model. The cloud interoperability should be considered as cloud management on IaaS layer, exchange of data and services among different platforms (on PaaS layer) hosted on different infrastructures and exchange of structured data (on SaaS layer) among different applications deployed on platforms hosted on different infrastructures.

Although there are several standards and solutions on the data presentation level (data formats and communication protocols), still there are open issues on interoperability on systems and applications and there are no solutions when one wants to exchange structured data between providers. One can conclude that cloud interoperability on IaaS and PaaS levels has been addressed and several partial solutions exist, while the cloud interoperability on the SaaS level is still in an infant development.

Acknowledgment

This work was partially financed by the Faculty of Computer Science and Engineering at the "Ss. Cyril and Methodius" University, Skopje, Macedonia.

References

[1] A. Parameswaran and A. Chaddha, "Cloud interoperability and standardization," *SETlabs briefings*, vol. 7, no. 7, pp. 19–26, 2009.

[2] M. Kostoska, M. Gusev, and S. Ristov, "An overview of cloud portability," in *Future Access Enablers for Ubiquitous and Intelligent Infrastructures*, ser. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, V. Atanasovski and A. Leon-Garcia, Eds. Springer International Publishing, 2015, vol. 159, pp. 248–254. ISBN 978-3-319-27071-5. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-27072-2_32

[3] D. Petcu, "Portability and interoperability between clouds: challenges and case study," in *Towards a Service-Based Internet*. Springer, 2011, pp. 62–74. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-24755-2_6

[4] IEEE, "610.7-1995 Standard Glossary of Computer Networking Terminology," *Available online: http://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=3284*.

[5] National Institute of Standards and Technology, "NIST cloud computing standards roadmap," *Available online: http://www.nist.gov/itl/cloud/upload/NIST_SP-500-291_Version-2_2013_June18_FINAL.pdf*, Jul. 2013.

[6] UKOLN, "Looking at interoperability," *Available online: http://www.ukoln.ac.uk/interop-focus/about/flyer-interoperability.pdf*.

[7] H. Kubicek and R. Cimander, "Three dimensions of organizational interoperability," *European Journal of ePractice*, vol. 6, 2009.

[8] European Union, "European interoperability framework for Pan-European eGovernment services," *Available online: http://ec.europa.eu/idabc/servlets/Docd552.pdf?id=19529*.

[9] H. van der Veer and A. Wiles, "Achieving technical interoperability," *European Telecommunications Standards Institute*, 2008.

[10] Z. Zhang, C. Wu, and D. W. Cheung, "A survey on cloud interoperability: taxonomies, standards, and practice," *ACM SIGMETRICS Performance Evaluation Review*, vol. 40, no. 4, pp. 13–22, 2013. doi: 10.1145/2479942.2479945. [Online]. Available: http://doi.acm.org/10.1145/2479942.2479945

[11] R. Prodan and S. Ostermann, "A survey and taxonomy of infrastructure as a service and web hosting cloud providers," in *Grid Computing, 2009 10th IEEE/ACM International Conference on*. IEEE, 2009. doi: 10.1109/GRID.2009.5353074 pp. 17–25.

[12] B. P. Rimal, E. Choi, and I. Lumb, "A taxonomy and survey of cloud computing systems," in *INC, IMS and IDC, 2009. NCM'09. Fifth International Joint Conference on*, 2009. doi: 10.1109/NCM.2009.218 pp. 44–51.

[13] B. Kumar, J. C. Cheng, and L. McGibbney, "Cloud computing and its implications for construction IT," in *Computing in Civil and Building Engineering, Proceedings of the International Conference*, vol. 30, 2010, p. 315.

[14] Open Grid Forum, "OCCI, Open Cloud Computing Interface," *Available online: http://occi-wg.org/*, 2011. [Online]. Available: http://occi-wg.org/

[15] Distributed Management Task Force, "Cloud infrastructure management interface (CIMI) model and RESTful HTTP-based protocol," *Available online: http://dmtf.org/sites/default/files/standards/documents/DSP0263_2.0.0c.pdf*, Mar. 2015.

[16] Google, "Unified cloud," *Available online: http://code.google.com/p/unifiedcloud/wiki/ UCI_Architecture*, 2010.

[17] F. Moscato, R. Aversa, B. Di Martino, T. Fortis, and V. Munteanu, "An analysis of mosaic ontology for cloud resources annotation," in *Computer Science and Information Systems (FedCSIS), 2011 Federated Conference on*, Sept 2011, pp. 973–980.

[18] Distributed Management Task Force, "Open virtualization format specification version 2.1.0," *Available online: http://www.dmtf.org/sites/ default/files/standards/ documents/DSP0265_1.0.0.pdf*, Jan. 2014.

[19] M. Carlson, M. Chapman, A. Heneveld, S. Hinkelman, D. Johnston-Watt, A. Karmarkar, T. Kunze, A. Malhotra, J. Mischkinsky, A. Otto *et al.*, "Cloud application management for platforms," *Available online http://cloudspecs. org/camp/CAMP-v1.0.pdf*.

[20] R. Han, M. M. Ghanem, and Y. Guo, "Elastic-TOSCA: Supporting elasticity of cloud application in TOSCA," in *CLOUD COMPUTING 2013, The Fourth International Conference on Cloud Computing, GRIDs, and Virtualization*, 2013, pp. 93–100.

[21] M. Gusev, M. Kostoska, S. Ristov, and A. Donevski, "P-TOSCA portability of SOA applications," in *Proceedings of 5th International Conference on Cloud Computing and Service Science (CLOSER)*, Lisbon, Portugal, 2015, pp. 71–78. [Online]. Available: http://closer.scitevents.org/?y=2015

[22] S. Ristov, M. Kostoska, and M. Gusev, "P-TOSCA portability demo case," in *2014 IEEE 3rd International Conference on Cloud Networking (IEEE CLOUDNET)*, 2014. doi: 10.1109/CloudNet.2014.6969002 pp. 269–271.

[23] M. Gusev, M. Kostoska, and S. Ristov, "Cloud P-TOSCA porting of N-tier applications," in *Proceedings of the 22nd International TELFOR Forum, IEEE Conference Publications*, 2014. doi: 10.1109/TELFOR.2014.7034559 pp. 935–938.

[24] OASIS, "Open Data Protocol (OData) 4.0," Feb. 2014, http://docs.oasis-open.org/.

[25] SNIA, "Cloud Data Management Interface (CDMI) v1.1.1," *Available online: http://www.snia.org/sites/default/files/ CDMI_Spec_v1.1.1.pdf*, Mar. 2015.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 877–881
DOI: 10.15439/2016F486
ACSIS, Vol. 8. ISSN 2300-5963

The Column-oriented Data Store Performance Considerations

Artur Nowosielski[1,2]

[1] PhD Candidate
Systems Research Institute
Polish Academy of Sciences
ul. Newelska 6, 01-447 Warsaw, Poland
Email: artnowo@gmail.com

[2] Findwise Sp. z o.o.
ul. Wspólna 35/16, 00-519 Warsaw, Poland

Piotr A. Kowalski[3,4], Piotr Kulczycki[3,4]

[3] Faculty of Physics and Applied Computer Science
AGH University of Science and Technology
al. Mickiewicza 30, 30-059 Cracow, Poland
Email: {pkowal,kulczycki}@agh.edu.pl

[4]Systems Research Institute
Polish Academy of Sciences
ul. Newelska 6, 01-447 Warsaw, Poland
Email: {pakowal,kulczycki}@ibspan.waw.pl

***Abstract*—The massive amounts of data processed by information systems raise the importance of detailed database performance analysis. Column-oriented data stores are becoming increasingly popular in big data appliances. This paper identifies database performance factors on the basis of empirical studies on a custom implementation. To summarize the research, a simple performance mathematical model has been created.**

I. Introduction

THIS article is the result of experiments performed on a custom column-oriented database management system. Performance studies presented in this paper are a part of a broader research initiative about an optimization of columnar data store sharding with the use of a natural computing algorithms. The research has been overviewed in the previous FedCSIS conference paper. [1] Performance modeling and discovering what determines a single database's behaviour is considered as the first step towards creating a more sophisticated partitioned database model. Such a model will be a subject of optimization by metaheuristic algorithms. Columnar data store performance studies are considered important because non-relational data stores are gaining popularity and more and more applications. However, the goal of this research is not a comparative study of the column-oriented database model versus other data models, nor CODB custom implementation versus other columnar data stores. It aims to discover which factors determine performance and the relationships between these factors' impact. Absolute values are considered less important in this study since it is intended to be a foundation of the weights (importance) estimation of the specific factors.

The paper is structured as follows. Section II contains a general description of a column-oriented database model relatively to the relational model for the sake of better understanding, and the custom implementation of columnar data store is introduced. The next section III presents how the performance of a Java application (specifically a database) can be measured and expressed. The second half of the article comprises of sections IV, V and VI, which include assumptions and conclusions brought on the basis of the experiments' results.

Because the research and benchmarking is performed on a custom database management system, the article has relatively unique character. However, available literature offers examples of research initiatives driven by similar ideas, such as [2] or [3]. Since this is a short paper, it presents a general overview and the most important facts of the research.

II. Column-oriented DBMS

The column-oriented database model has been around for almost as many years as the most popular, row-oriented relational model by E.F. Codd [4]. One of the first concepts regarding column-oriented storage are transposed files databases from late 1970s [5].

After an initial rush, columnar databases remained in their own, narrow niche for more than 20 years, while the relational model dominated a majority of applications. Relational model's strength came from the strong mathematical foundation based on the set algebra and focus on the data consistency and reliability. Despite this, in the recent decade, alternative models have gained more interest in the commercial world with the rise of the non-relational database trend. Databases that focus on other aspects than traditional Relational Database Management Systems (RDBMS) started gaining more attention. This trend has been called the *not only SQL* or *NoSQL* and was one of the outcomes of a rise of interactive, especially social, web services within the *web 2.0* movement. [6] The most significant developments in the area of columnar data stores are the C-Store [2] [7] and MonetDB [7].

Viewed from some angles, it can be said the column-oriented model is essentially only a physical-tier modification of the relational model. [3] However, experiments prove that implementing adequate modifications to the row-oriented Database Management System (DBMS) storage tier is not enough to rival the columnar store in some applications [8]. Nevertheless, some column database management systems offer roughly the same interface as the relational ones, hiding all the internal differences [9]. Certainly, there are analogies between fundamental terms of both domains. Keyspace parallels a database from the RDBMS domain. Relational table (or

978-83-60810-90-3/$25.00

relation) is roughly the same as a column family in columnar stores. Correlation of values with the same key from different columns in columnar store leads to construction of a tuple, which is comparative to a record in the relational database.

Despite obvious analogies, there is a fundamental difference in the storage structure. In row-oriented stores there is a single data file for the whole table and data is stored in a tuple-by-tuple manner where tuples are stored column-by-column. Thanks to this, it is cheap to iterate over records and read all the values in order to construct a complete tuple. Moreover, it is cheap to write the whole record at once by just appending or replacing another entry. However, this architecture has significant disadvantages too. Firstly, the table schema is effectively immutable - changing the column set in a table requires rebuilding the whole data file and all of the indices. Secondly, iterating through values in a single column from all the records requires long jumps within a data file, which involves huge I/O overhead.

These issues are addressed by column-oriented architecture. In a classical approach with explicit record ID storage [7], each column is a sequence of id-value pairs stored in a separate file. Database schema is flexible, columns can be freely added or removed and if a given record does not have value in a given column, it does not take any space to record such an information. Iteration over every value in a column is simple, so that columnar stores are effective in terms of generating aggregations, summaries and other read-intensive applications [3] [9]. But this schema also has a significant drawback. It is expensive to collect all the distinctive elements of a record altogether. Such an operation requires searching of all the data types (columns).

Taking this into consideration, in a domain of columnar data stores, a concept of the record does not actually exists. At the lower tiers of the system, there is no such concept and a specific values in a *record* can be associated to others only by matching the row id. This is the most fundamental conceptual difference between the two models. The columnar model storage is organized around pieces of information of the same type, not of the same entity.

This paper relies solely on the CODB database management system. The implementation has been partially reviewed in [1]. Database logic is comprised of execution of a set of processes touching a physical storage tier. A given part of an algorithm is considered as a significant in terms of performance when its execution time or resource consumption depends on any external factor. For example, looking for a value in the value storage file depends on a number of currently stored values in the set, whereas appending new value in some cases does not, in case when it is performed at EOF (end of file - the very last possible offset in a file), given that the EOF can be obtained instantly from the filesystem. In order to decrease space waste and speedup operations, a concept of storage maps has been implemented. Storage maps are responsible for keeping track of free space chunks. When any operation needs to allocate a bit of space, it asks the storage map first, and jumps to EOF only if free space does not offer an adequate fragment. For the same reason, a key storage data file operates on value hashes instead of actual values. Hashing overhead is orders of magnitude lesser than the potential overhead caused by moving around bigger pieces of data.

III. Performance analysis

A very important aspect of a performance measurement is an overhead. In conformity with common sense and intuition, it must be predictable and have a minimal possible impact on the measurement's result.

Another critical concern, when it comes to software performance measurement, is concurrency and parallelism. In a classical, single-threaded sequential program execution, the matter is trivial. The execution time is proportional to a cycle per instruction (CPI) value, whereas CPI is an inversion of the instructions per cycle (IPC) value with constant, known cycle time. [10] In multi-threaded or parallel conditions none of these assumptions are true. At the time of writing this paper, the test CODB system operates in a single thread for the most of the time. The only multi-threaded parts are Java parallel streams used for processing some of the internal collections. This does not affect database logic, which is discussed in this paper. Taking this into consideration, sequential processing measurement techniques were used.

There are many execution parameters which can be measured. For the sake of the research, the following parameters were chosen: execution time, CPU workload and a heap size. Such a selection lets us to take two important perspectives of the system's performance: the user view (time-oriented) and the system view (resource-oriented). The user perspective is connected to considering system as efficient, it determines system's capacity as well. The faster requests are processed, the more of them can be handled in a unit of time. This aspect is particularly important in interactive applications, requiring fast responses for a massive amount of requests. A fundamental time metric is the execution time. The other perspective, a system one, is resource-oriented. Resource is a part of a system, which serves for the other elements of the same system. [11]

Metric is defined as a way to determine whether a system has given property or not and to what extent. Specifically, in the performance engineering domain, metrics provide information about performance parameters with regards to time and amount of computational work. Application context is crucial for interpretation of a metric's result. For time-sensitive applications, like real-time systems, time domain is fundamental, whereas applications processing huge volumes of data will put a pressure on throughput, regardless of resource usage.

In the literature of the subject there is no generally-adopted standard on performance metrics. They are rather ad hoc, defined for each application or class of applications. However, there are common types of metrics used, compliant with two major perspectives mentioned earlier. In this research, the following metrics have been used: Response time [ms] - the total time of a request execution; Throughput (capacity) - the number of requests completed in a given time unit; Resource

consumption [MB] - the highest JVM heap size during the test execution. In terms of scalability, the most important metric is the capacity.

A custom, configurable workload generator has been used. It enabled the configuration of the following aspects of the generated requests:

- read/write requests ratio;
- data granularity: defined or random size of entries;
- data entropy: a number of generated values or a unlimited randomness;
- whether the generated requests should touch many columns or not;

CODB Benchmark application configures a dedicated loggers for the sake of collecting execution time, throughput and heap measurements. Measurement data is written in CSV (comma-separated values). Execution time is calculated as a difference between two consecutive `System.nanoTime()` invocations, one just before measured method invocation and one just after. JVM heap consumption is measured using the `java.lang.Runtime` class.

IV. Performance factors

This section presents a set of factors which are suspected to impact on database performance. They were chosen based on expertise, but do not necessarily have a real, significant influence. Verification of these suspicions is the key objective of benchmarking.

First category of performance factors are the universal factors, related to data processed by a database, but unrelated to specific technology concerns.

Read and write operations ratio can determine a lot of performance-impacting elements. Firstly, write operations require synchronization effort. For obvious reasons, insert, update or delete requests imply I/O operations, which need to be enqueued and buffered on many tiers of the system down to the physical layer, where they are really executed.

Assuming a constant amount of data to be written; fewer, big portions are expected to be processed faster than a larger number of small portions. With the use of a benchmark, it needs to be measured to what extent such a prediction is true.

Due to CODB storage file structure, which is essentially a RLE-like-compressed structure, entropy is predicted to have a significant influence on performance of the system. RLE, or run-length encoding, is a simple way of data compression based on substitution of numerous occurrences of a term with a single instance and a number value which represents the number of occurrences. The lesser the entropy is, the better performance should be, because as stated before, when a single column is discussed, appending a new entry with an already existing value may require as little as increasing a single 8B counter and writing a 16B key.

Requests are issued by multiple threads, but the threshold on which synchronization and context switching between threads becomes bigger than concurrency performance gain is unknown. Benchmarking may provide a reasonable empirical information regarding how many threads is too many and, specifically, how that number is related to the CPU's core number and the CPU's hyper threading capabilities.

Besides the technology-independent factors, performance can be affected by technology-specific components. From the wide variety of candidates, two were chosen for the research as potentially having the biggest impact on results.

The Java Virtual Machine platform, and thus the Java programming language, memory model is based on indirect memory management. The application does not allocate and release the memory occupied by the objects on its own, but it is done implicitly by the JVM within a part of the memory called the heap. A critical component of the memory management facility in Java is the garbage collector, a module responsible for removing unused objects from heap. According to the official documentation [12], the HotSpot JVM v.1.8 provides four garbage collector implementations: serial, parallel, concurrent mark-sweep (CMS) and G1.

As CODB is executed on the JVM, it's state may (or may not) have an impact on performance. For example, some internal data collections or buffers are expanded exponentially, so that at the beginning (*cold* state) it will happen more frequently than later (*hot* VM). Hot tests are performed by issuing 1000 write/read requests before starting measurement. After the warm-up all the database internal structures are cleared in order to avoid performance impact by having prefilled collections or buffers.

V. Results and discussion

Each test was performed 4 times and consisted of issuing 50 000 requests. Tables I, II and III present the measurement results. In all the tables, extreme values which are to be discussed further are highlighted. Every table header contains information about what values are desired (low or high).

Each section in result tables contains results with different values of a single factor. Unless a given factor is tested, the following values were stated as defaults for each test: parallel garbage collector, hot JVM, r/w ratio = 0.5, low entropy and two columns in use.

Testing environment was: Intel Core i7-4600U CPU with 12GB DDR3-1600 RAM and SSD drive with the ext4 filesystem. The operating system was a 64-bit GNU/Linux 4.2.0 with Oracle HotSpot 64-bit JVM v1.8.0-74. Benchmark was started using Maven Exec Plugin.

Garbage collector implementation has very low impact on request execution time. Results for all the implementations are similar and of similar stability (almost the same standard deviation). Serial GC performed the best, probably because of relatively low data volume and single threaded testing. Stopping a single thread is less harmful in terms of performance than stopping multiple threads. For heap usage levels GC has an obviously fundamental impact, although some patterns are visible here. Parallel GC achieved the lowest minimum heap size of as little as 15MB. This result probably is an outlier, because of distance from all the other implementations and needs a deeper investigation. The highest standard deviation also may be skewed by one or more outliers. Differences in terms of

TABLE I
BENCHMARKING RESULTS: REQUEST EXECUTION TIME (THE LOWER THE BETTER)

Factor	Name	Min[ms]	Max[ms]	Avg[ms]	Std deviation	Impact
GC type	serial	0.004	186.739	0.060	0.483	low
	parallel	0.004	189.110	0.067	0.654	
	CMS	0.004	188.975	0.059	0.485	
	G1	0.004	188.600	0.059	0.633	
JVM state	hot	0.004	189.110	0.067	0.654	immaterial
	cold	0.004	189.188	0.0623	0.816	
Read/write ratio	only reads	0.002	**11.171**	**0.006**	0.036	high
	50% writes	0.004	189.110	0.067	0.654	
	only writes	**0.027**	**189.935**	**0.087**	0.728	
Data entropy	low	0.004	189.110	0.067	0.654	high
	high	**0.001**	**30.246**	0.060	0.306	
Multicolumn	yes	0.004	189.110	0.067	0.654	low
	no	0.004	183.657	0.062	0.480	

TABLE II
BENCHMARKING RESULTS: HEAP USAGE (THE LOWER THE BETTER)

Factor	Name	Min[MB]	Max[MB]	Avg[MB]	Std deviation	Impact
GC type	serial	29.50	**425.52**	116.16	81.96	high
	parallel	**14.98**	417.96	124.32	**103.20**	
	CMS	28.97	347.42	**141.60**	76.67	
	G1	30.25	266.02	135.67	64.61	
JVM state	hot	14.98	417.96	124.32	103.20	high
	cold	17.73	236.70	93.38	62.57	
Read/write ratio	only reads	**61.77**	125.05	93.40	**28.80**	high
	50% writes	14.98	**417.96**	124.32	103.20	
	only writes	17.04	373.58	139.91	97.91	
Data entropy	low	14.98	417.96	124.32	103.20	high
	high	14.12	**142.84**	**58.10**	**48.46**	
Multicolumn	yes	14.98	417.96	124.32	103.20	high
	no	14.89	**300.54**	**89.75**	67.85	

TABLE III
BENCHMARKING RESULTS: THROUGHPUT (THE HIGHER THE BETTER)

Factor	Name	Min[req/sec]	Max[req/sec]	Avg[req/sec]	Std deviation	Impact
GC type	serial	**10518**	18237	12771	2373	moderate
	parallel	**6986**	17600	11475	3125	
	CMS	8492	18608	12891	2707	
	G1	8834	**18924**	**13286**	2825	
JVM state	hot	6986	17600	11475	3125	low
	cold	5920	18975	12095	3380	
Read/write ratio	only reads	**41400**	**50000**	**47850**	3724	high
	50% writes	6986	17600	11475	3125	
	only writes	**5491**	**14693**	**9617**	3153	
Data entropy	low	6986	17600	11475	3125	low
	high	**10455**	**13940**	**12045**	**1337**	
Multicolumn	yes	6986	17600	11475	3125	low
	no	**8699**	16381	**12277**	2708	

maximum and average recorded usage are much more stable. Execution with the serial GC consumed the highest amount of memory, which seems to be a trade-off of its simplicity and speed. In terms of request processing throughput, Serial GC has the best minimum recorded throughput, but the true winner is the G1 GC, which offered the highest maximum result and the highest average throughput. Results are similar and the impact is relatively low, though.

The JVM state variable shows a little impact on performance in both time-oriented metrics. It has much higher impact on the heap usage levels, but that would probably make sense - the longer the program is running, the higher is heap usage. In connection with similarly low standard deviation, it brings a conclusion that the JVM state is not very important for the performance. In general, technology-specific factors turned out to have much less impact on a database performance.

When it comes to the data oriented metrics, their impact is much more visible in performance results. Read-write ratio, according to intuition, showed that read operations are performed much faster than write operations. A clearly visible pattern is present in both time-oriented metrics. Starting from the 50% share of write operations, results are stable. This may represent the logarithmic dependency. In terms of heap usage, 100% read pattern showed surprisingly high minimum recorded usage, which may be an outlier. The highest maximum usage was registered for 50-50 pattern. Probably the pattern here is the higher diversity of objects, the higher memory usage is. This requires a further investigation, but is not very important in the research context. Relatively low standard deviation in time-oriented metrics shows that results

are quite stable.

Data entropy presents very interesting case in the request execution time results. The request processed in 0.001ms may be a previously discussed corner case with the insertion of a new instance of already existing value that requires to only append 16B and increment a single long variable. Enriching the execution time log with additional information about request type would help to verify this prediction. High entropy resulted also in a very good maximum request processing time. The reason is probably again an appending at EOF. High entropy tests also proved to have much more stable result than low entropy tests. In terms of heap usage, pattern is similar, high entropy has much better and much more stable results. Throughput metric repeats discoveries from request execution time, as these metrics are related to each other, in a sense. The entropy results are surprising, and require a deeper data mining in order to bring more specific conclusions on its impact.

The last analyzed variable was a multi-column vs. single-column mode. Single column mode showed better results in all of the metrics, although the impact on request execution time results is relatively low. Influence on throughput is moderate, whereas the highest impact is put on the heap usage levels. Standard deviations are similar, but relatively high.

In the context of a horizontal scalability, especially important for this research, data-oriented observations are viable. An optimum case of a single-column with a high entropy and a relatively low share of write operations emerges from the results.

VI. Performance model

In order to produce the mathematical model, chosen factors presented in sections IV and V needs to be converted to a mathematical value. The model is necessary to estimate a performance of a database instance with the specific parameters. This is the first approach to a simplified model and it is to be refined in subsequent work. Weights are determined on the basis of impact displayed in tables I and III. Weights sum must be equal to 1. Other variables were considered immaterial and thus are not present in the formula. Garbage collector implementation also was skipped, as its impact on time-oriented metrics is much lower than data-oriented factors. P metric estimates database instance performance, the higher is the better.

$$P = 0.2 * MC + 0.4 * E + 0.4 * \frac{1}{ln(RW + 1.1)} \quad (1)$$

MC represents a single- or multi-columnar mode. At the moment it is defined as a binary factor with values of 0 for a multicolumn, and 1 for a single column mode. In future research the MC factor may need to be refined to a functional form, depending on the number of columns involved. E takes a values from range $[0, 1]$ where 0 is using the same value all the time and 1 means a total randomness. RW is a ratio between the read and write operations, in range $[0, 1]$ where 0 is the read only and 1 write only. Taking everything into consideration, P metric can take values from range approx $[1.35, 4.80]$. The lowest value is a write-only, multi-column instance with a low entropy, whereas the highest is achieved for a single-column read-only instance with a high entropy. This model will be validated and enhanced during the further work.

VII. Summary

This paper is the very first phase of a performance analysis of column-oriented database management system. Column-oriented databases were described in details, in relation to the popular relational model. Some of the CODB implementation details were presented, putting special emphasis on the data layout. Then, performance engineering concerns were reviewed along with performance metrics. Consecutively, different components, both technological and data-originated, with potential influence on performance results were discussed and verified. Finally, a first approach to a database mathematical performance model was created and discussed, on the basis of the results.

References

[1] A. Nowosielski, P. A. Kowalski, and P. Kulczycki, "The column-oriented database partitioning optimization based on the natural computing algorithms," in *2015 Federated Conference on Computer Science and Information Systems, FedCSIS 2015, Łódź, Poland, September 13-16, 2015*, 2015. doi: 10.15439/2015F262 pp. 1035–1041. [Online]. Available: http://dx.doi.org/10.15439/2015F262

[2] M. Stonebraker, D. J. Abadi, A. Batkin, X. Chen, M. Cherniack, M. Ferreira, E. Lau, A. Lin, S. Madden, E. O'Neil, P. O'Neil, A. Rasin, N. Tran, and S. Zdonik, "C-store: a column-oriented DBMS," *VLDB Conference*, pp. 553–564, 2005. doi: 10.1007/BF02443652. [Online]. Available: http://dl.acm.org/citation.cfm?id=1083592.1083658

[3] A. Lübcke, "Challenges in workload analyses for column and row stores," *CEUR Workshop Proceedings*, vol. 581, pp. 5–8, 2010.

[4] E. F. Codd, "A relational model of data for large shared data banks," *Commun. ACM*, vol. 26, no. 6, pp. 64–69, 1983. doi: 10.1145/357980.358007. [Online]. Available: http://dx.doi.org/10.1145/357980.358007

[5] P. Svensson, "On search performance for conjunctive queries in compressed, fully transposed ordered files," in *Very Large Data Bases, 1979. Fifth International Conference on*. IEEE, 1979, pp. 155–163.

[6] K. Grolinger, W. a. Higashino, A. Tiwari, and M. A. Capretz, "Data management in cloud environments: NoSQL and NewSQL data stores," *Journal of Cloud Computing: Advances, Systems and Applications*, vol. 2, p. 22, 2013. doi: 10.1186/2192-113X-2-22. [Online]. Available: http://www.journalofcloudcomputing.com/content/2/1/22

[7] D. Abadi, "The Design and Implementation of Modern Column-Oriented Database Systems," *Foundations and Trends{®} in Databases*, vol. 5, no. 3, pp. 197–280, 2012. doi: 10.1561/1900000024. [Online]. Available: http://dx.doi.org/10.1561/1900000024

[8] D. J. Abadi, S. R. Madden, and N. Hachem, "Column-Stores vs. Row-Stores: How Different Are They Really?" *Sigmod*, vol. June 9-12, pp. 967–980, 2008. doi: 10.1145/1376616.1376712

[9] D. J. Abadi, D. S. Myers, D. J. DeWitt, and S. R. Madden, "Materialization strategies in a column-oriented DBMS," in *Proceedings - International Conference on Data Engineering*, 2007. doi: 10.1109/ICDE.2007.367892. ISBN 1424408032. ISSN 10844627 pp. 466–475.

[10] S. Eyerman and L. Eeckhout, "System-level performance metrics for multiprogram workloads," *IEEE Micro*, vol. 28, no. 3, pp. 42–53, 2008. doi: 10.1109/MM.2008.44

[11] M. Woodside, G. Franks, and D. C. Petriu, "The Future of Software Performance Engineering," in *Future of Software Engineering (FOSE '07)*, 2007. doi: 10.1109/FOSE.2007.32. ISBN 0-7695-2829-5 pp. 171–187. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4221619

[12] Oracle, "Java Platform, Standard Edition HotSpot Virtual Machine Garbage Collection Tuning Guide," 2016.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 883–888
DOI: 10.15439/2016F91
ACSIS, Vol. 8. ISSN 2300-5963

Big Data Techniques, Systems, Applications, and Platforms: Case Studies from Academia

Atanas Radenski*, Todor Gurov†, Kalinka Kaloyanova‡‖, Nikolay Kirov§‖
Maria Nisheva‡‖, Peter Stanchev¶‖, and Eugenia Stoimenova†‖

* Schmid College of Science and Technology, Chapman University
Orange, CA, USA, Email: radenski@chapman.edu
† Institute of Information and Communication Technologies
Bulgarian Academy of Sciences, Sofia, Bulgaria
Email: (gurov,jeni)@parallel.bas.bg
‡ Faculty of Mathematics and Informatics, Sofia University, Sofia, Bulgaria
Emails: (kkaloyanova,marian)@fmi.uni-sofia.bg
§ Department of Informatics, New Bulgarian University, Sofia, Bulgaria
Email: nkirov@nbu.bg
¶ Department of Computer Science, Kettering University, Flint, MI, USA
Email: pstanche@kettering.edu
‖ Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Sofia, Bulgaria
Emails: (nkirov,stanchev)@math.bas.bg

***Abstract*—Big data is a broad term with numerous dimensions, most notably: big data characteristics, techniques, software systems, application domains, computing platforms, and big data milieu (industry, government, and academia). In this paper we briefly introduce fundamental big data characteristics and then present seven case studies of big data techniques, systems, applications, and platforms, as seen from academic perspective (industry and government perspectives are not subject of this publication). While we feel that it is difficult, if at all possible, to encapsulate all of the important big data dimensions in a strict and uniform, yet comprehendible language, we believe that a set of diverse case studies – like the one that is offered in this paper – a set that spreads over the principal big data dimensions can indeed be beneficial to the broad big data community by helping experts in one realm to better understand currents trends in the other realms.**

***Index Terms*—Metadata, semantic annotations, Spark, NoSQL, data-intensive applications.**

I. Introduction

The principle dimensions of big data include its defining characteristics (such as volume, velocity, variety, and veracity), techniques (such as data mining, machine learning, natural language processing, neural networks, clustering, pattern recognition, sentiment analysis, predictive modeling, supervised learning, time series analysis, to mention a few), software systems (such as Hadoop, Spark, NoSQL DBMSs), applications (such as business analytics, marketing, healthcare, research, performance optimization, security, law enforcement, transportation, and many others), computing platforms (such as clusters, NUMA in-memory database servers, and cloud computing platforms), and big data milieu (such as industry, government, academia).

The big data dimensions are not only broad but also in perpetual change. This is why the task of compiling and maintaining a specification that is rigorous yet comprehensible seems impractical. Instead, we believe that reports like this one, presenting case studies with broad coverage of the big data realm can be beneficial for the broad big data community. Our broad collection of case studies can potentially help experts in one big data dimension expand their understanding of other dimensions.

The technical core of this paper comprises seven case studies. In section II, Techniques, we illustrate the potential of ontologies to semantically enhance data (subsection II.A) and metadata to facilitate image data mining (subsection II.B). In section III, Software Systems, we focus on some of the forces behind the transition from relational to NoSQL DBMS (subsection III.A) and from Hadoop MapReduce to Spark. In section IV, Application Domains, we discuss applications in astronomy and earth science (subsection IV.A) and in biomedical research (subsection IV.B). Finally, section V, Platforms, highlight the convergence of HPC and big data, as seen in the Avitohol computers system (subsection V.B).

All case studies are based on the authors' own research projects.

II. Techniques

A. Semantic Enhancement of Data with Ontologies

The characteristics of big data discussed above create a number of challenges to the methods and tools for their utilization. For example, the volume of data to be processed requires an ability to abstract the data in a form that summarizes the situation and is actionable from the point of view of humans and

978-83-60810-90-3/$25.00

decision-making software systems [16]. This requirement for *semantic scalability* is also important in the context of variety of data formats. The latter implies an additional requirement for an ability to integrate and interoperate with heterogeneous data – "to bridge syntactic diversity, local vocabularies and models, and multimodality" [19]. The velocity, i.e. the rapid appearance and change of data, requires the ability to focus on the relevant data and to process it quickly. Data veracity requires the capability of finding anomalies in it and making some types of reasoning based on proper domain knowledge. Extracting value using data analytics methods on various kinds of data creates the need of ability to extract knowledge from data and integrate it with existing knowledge bases.

Such challenges need be overcome to permit big data's full-scale potential for the user. Most traditional utilization approaches do not work in a satisfactory way for big data, so more agile utilization paradigms are needed in this case.

The main idea of the so-called *semantic utilization of big data* is to provide a kind of *semantic enhancement of data* that can be realized with the help of proper ontologies used to annotate data.

An *annotation* is a form of metadata attached to a specific dataset, to a particular database field or to a particular section of a document content. An annotation provides additional information (metadata) about an existing piece of data. Compared to tagging, which speeds up searching and helps one to find relevant and precise information, *a semantic annotation* goes one level deeper: It enriches the unstructured or semi-structured data with a context that is further linked to the available structured domain knowledge and makes it possible to process complex filter and search operations expecting results that are not explicitly related to the original search queries.

Ontologies [6] are the only widely accepted paradigm for the representation and management of open, sharable, and reusable knowledge in a way that allows automatic interpretation and inference. They provide semantic enhancement of data suggesting controlled vocabulary for annotations and thus permitting agile integration and semantic interoperability.

This kind of semantic enhancement of data may be characterized as an "arm's length approach" [15] – it presumes no change of data but association of each database field with an entire knowledge base. Data should be leaved as they are but incrementally tagged with terms from a consistent and non-redundant set of ontologies.

The successful implementation of the discussed approach depends on the creation of a shared resource (for example, a shared repository) of ontologies that could be used for annotation purposes. Moreover, it will be necessary to build an agile methodology for dynamic creation, application and extension of ontologies to annotate new sources of streaming data [15]. Such methodology should define a simple, repeatable process for ontology development and change management as well as an unambiguous process for data annotation using available ontologies.

B. Metadata in Image Data Mining

A most commonly accepted definition of "data mining" is the discovery of "models" for data. A "model", however, can be one of several things [12]. There are different approaches to modeling data. For thousands of years science was empirical. It was only in the last few hundred years that the theoretical paradigm emerged. The data-driven scientific inquiry came with data mining. The typical feature-based model looks for the most extreme examples of a phenomenon and represents the data by these examples. Some of the important kinds of feature extraction from large-scale data are:

1) *Frequent Itemsets* – a model makes sense for data that consists of "baskets" of small sets of items;
2) *Similar Items* – data looks like a collection of sets, and the objective is to find pairs of sets that have a relatively large fraction of their elements in common.

Data mining can be viewed as a result of the natural evolution of information technology. Data mining has incorporated many techniques from other domains such as statistics, machine learning, pattern recognition, database and data warehouse systems, information retrieval, visualization, algorithms, high-performance computing, and many application domains. In the field of image data mining, we developed an approach for extending the learning set of a classification algorithm with additional metadata. It is used as a base for the assignment of appropriate names to found regularities. The analysis of the correspondence between connections established in the attribute space and existing links between concepts can be used as a test for the creation of an adequate model of the observed world. Meta-PGN classifier is suggested as a possible tool for establishing these connections. We applied this approach to the field of content-based image retrieval of art paintings by designing system architecture for the extraction of specific feature combinations, which represent different sides of artists' styles, periods and movements [8]. Technically, we provide the system with a description of the real world and the systems then follows our mental model to generate appropriate names of detected concepts. The system interacts with the user, displaying those parts of the mental model that are utilized in the name generation process. This interaction is used to further improve and extend the mental model.

III. Software Systems

A. NoSQL versus RDBMS

The huge amounts of data that needs to be stored and processed on multiple servers is a recognized challenge of big data.

To manage the integrity of data, the classical database decisions (mostly relational databases) are based on transactions and support the main transactional characteristics – **A**tomicity, **C**onsistency, **I**solation, **D**urability, also known as ACID property.

But relational databases are difficult to scale, and they cannot guarantee increasing expectations for the performance

and availability when it comes to managing huge volumes of data on different servers.

Published by Eric Brewer in 2000, the CAP Theorem sets the basic requirements for a distributed system – Consistency, Availability, and Partition Tolerance [5]:

- Consistency – all the servers in the system have the same data;
- Availability – the system always responds to a request;
- Partition Tolerance – the system continues to operate as a whole even if an individual server fails.

The CAP Theorem postulates that only two of the three different aspects of scaling out can be fully achieved at the same time.

Traditional relational databases (Oracle, MS SQL, IBM DB2, etc.) are architected to run on a single machine and use strong, schema-based approach to modeling data that rely on consistency. So they represent the first group – Consistent, Available (CA) systems. Some column stores like Vertica, etc., also belong in that group.

NoSQL database decisions, on the other hand, are considered as an alternative to relational databases at times when organizations like Google and Amazon recognize that operating at scale is more effective using clusters of servers, and a schema-less data models are a feasible alternative in case of variety of data.

When distributed data stores are used, at the time of network partition, it is not possible to have both Consistency and Availability. So while traditional data-bases focus on Consistency, the NoSQL systems try to focus on Availability. In that case Consistency could be replaced by Eventual Consistency – data is not consistent at all time, but will be at given time or Local Consistency (the consistency is assured only within one single node and not throughout the cluster). Thus most NoSQL databases rely on properties less strict than the ACID ones, which are called BASE – Basic Availability, Soft state and Eventual consistency [13].

Consistent and Partition-Tolerant (CP) systems like HBase, MongoDB, and Big Table have difficulty to achieve data consistency over partitioned nodes, while Cassandra, Couch DB and Dynamo that support AP (Availability, Partition-Tolerance) achieve ”eventual consistency” through replication and verification.

Furthermore, the first group contains systems that support even ACID properties – like Dynamo, but most systems conform to BASE properties.

In the majority of cases non-traditional systems yield a better performance when ordinary data operations are measured. Our experiments on Oracle, Vertica, and Mongo DB platforms present particular results that confirm this thesis [10], [11].

B. From Hadoop MapReduce to Spark

Spark, initiated at the University of California, Berkley in 2009, was donated in 2013 to Apache. As an Apache project, Spark has gained popularity as a flexible and efficient in-memory implementation of the map-reduce distributed computing model and has already emerged as a faster substitute for the original Hadoop MR in-disk engine. Early Apache projects, such as Hive and Mahout, which originally compiled into Hadoop MR [27], have now been implemented to run on Apache Spark. Besides speed, Apache Spark's advantages to Hadoop MR include capabilities for interactive computing, stream processing, and sensor data processing (which are all lacking from the rigid two-stage Hadoop MR engine). While Apache Spark can be used within Apache Hadoop, it can also run independently, together with its own libraries, such as Spark SQL, Spark Streaming, Spark GraphX, and MLib (machine learning).

It has been broadly acknowledged that Spark has a pronounced efficiency edge over MR, but strict performance comparisons and analysis were scarce before 2014. In late 2014, Databricks, the company founded by the creators of the original Spark, released big data benchmark results that illustrate the speed advantages of Spark over Hadoop MR [26]. Spark was reported to perform three times faster than Hadoop MR on a 100 TB sort workload, and four times faster on a 1 PB workload, using – in both cases – significantly less hardware.

In early 2015, [14] reported performance experiments with codon count algorithms on nucleotide sequence data on AWS (the Amazon cloud computing platform). To do so, the authors measured the performance of a basic Spark codon count algorithm and compared it to a couple of Hadoop MR algorithms: a basic Hadoop MR codon count algorithm and an optimized ”local aggregation” Hadoop MR algorithm. The experiments confirmed that basic codon count with Spark is 15 times faster than basic codon count with MapReduce, a result that is unsurprising. Unexpectedly, however, basic codon-count with Spark remains about two times slower than optimized ”local aggregation” codon count with Hadoop MR. This result hints that properly optimized Hadoop MR code can be faster than the same analysis with Spark. The authors therefore suggest that available optimization techniques, such as local aggregation, be considered for speeding-up of legacy Hadoop MR applications in place of their eventual re-implementation in Spark for performance gains.

IV. Application Domains

A. Astronomy and Earth Science

With the current emergence of terabyte– and very soon petabyte-scale astronomical and Earth observation systems, the traditional approach to basic functions such as data searching, analytics and visualization are becoming increasingly difficult to handle. Simple database queries can result now in data subsets so large that they are incomprehensible, slow to handle, and impossible to visualize with commodity visualization tools. Astronomy and remote sensing complement each other, as they are on the quest for new big data interpretation capabilities: both disciplines have peculiar data, typical data processing and analysis chains, and specific models to be fed with data. However, both disciplines lack the capabilities for easily accessible semantics-oriented browsing in large data archives. Therefore, joint efforts to design and develop

innovative big data tools should help users in many different fields and set new standards for many communities. Several broad challenges to this line of reasoning that demand a multidisciplinary approach through international networking of experts and professionals have been identified. These challenges would then be channeled into COST Action TD1403 Objectives [1]:

- Digital curation and data access;
- New frontiers in visualization;
- Adaptation to new high performance computing technologies;
- New generation of interdisciplinary scientists.

The COST Action TD1403 was launched in April 2015 and will last for two years with a possible extension of 2 more. Now it involves 26 European countries. BigSkyEarth COST Action is organizing meetings, workshops, training schools and conferences. Also it supports Short Term Scientific Missions – exchange visits for individual mobility, strengthening existing networks and fostering collaboration between researchers.

The Bulgarian participation in the Action is in connection with a group of experts in astro-informatics – astronomers, mathematicians and computer science specialists [9]. Our fields are digitization of widefield (larger than or equal to 1 square degree) astronomical photographic plates, image processing and image compression.

The Wide-Field Plate Database [24], established in 1991, is the basic source of data for the wide-field plates obtained with professional telescopes world-wide [22]. It consists of four parts:

- Catalogue of Wide-Field Plate Archives with data for over 500 instruments (telescopes, cameras, etc.);
- Catalogue of Wide-Field Plate Indexes with descriptions of about 600 000 plates;
- Data Bank of digitized plate images (at low resolution for quick plate visualization and easy on-line access, and at high resolution intended for photometric and astrometric measurements);
- Links to online services and cross-correlation with other existing catalogues and journals.

We have identified more than 2 400 000 wide-field plates [21]. They allow one to obtain information of celestial objects over the past 133 years (1872-2005). At present over 300 000 wide-field plates have been digitized with total data about 30 TB.

Digitized photographic plates are irreplaceable sources for:

- Studies of the stellar long term brightness changes, as a result of observations conducted in different observatories;
- Studies of the long term variability of active galaxies;
- Searching and identification of potentially hazardous asteroids and comets which might cause catastrophic events by their collision with Earth.

B. Biomedical Research

Stimulated by the progress in computer technology and electronics data acquisition, recent decades have seen the growth of huge databases in biomedical sciences. For instance, Next generation sequencing (NGS) is a significant technological advance in biomedical sciences. It generates massive genomic datasets that play a key role in the big data phenomenon that surrounds us today. Advancing machine learning, data mining and statistical techniques for processing of big data are the key to transforming big data into actionable knowledge. One major problem with big data is that the standard methods of applied statistics are not really relevant for big data analysis. To extract information from high-dimensional data sets and make valid statistical inferences and predictions, novel data analytic and statistical techniques are needed. Here are some modern issues that we focus on.

Current advances in biomedical research technology, expression and SNP microarrays yield big data sets for many thousands of transcripts, genes or SNPs. Researchers are often interested in finding differences among these features between two separate groups, e.g. patients and controls, treatment and control groups; different strands; different tissues etc. Due to the differences of the underlying technologies and their biophysical and biochemical processes, scientists need to use statistical data analysis methods designed specifically for the particular technology. These tests often employ multiple comparison designs, where each gene, transcript or SNP is separately tested for significance and in many situations these tests are conservative. In complex multiple testing hypotheses, the classic statistical tests overestimate the p-values, leading to both loss of statistical power and increased experimental costs. One really common choice for correcting for multiple testing is to use the false discovery rate to control the rate at which things you call significant are false discoveries. There has been recent interest in developing efficient algorithms for multiple comparison to increase the statistical power and reduces the experimental costs. A computationally efficient technique has been proposed recently [4] that increases the statistical power, while controlling the False Discovery Rate of the statistical tests. This technique is applied to DESeq – a popular method for finding differentially expressed genes using RNA-sequence data. The statistical power increase is particularly high in small sample size experiments, often used in preliminary experiments and funding applications.

Some other issues arise from the method for finding differentially expressed gene. Apart from the DESeq method there are a few more like edgeR and limma frequently used by biomedical researchers [17]. These methods often produce similar, but not identical results. At the same time, due to randomness, even the same method can produce slightly different results on a data simulated from the same model. Therefore we are interested whether the slightly different results produced by different models can be attributed to randomness or to an underlying difference in the methods. Since the gene sequence is very long, we might not be interested in the full ranking

of the p-values but in some incomplete or partial orderings of them. Consequently, we want to compare such partial orderings. For example we can split the genes into several groups according to the size of their p-values lying in the sub-intervals $[0, 0.001]$, $[0.001, 0.01]$, $[0.01, 0.05]$, $[0.05, 0.1]$, $[0.1, 1]$. Then we construct a distance measure to compare how similar/dissimilar two incomplete rankings are based on the number of items present in the same ordered groups in both rankings [18]. Based on simulated large number of rankings and computed distance between any two of them, we can make inferences about the significance of a particular distance. That is to estimate the similarity between the corresponding incomplete rankings. Scientific computing is involved in all of these steps: simulating incomplete rankings, applying the method for finding differentially ex-pressed genes, computing all distances and estimating the distribution of the distance. We use the advanced computing resources at the Institute of Information and Communication Technologies (IICT) [7].

V. Platforms

Academic organizations are already moving towards unifying their HPC and big data processes within integrated HPC/big-data platforms, as observed in the development of the Avitohol platform at the Institute of Information and Communication Technologies at the Bulgarian Academy of Sciences.

As owner and manager of the Advanced Computing and Data Centre [7], IICT provides advanced computing resources and expertise thus helping Bulgarian science to come at the forefront of development worldwide.

The new multifunctional High Performance Computing system – Avitohol, forms the core of the computing infrastructure in the Institute. It consists of 150 computational servers HP SL250s Gen8, equipped with two Intel Xeon E5-2650v2 CPUs and two Intel Xeon Phi 7120P coprocessors, 64 GB RAM, two 500 GB hard drives, interconnected with non-blocking FDR InfiniBand running at 56 Gbp/s line speed. The total number of cores is 20700 and the total RAM is 9600 GB, respectively. The servers are deployed in 4 dual racks HP MCS 200, which have water cooling and can deliver up to 50 kW of power per rack. A central rack contains most of the storage, management servers and the central communication switches.

The system is controlled by two management (head) nodes and 4 I/O nodes. All those nodes are of the type HP ProLiant DL380p Gen8 with 2 Intel Xeon E5 2650v2 CPUs and 64 GB RAM. The I/O nodes provide access to 96 TB of raw storage capacity (24 disks of 4 TB each), which is provided by a SAN system.

The theoretical peak performance of the system is estimated at 412.3 TFlop/s in double precision while the RMAX Performance according the LINPACK benchmark is 264.2 TFlop/s. The Avitohol HPC system has been operational since June 2015 and it is ranked on 388th place according the 46th T0P500 list [20].

The second advanced computing system at IICT is the heterogeneous High Performance Computing Grid (HPCG)

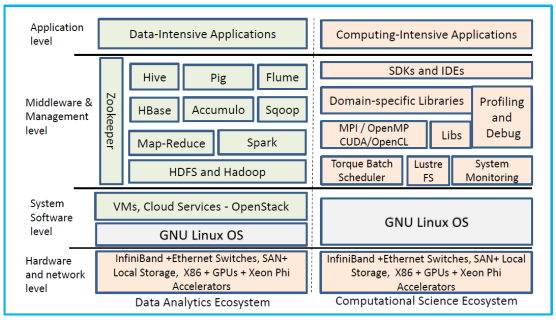


Fig. 1. Data analytics and computational ecosystems.

cluster [7]. It has been operational since 2010 and consists of HP Cluster Platform Express 7000 enclosures with 36 blades BL 280c (Total 576 CPU cores), 24 GB RAM per blade; 8 controlling nodes HP DL 380 G6 with dual Intel X5560 2.8 GHz, 32 GB RAM (total 128 CPU cores); 2 HP ProLiant SL390s G7 4U servers with 16 NVIDIA Tesla M2090 graphic cards (total 8192 GPU cores); 2 HP SL270s Gen8 4U servers with 8 Intel Xeon Phi 5110P Coprocessors each (total 480 cores, 1920 threads); 3 SAN storage systems with 132 TB total storage. All servers are interconnected with FDR InfiniBand running at 56 Gbps line speed. The theoretical peak performance of the system is estimated at 21.92 TFlops/s in double precision.

A dedicated optical network link has been established between the two systems, as well as between the Avitohol system and the main node of the Bulgarian Research and Educational Network (BREN) [2].

The existing computing facilities at IICT are involved in two computational infrastructures: the European Grid Infrastructure [3] and regional VI-SEEM infrastructure [23].

Based on their experiences in the last 10 years, the scientific and support staff in the center for HPC and Data computing at IICT [7] is dedicated to providing full support for the development and deployment of innovative scientific and industrial applications with substantial needs for computing power and data storage and transfer. The system can be considered to have two equally important sides. On the one hand it allows for state-of-the-art high performance computations, providing a full stack from base operating system software and libraries up to specially configured and deployed applications that make full use of the special capabilities of the systems, e.g. the Xeon Phi accelerators and CUDA GPGPU devices and the InfiniBand network. On the other hand, the storage systems provide access to data using various base protocols. The Lustre file system is the most frequently used for HPC workloads, while protocols like iSCSI are used for Cloud provisioning and other types of data processing. The data processing capabilities of the new Avitohol system are currently under configuration and testing, with the aim to build-up the full ecosystem with Hadoop and HDFS as the base layer and the components like Hive, Spark, Pig, etc., working on top of it. We plan to allow these components access to the Xeon Phi coprocessors

for advanced deep learning and related algorithms. Such new data processing capabilities will foster the development of integrated scientific applications that are based on realtime data, coming from local and international sources. Fig. 1 shows a schematic representation of the hardware and software components that are available or planned for deployment. System architects and engineers are developing a mixed HPC/Big data cluster, to provide for the convergence of HPC and Big data computing into a "single" environment.

The current goal of IICT is to achieve petaflop and petabyte-level of computing and storage capability, coupled with a developed software and middleware stack and services, opening the way to new forms of scientific research, more directly connected with the national industry and the societal challenges.

VI. Conclusion

This paper offers seven case studies that span several dimensions of big data: techniques, systems, applications, and platforms. All case studies are extracted from current research projects of the authors themselves. The case studies cover various aspects of big data and can, hopefully, be beneficial to the broad big data community by helping experts in one realm get acquainted with current cases in other realms.

The main contributions of the individual authors can be described as follows. M. Nisheva presented the potential of ontologies to semantically enhance data (section II.A). P. Stanchev discussed the use of metadata in image data mining (section II.B). K. Kaloyanova reviewed the capabilities of NoSQL databases as opposed to RDBMS (section III.A). N. Kirov described utilization of big data in Astronomy and Earth Science (section IV.A), while E. Stoimenova highlighted the specifics of statistical applications in biomedical research (section IV.B). T. Gurov focused on the convergence of HPC and big data, as realized with the Avitohol platform (section V.B). A. Radenski discussed the transition from Hadoop MapReduce to Spark (section III.B). A. Radenski also planned the overall structure of this publication and drafted the abstract, the introductory section I (minus the specification of the big data characteristics), the background section V.A, and the concluding section VI.

Acknowledgment

The work of the first author (A.R.) was supported by an AWS in Education 2014-2015 award from Amazon. The work of the second author (T.G.) was partly supported by the National Science Fund of Bulgaria under Grant DFNI-I02/8 and by EC Programme Horizon 2020 under project VI-SEEM project, Grant Agreement No: 675121. The work of the last author (E.S.) was supported by the National Science Fund of Bulgaria under Grant DFNI-I02/19.

References

[1] Big Data Era in Sky and Earth Observation (Big-SkyEarth) COST Action TD1403, http://bigskyearth.eu/, (Retrieved January, 2016).
[2] Bulg. Research and Educational Network (BREN), http://www.bren.bg/.
[3] EGI, 2016, European Grid Infrastructure, www.egi.eu .
[4] J.P. Ferguson, D. Palejev, "Calibration of p-values for multiple testing problems in genomics", *Stat. Appl. Genet. Molec. Biol.*, vol. 13(6), 2014, pp. 659–73, doi: 10.1515/sagmb-2013-0074.
[5] S. Gilbert, N. Lynch, "Perspectives on the CAP Theorem", *Computer*, vol. 45, no. 2, 2012, pp. 30–36, http://doi.ieeecomputersociety.org/10.1109/MC.2011.389.
[6] T. Gruber, "Toward Principles for the Design of Ontologies Used for Knowledge Sharing", *International Journal of Human-Computer Studies*, Vol. 43, 1995, pp. 907–928, doi:10.1006/ijhc.1995.1081.
[7] Advanced Computing and Data Center, IICT, http://hpc.acad.bg/.
[8] K. Ivanova, I. Mitov, P. Stanchev, Ph. Ein-Dor, K. Vanhoof, "Establishing Correspondences between Attribute Spaces and Complex Concept Spaces Using Meta-PGN Classifier", *Proc. of the 2nd Int. Conf. "Digital Preservation and Presentation of Cultural Heritage"*, V. Tarnovo, Bulgaria, IMI-BAS, Sofia, 2012, ISSN 1314-4006, pp.71–77.
[9] O. Kounchev et al, Astroinformatics: "A Synthesis between Astronomical Imaging and Information & Communication Technologies", *In: Modern Trends in Mathematics and Physics*, ed. S.S. Tinchev, Heron Press, Sofia, 2009, pp. 60–69.
[10] H. Kyurkchiev, K. Kaloyanova, "Performance Study of Analytical Queries of Oracle and Vertica", *Proc. of the $7^{t}h$ International Conference "Information Systems & Grid Technologies"*, Sofia, 2013, pp. 127–139, DOI:10.13140/2.1.3667.0726.
[11] H. Kyurkchiev, E. Mitreva, "Performance Study of SQL and NoSQL Solutions for Analytical Loads", *Proc. of the Doctoral Conference in "Mathematics, Informatics and Education" (MIE2013)*, Sofia, 2014, pp. 49–57, DOI: 10.13140/2.1.1307.7766.
[12] J. Leskovec, A. Rajaraman, J. D. Ullman, *Mining of Massive Datasets*, 3rd Edition, Cambridge University Press, 2014.
[13] E. Mitreva, K. Kaloyanova, "NoSQL solutions to handle big data", *Proc. of the Doctoral Conference in Mathematics, Informatics and Education (MIE 2013)*, Sofia, 2013, pp. 77–85.
[14] A. Radenski, L. Ehwerhemuepha, K. Anderson. "From in-disk to in-memory big data with Hadoop: Performance experiments with nucleotide sequence data", *Proc. ABDA'15, the International Conference on Advances in Big Data Analytics*, CSREA Press (H. Arabnia and M. Yang, Ed.), 2015, pp. 34–40.
[15] D. Salmen, T. Malyuta, A. Hansen, S. Cronen, B. Smith, "Integration of Intelligence Data through Semantic Enhancement", *Proceedings of the Conference on Semantic Technology in Intelligence, Defense and Security STIDS 2011*, CEUR, Vol. 808, 2011, pp. 6–13.
[16] A. Sheth, "Transforming Big Data into Smart Data: Deriving Value via harnessing Volume, Variety and Velocity using semantics and Semantic Web", *Keynote at the 21st Italian Symposium on Advanced Database Systems* 2013. http://j.mp/SmatData, visited on December 23, 2015.
[17] C. Soneson, M. Delorenzi, "A comparison of methods for differential expression analysis of RNA-seq data", *BMC bioinformatics*, vol. 14(1), 2013, 1, DOI: 10.1186/1471-2105-14-91.
[18] E. Stoimenova, D. Palejev, "Comparison of incomplete ranked lists with application to RNA-seq differential expression methods", Preprint, 2016.
[19] K. Thirunarayan, A. Sheth, "Semantics-Empowered Approaches to Big Data Processing for Physical-Cyber-Social Applications", Association for the Advancement of Artificial Intelligence (AAAI) Technical Report FS-13-04, 2013.
[20] TOP500 list, November 2015, http://www.top500.org/site/50586.
[21] M. Tsvetkov, K. Tsvetkova, N. Kirov, "Technology for scanning of astronomical photographic plates", *Serdica Journal of Computing*, vol. 6 (1), 2012, pp. 77–88.
[22] M. Tsvetkov, "Wide-Field Plate Database: a Decade of Development", *In: Observatory: Plate Content Digitization, Archive Mining and Image Sequence*, Processing, iAstro workshop, Sofia, Bulgaria, Eds. M. Tsvetkov, F. Murtagh, R. Molina, 2006.
[23] VI-SEEM, 2016. https://vi-seem.eu/.
[24] Wide-Field Plate Database, http://www.wfpdb.org, (Retrieved January, 2016).
[25] T. White, *Hadoop: The Definitive Guide*, O'Reilly, 2012.
[26] A. Woodie, Spark Smashes MapReduce in Big Data Benchmark. Datanami, October 10, 2014, http://www.datanami.com/2014/10/10/spark-smashes-mapreduce-big-data-benchmark/.
[27] E. Zdravevski, et al, "Parallel computation of information gain using Hadoop and MapReduce", *Proc. of the 2015 Federated Conf. on Comp. Sci. and Inf. Systems" (FedCSIS2015)*, Lodz, Poland, 2015, pp. 181–192, DOI: 10.15439/2015F89.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 889–898
DOI: 10.15439/2016F498
ACSIS, Vol. 8. ISSN 2300-5963

Superlinear Speedup in HPC Systems: why and when?

Sasko Ristov, Radu Prodan
University of Innsbruck
Innsbruck, Austria
Email:{sashko, radu}@dps.uibk.ac.at

Marjan Gusev
University Ss Cyril and Methodius, FCSE
Skopje, Macedonia
Email:marjan.gushev@finki.ukim.mk

Karolj Skala
Ruđer Bošković Institute
Zagreb, Croatia
Email:Karolj.Skala@irb.hr

***Abstract*—The speedup is usually limited by two main laws in high-performance computing, that is, the Amdahl's and Gustafson's laws. However, the speedup sometimes can reach far beyond the limited linear speedup, known as superlinear speedup, which means that the speedup is greater than the number of processors that are used. Although the superlinear speedup is not a new concept and many authors have already reported its existence, most of them reported it as a side effect, without explaining why and how it is happening.**

In this paper, we analyze several different superlinear speedup types and define a taxonomy for them. Additionally, we present several explanations and cases of superlinearity existence for different types of *granular* algorithms (tasks), which means that they can be divided into many sub-tasks and scattered to the processors for execution. Apart from frequent explanation that having more cache memory in parallel execution is the main reason, we summarize other different effects that cause the superlinearity, including the superlinear speedup in cloud virtual environment for both vertical and horizontal scaling.

***Index Terms*—Cache memory, load, parallel and distributed processing, performance.**

I. Introduction

THE goal of today's world of parallel and distributed systems is to achieve the greatest speedup, represented either as the lowest time for execution of a single task (High Performance Computing), or to execute as many tasks as possible for a given time (High Throughput Computing), when the task(s) are executed on scaled resources. Many new algorithms and computing paradigms appeared in the last decade, and new challenges have emerged to solve more complex problems faster, or to achieve greater speedup, as much as possible [1].

The speedup is usually defined as a ratio of the wall times of sequential and parallel execution of an algorithm. The target of the parallelization is to achieve the lowest execution time in order to maximize the speedup against the best sequential algorithm. Increasing the number of computing resources will increase the intra-resource's communication and requires additional operations, such as reduction operations.

Most of the authors analyze the computer only as a processing unit, focusing on the processing power, without analyzing the details of the computer as a complex system with memory and I/O devices as resources. Actually, these resources limit the speedup, or can boost its performance.

According to the Gustafson's Law [2], the speedup is limited with the number of processors, when the linear speedup is achieved. However, beyond the limits, the superlinear speedup happens in reality for plenty of reasons and it allows an increased utilization of parallel systems [3].

Many authors reported the existence of a superlinear speedup, but most of them only mentioned it as a side effect [4]. Besides reporting a superlinearity, other researchers briefly presented that the reason for achieving a superlinear speedup is because of the greater amount of cache memory in the parallel execution compared to the sequential [5]. However, these explanations are insufficient. For example, all currently produced multiprocessors contain a multi-level cache, but a superlinear speedup is not reported always. Also, it is not reported for all algorithms. Sometimes it is limited to the problem size or the number of used multiprocessors.

In this paper, we present a systematic overview of the reasons why the superlinear speedup appears. The analysis approach is to focus on granular algorithms, in both traditional and cloud virtual environments. Tthe superlinearity is reported in both environment types, explaining the reasons summarized in this paper. Data- or code-parallelism divides a single task into threads or processes and sends them for execution on different processors, thus aspiring to become a high-performance computing system with a goal to finish the task as fast as possible. On the other hand, today's service oriented architectures offer scalable web services to their customers using elastic cloud resources. The latter approach tends toward a high throughput computing system aiming at serving as many possible customers for a certain time, without reducing the service performance.

Due to its elasticity and the linear pay-as-you-go model, the cloud is preferred platform both for high performance algorithms, especially if they are low communication-intensive, such as scientific applications [6], [7]. Many scientific applications are moving from computation-intensive to data-intensive, that is, they require high throughput computing, rather than high-performance computing. This is a huge challenge in the cloud because the data transfer between the cloud compute nodes and storage is a bottleneck [8]. Despite the additional virtualization layer, a superlinear speedup is also reported for granular algorithms [9].

978-83-60810-90-3/$25.00

The rest of the paper is organized as follows. The speedup limits in parallel executions are described in Section II. Section III elaborates when and how a superlinear speedup can be achieved for a parallel implementation of some algorithm. Examples of obtained superlinear speedup for high performance algorithms are presented in Section IV. Despite the virtualization layer, the cloud environment can also achieve a superlinear speedup, as discussed in Section V. Section VI discusses several paradoxes, as well as further challenges. Finally, we conclude the paper and present the plans for future work in Section VII.

II. Speedup Limitations

This section briefly explains the two main laws in the computer architecture about the limit of the maximal speedup that can be achieved when an algorithm is executed parallel with more computing resources, that is, Amdahl's [10] and Gustafson's laws. The former targets the speedup for problems with fixed problem size while the latter the algorithms that require intensive parallel processing.

Speedup $S(p)$ is defined as a ratio of the execution times of the best sequential algorithm $T(1)$ and the parallel implementation on p processors $T(p)$, as presented in (1). However, this definition holds only for fixed-time algorithms. When analyzed more broadly, the speedup should be defined as a ratio of speeds, and not of times, as defined in (2). Note, that for fixed-time algorithms, the amount of work is constant, which results in (1).

$$S(p) = \frac{T(1)}{T(p)} \tag{1}$$

$$S(p) = \frac{(\frac{ParallelWork}{ParallelTime})}{(\frac{SerialWork}{SerialTime})} \tag{2}$$

Fig. 1 presents the theoretical or ideal speedup for both laws, depending on the number of processors used in the parallel execution. The Amdahl's Law limits the speedup to the value $1/s$, as defined in (3), where s is the portion of the serial part for the fixed size program. The conclusion is that the speedup is limited regardless of the number of processors, when the problem is fixed.

$$S^{max}_{Amdahls}(p) = 1/s \tag{3}$$

The Gustafson's Law, on the other side, shows that if the problem is executed within a fixed time, the maximum value of the speedup is *linear* limited by the number of processors, as defined in (4), assuming that the problem size increases and the serial portion becomes negligible. However, in real executions, due to communication, synchronization, and resource sharing, the speedup is *sublinear*, or $S(p) < p$.

$$S^{max}_{Gustafson}(p) = p \tag{4}$$

Both the Amdahl's and Gustafson's laws calculate the maximum speedup, that is, the speedup limit of a parallel algorithm or program; they both consider that the serial part of the algorithm does not depend on the number of processors. However, this is in the ideal condition, while in a real situation, each processor does not start and finish in the same time, and the communication overhead and synchronization can harm the parallel execution when the number of processors increases.

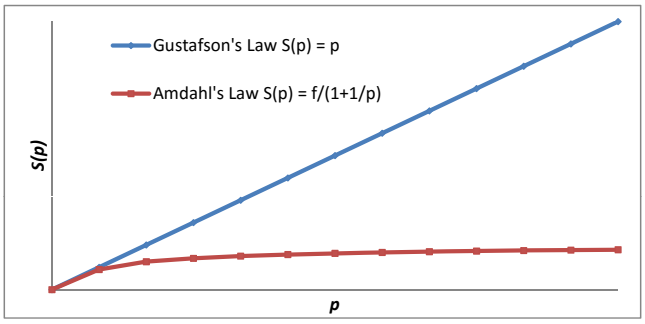


Fig. 1. Speedup for Amdahl's and Gustafson's laws

Karp and Flatt [11] introduced the *scaled serial fraction* f_k of an algorithm as defined in (6), where p is the number of processors, and s_k is the speedup that calculates the overhead (5) as a number of the executed additional arithmetic operations for parallel execution. Parameter k represents the scaling factor for the overhead in a parallel implementation using p processors.

$$s_k = \frac{k \cdot T(1,1)}{T(p,k)} \tag{5}$$

$$f_k = \frac{1/s_k - 1/p}{1 - 1/p} \tag{6}$$

Let's discuss the relation for the scaled serial fraction. If s_k = p, then $f_k = 0$, which yields to the Gustafson's Law. The results of the parallelization will still be good even if the parallel implementation achieves a small speedup, while f_k retains to a some constant value, because the algorithm has limited parallelization.

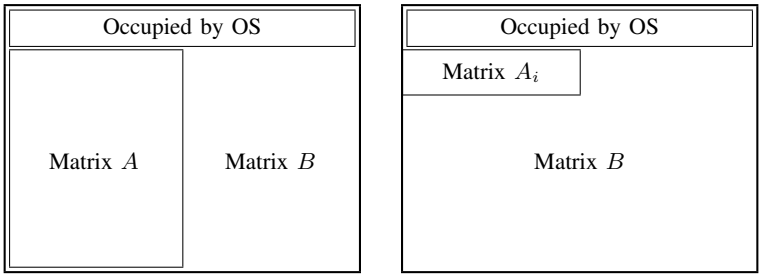
Let's rewrite (6) as (7) in order to determine the speedup that calculates the overhead s_k and yield special cases, that is, the Amdahl's and Gustafson's law.

$$s_k = \frac{1}{f_k \cdot (1 - 1/p) + 1/p} \tag{7}$$

If the scaled serial fraction $f_k = 0$, then $s_k = p$, which yields toward Gustafson's Law, while if $p \to \infty$, then $s_k = 1/f_k$, as Amdahl's Law states. For each scaled system with $f_k > 0$, the scaled speedup that calculates the overhead is sublinear, i.e. $s_k < p$.

III. Beyond the Speedup Limits. Why and When?

Although the limits are given by the Gustafson's Law, the speedup achieved by executing some algorithms on parallel configurations goes beyond it, achieving a superlinear speedup. This section presents several such cases, along with detailed



a) sequential version b) parallel version

Fig. 2. Cache occupancy in sequential and parallel execution

explanation about the reason for various superlinear speedup appearances.

Let's analyze when a superlinear speedup can be achieved while parallelizing a sequential problem. The CPU execution times for sequential algorithm T_s and parallel algorithm T_p are respectively defined in (8) and (9), where CC and MC represent the clock cycles required by the processor for execution of operations and memory accesses, correspondingly, and CT the time period of a single clock cycle. In a homogeneous environment, CT will be the same for both implementations.

$$T_s = (CC_s + MC_s) \cdot CT \tag{8}$$

$$T_p = (CC_p + MC_p) \cdot CT \tag{9}$$

Shi [12] classified the parallel versions of algorithms to be either structure persistent or non-persistent. The former means that the number of total operations that the algorithm executes is same both for the sequential and parallel implementation, for the same input. The latter's parallel implementation does not execute all operations, that the compatriot sequential algorithm would. A formal notation of (7) means that the scaled serial fraction $f_k < 0$, which yields toward a superlinear speedup $s_k > p$.

Gusev and Ristov [13] defined the condition when a superlinear speedup can be obtained for a shared memory multiprocessor, which is presented in (10), where a positive number ϵ exists, such that $0 \leq \epsilon \leq p$ and $CC_s = CC_p \cdot (p - \epsilon)$. The parameter ϵ represents the effect of parallelization overhead and synchronization and p the number of scaled computing resources.

$$MC_s > p \cdot MC_p + \epsilon \cdot CC_p \tag{10}$$

The superlinearity was defined for cache-intensive algorithms only (algorithms where the cache-intensive complexity represented by the average reuse of an element c is greater than 1 [14]). For example, the dense matrix-matrix multiplication algorithm has cache-intensive complexity $c = O(N)$ because each element of matrices is accessed N times for different computations. On the other hand, the cache-intensive complexity of the scalar product is $c = 1$, which yields that a superlinear speedup cannot be achieved. We must note that the cache-intensive complexity defines the level of superlinearity, that is, an algorithm with greater cache-intensive complexity ($c \gg 1$) will achieve a greater superlinear speedup.

A. Superlinear speedup for non-persistent algorithms

Typical examples of non-persistent algorithms are searching algorithms, which finish when one of the processors finds the solution, and together with all the other processors stop the execution, without finishing all operations. In this case, a superlinear speedup usually appears because CC_p is smaller than CC_s, thus compensating the overhead of parallelization. This case can be better presented if the total number of clocks are presented through the number of instructions I and CPI (clocks per instruction), as presented in (11) and (12) [15]. I_p will be smaller than I_s, which will cause a spurious superlinear speedup.

$$CC_s = I_s \cdot CPI_{CC}; \tag{11}$$

$$CC_p = I_p \cdot CPI_{CC} \tag{12}$$

Many examples can be found in the literature for superlinear speedup of the non-persistent algorithms. For example, parallel shortest path planning [16].

B. Superlinear speedup for persistent algorithms

The total number of instructions of the sequential and parallel implementations of structure persistent algorithms is the same, that is, $I_p = I_s$, which means that superlinear speedup appears due of the memory clock cycles, that is, by reducing the number of clocks per instruction for memory access CPI_{MC} in the parallel implementation. There are several different cases when CPI_{MC} in parallel implementation will be smaller than its serial compatriot. Let us explain all these cases.

1) More cache for parallel execution: The case when the parallel execution of a structure persistent algorithm can obtain a superlinear speedup due to utilizing more cache memory is the mostly reported by the researchers [17].

Since more cache memory is used in parallel execution, for some region of problem size, it can store the whole problem

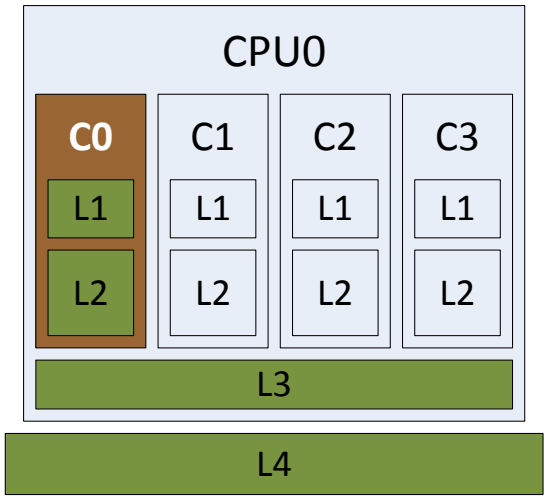


Fig. 3. Utilized memory for sequential execution

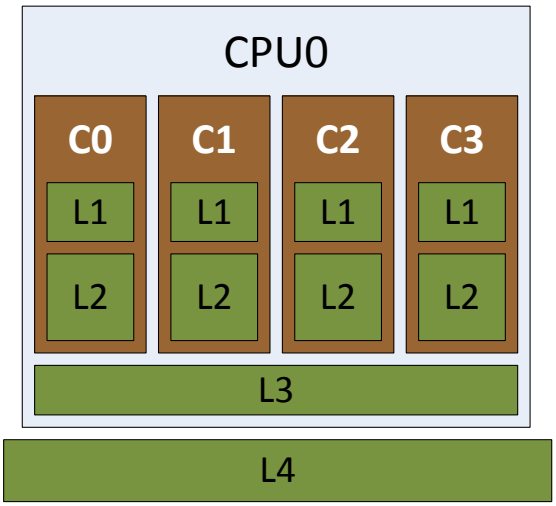


Fig. 4. Utilized multi tiered memory for loosely coupled processors for parallel execution

size, while the sequential execution cannot, as presented for storing matrices in Fig. 2 [13].

Fig. 3 [18] presents an example for utilized memory tiers of a typical multiprocessor with four cores, each with private L1 and L2 cache memory, shared L3 cache memory, and main memory represented by L4.

Velkoski et al. [18] went beyond this analysis. They have analyzed the impact of loosely and tightly coupled cores for parallel implementation and concluded that the former is superior for naive dense matrix-vector multiplication. The utilization of memory tiers of a typical multiprocessor for loosely and tightly coupled cores in parallel execution is presented in figures 4 and 5 [18]. The tightly coupled case uses all four cores on one chip, while the loosely coupled case uses one CPU core of four chips on a shared memory multiprocessor. The results show that a superlinear speedup region appears for both loosely and tightly coupled processors, which starts for the same matrix size, but the former's region is wider, as well as it achieves a greater superlinear speedup. These results clearly present that the use of more L2 cache memories for parallel execution yields a superlinear speedup in the tightly coupled processors, while more L3 cache memory generates even a greater superlinear speedup and region, despite the increased overhead of the inter-chip communication, compared to the intra-chip for tightly-coupled systems.

Another interesting example was reported by Djinevski et al. [19] achieving a superlinear speedup region on GPU, when they used one loosely coupled processing unit of all streaming multiprocessors (SMs) for parallel execution, and a single processing unit of one SM for sequential execution. The superlinear speedup regions are achieved regardless of the used number of SMs.

2) Shared cache for parallel execution: Although most of the reported superlinear speedups are obtained because of the increased cache memory in a parallel execution, a superlinear speedup is achieved in those some algorithms addressing a common shared cache. That is, a superlinear speedup can be achieved even in the tightly coupled processors.

For example, this is the case for an algorithm where the same variables (data) are used by several or all shared memory multiprocessors. If these variables are defined as shared, then fetching a variable by one processor will load it in the upper memory tier (for example, from RAM to the shared L3 cache), thus reducing the access time for the same variable by other processors.

Next, let's explain the difference when multiprocessors, which use private per core cache or shared cache, access the data from the memory. Without loosing generality, assume that the multiprocessor has one cache level and the accessed memory location is not present in the cache.

Fig. 6 presents how two multiprocessors, each with a private cache, access the same memory location. Let's assume that the instruction $Read(X)$ is executed by the processor A earlier. It will generate a cache miss, and pay the penalty for that. After fetching the variable X from the memory into its private cache, the processor B will do a similar sequence. This means that in this case, two cache misses and two memory accesses will happen.

Accessing the data in the memory by two multiprocessors that use a shared cache is presented in Fig. 7. In this environment, when the processor A accesses the variable X, a cache miss will be generated and one memory access. Now, when the processor B will require the same variable X in the near future without replacing it from the cache, a cache hit will be generated without a cache miss penalty and an additional memory access.

We can conclude that a tightly coupled multiprocessor (that uses a shared cache) can benefit when shared variables are used by reducing the cache misses and memory accesses.

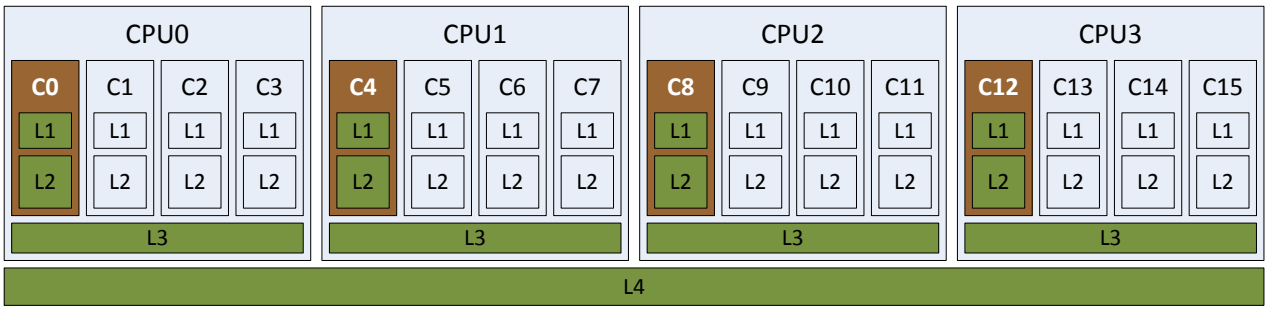


Fig. 5. Utilized multi tiered memory for loosely coupled processors for parallel execution

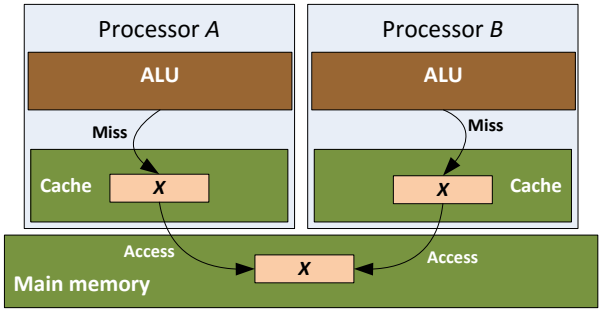


Fig. 6. Accessing the data from the memory by two multiprocessors that use private cache. Two cache misses and two memory access will happen.

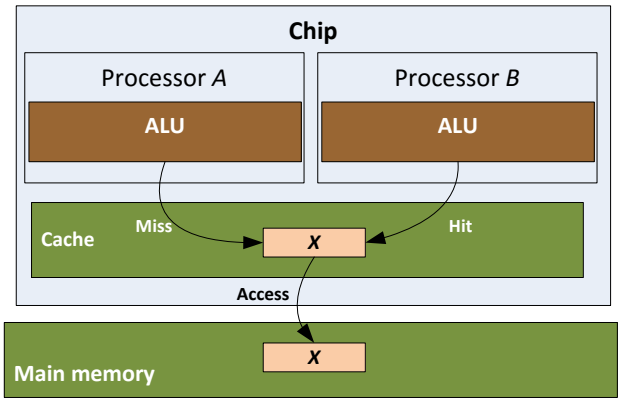


Fig. 7. Accessing the data from the memory by two multiprocessors that use shared cache. Only one cache miss, one cache hit and only one memory access will happen.

The precondition is that the time and space locality should be utilized by all processors.

Anchev et al. [20] reported a superlinear speedup for dense matrix-matrix multiplication. The reason for superlinearity is the use of a shared L3 cache, or, as we explained earlier, the implicit prefetch of the data (matrix elements).

However, we must note that the superlinear speedup was reported only for AMD Opteron, while Intel's i7 obtained a sublinear speedup only. We believe that the reason for this is due to the fact that the frequency gap between L3 and main memory is much bigger for AMD Opteron, and thus reducing the cache miss ratio and penalties, which compensates the parallelization overhead, and generates a superlinear speedup.

3) Superlinear speedup in a heterogeneous environment: Superlinear speedup is reported in a heterogeneous environment that consists of three Intel Xeon CPU + one GPU NVIDIA FX Quadro, because the heterogeneous environment schedules the tasks better than the homogenous environment and thus reduces the impact of Amdahl's Law with a limited overhead in parallel execution [21].

IV. Superlinear Speedup Regions

This section overviews several examples of granular algorithms, where the existence of a superlinear speedup is reported. We define two different region types of a superlinear speedup: 1) for some range of the number of processors, usual

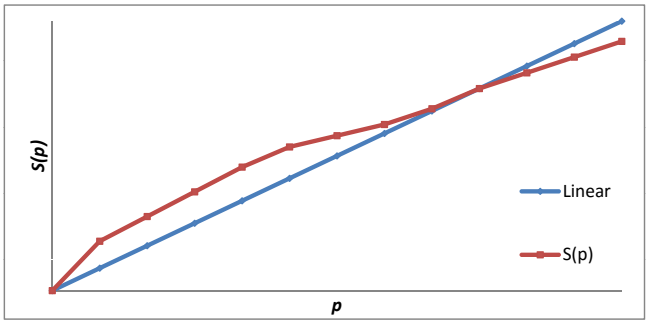


Fig. 8. Example of superlinear speedup for a particular range of number of processors (fixed problem size)

for fixed problem size, and 2) for a particular range of problem size, but fixed number of processors.

Fig. 8 presents a superlinear speedup for some range of the number of processors, usual for fixed problem size. The superlinearity usually starts even when two processors are used. However, it is lost as the scaling factor increases [22] due to the communication and synchronization overhead.

Another situation is to fix the number of processors, but

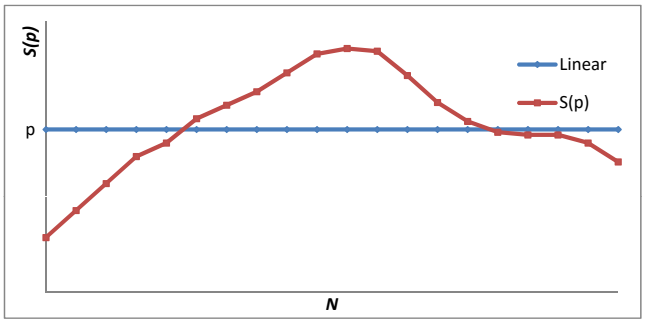


Fig. 9. Example of superlinear speedup for a particular range of problem size, but fixed number of processors

change the problem size, which can also impact the speedup, as presented in Fig. 9. We observe that there is a superlinear region for a specific range of problem size N where the speedup $S(p) > p$, while in other regions, it is sublinear.

We must note that sometimes, the speedup could be $S(p) < 1$, (sublinear speedup), which means that it is not speedup, but a slowdown. This could happen for several reasons. For example, for small problem size, which is negligible for good performance comparison, sequential execution will be faster than the time for forking threads. Another example is the case when the number of threads or processes is greater than the number of existing processors. Or more generally, a slowdown may happen due to the communication and synchronization time, or the overload of instruction in parallel execution. Further on, using the cache line for time and space locality in sequential execution can overcome the problem of the limited number of processors. Let's define it more formal, that is, the condition $MC_s < p \cdot MC_p$ will compensate $CC_p \leq p \cdot CC_s$, which will lead to a slowdown. In this case, the speedup could be achieved if problem size is divided into huge data chunks that will be scattered to the processing resources that will execute them sequentially.

Without losing generality, many authors use the *Efficiency* indicator calculated by $E(p) = \frac{S(p)}{p}$, which maps the limited speedup into the range [0, 1]. This parameter helps a lot to compare parallel executions with a different number of processors among each other. However, the value of $E(p)$ for superlinear speedup is $E(p) > 1$.

Gustafson [23] presented two cases of non-spurious superlinear speedup. Superlinear speedup can be obtained in distributed memory ensembles because of various memory speed. He also reported a superlinear speedup in cases when algorithms and tasks are with different speed.

Sometimes, parallel execution achieves a superlinear speedup because it partitions and reduces the data chunks, which can be placed in the cache memory, thus reducing the cache misses [24], [25], [26], [27], [28].

Many authors reported a superlinear speedup for parallel execution of some algorithms. However, most of them presented a likely explanation only for the superlinear speedup appearance. For example, one explanation is that the reason for achieving a superlinear speedup is because of having more cache in parallel execution. Still, in most cases the superlinear speedup is achieved for some range of the used number of processors, or for a specific problem size, or in a combination of both cases. For example, Monagan and Pearce [29] achieved a superlinear speedup for the parallel sparse polynomial division. However, they did not explain why a superlinear speedup has not appeared for extremely sparse problems, although a parallel execution uses the same amount of cache. Also, the same experiments have not reported a superlinear speedup on the Core 2 processor, although the level 3 cache has more cache than the sequential one.

Phillips et al. [30] reported a superlinear speedup, even when comparing parallel executions up to 26 processors with the one that uses two processors for continuous iterative guided spectral class rejection (CIGSCR). Peschlow et al. [31] achieved a superlinear speedup while simulating wireless networks, but only in a single range of a number of processors and for a specific number of nodes.

V. Analysis of a Superlinear Speedup in Cloud Environment

This section presents several cases where a superlinear speedup is achieved in cloud virtual environment for various types of scaling the resources.

Nowadays, cloud computing is being increasingly used for high-performance and high throughput applications. Its elastic on demand resources allow the customers to rent, for example, 1000 processors and execute a certain task, instead of building their own underutilized data center. Since the cloud's pricing strategy is linear, and expected speedup is also linear, it seems that customers will be charged fairly. In reality, the reported performance for communication-intensive high-performance algorithms shows that customers might feel that they are cheated. However, there are several cases where the superlinear speedup is achieved, despite the virtualization layer.

Customers can scale their rented resources horizontally, vertically or diagonally in the cloud. If the original configuration maps one process to a virtual machine (VM) instance hosted on a processor with one CPU core, as presented in Fig. 10 a), then Fig. 10 b), c) and d) present the three possible cloud scalings. The horizontal scaling presented in Fig. 10 d) increases the number of same VM instances and maps separate process (with a single thread) to a different VM instance. The vertical scaling presented in Fig. 10 b) increases the number of CPU cores per VM (resized VM) and maps separate threads of a single process to a different core on the same VM instance. A combination of the both scaling types yields a diagonal scaling presented in Fig. 10 c). To realize the vertical and diagonal scaling, the customer should use some API for parallelization, such as OpenMP, which will create parallel threads.

There are published papers that present a superlinear speedup in both the horizontal and vertical scaling. A superlinear speedup is reported and elaborated for cache-intensive algorithms [9] in the case of vertical scaling. Although sequential execution utilizes cache more, the superlinear speedup

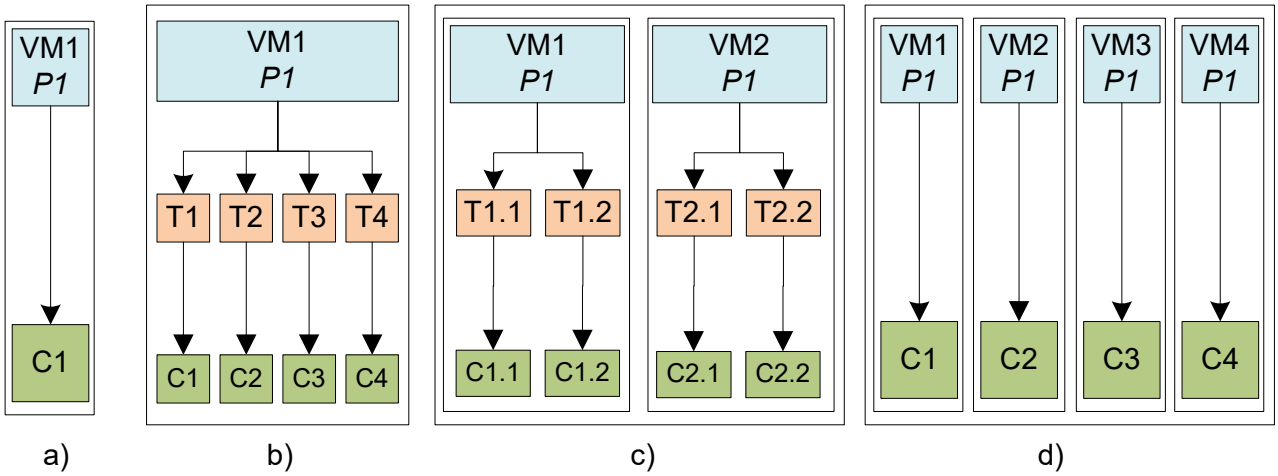


Fig. 10. Example of b) Vertical, c) Diagonal, and d) Horizontal scaling of nominal resources a)

is achieved also for horizontal scaling in the cloud, as well [14]. The authors have determined that the cloud environment handles the cases when the problem size can be fitted in the last level cache memory better, which is the reason why a superlinear speedup is achieved [32].

VI. Discussion

The superlinear speedup is achieved by many researchers, usually as a side effect without elaborating the theoretical background and explaining the details. In this paper, we have analyzed several aspects how to achieve a superlinear speedup, explaining why and when it can happen when various algorithms are executed on different platforms.

A. Superlinearity versus algorithm type

Mainly, the multi-tiered memory organization is the main reason to obtain a superlinear speedup for granular algorithms. We have classified two paradoxical cases; in the first case, the superlinearity appears because of the increased capacity of L2 and L3 cache memory, while in the second case, it is achieved because of the shared last level cache memory. A superlinear speedup is achieved in the first case, when the algorithm is executed on a loosely coupled system, while in the other case, the algorithm executed on a tightly coupled system.

The main reason is the difference of the algorithms. The loosely coupled parallel execution outperforms the tightly coupled for dense matrix-vector multiplication, in which, the matrix $A_{N \cdot N}$ is divided horizontally among processors for row-major memory, and implicit fetching is not used, while it is used only for the vector $B_{N\ cdot1}$, as presented in Fig. 11 a). However, its dimension is $O(N)$, while the matrix dimension is $O(N^2/p) \rightarrow O(N^2)$ is dominant, because the number of processors $p << N$.

On the other side, Fig. 11 b) presents the data that is accessed by the processor P_i for parallel execution of dense matrix-matrix algorithm, which shows that each processor uses the whole matrix B and therefore, implicit fetching yields a superlinear speedup. In this case, the size of shared data among all processors is bigger than the private chunks of matrix A, as well as compared to the vector's size in dense matrix-vector multiplication.

Another issue is the way of storing the matrices. Without loosing generality, we will assume that a row-major storing is used. Accessing the data of the matrix A is linearly, which utilizes the cache lines and thus reduces the cache misses regardless of the cache size. For example, when accessing the element $a_{i,j}$, the elements $a_{i+1,j}, a_{i+2,j}, \ldots a_{i+k,j}$ are also fetched in the cache. The size of k depends on the cache line and matrix element sizes. Therefore, cache misses are generated for the element $a_{i,j}$ only. Accessing the elements of the matrix B does not utilize the cache line, because a column of the matrix B is accessed linearly. In this case, the cache size is very important in order to store as much as possible a part of the matrix B.

We must note that although very naive examples of dense matrix-matrix and matrix-vector multiplications were presented, the generality is not lost. Our goal is not to prefer this algorithm for parallel execution, but just to show how and when a superlinear speedup can be obtained, paradoxically, for various algorithm - totally different reasons.

Using a multi-tiered memory is not the sine qua non for superlinearity. As we presented an example in Section III-B2, Intel i7 processor has not obtained a superlinear speedup for the same algorithm, as AMD Opteron. For example, a superlinear speedup is obtained on Cray XMT [33]. Intel Xeon achieved a superlinear speedup for two processors using the data-parallelism benchmarking (Black-Scholes), but only a sublinear speedup with dense matrix-vector multiplication [34]. Therefore, having cache memory is only one of the conditions for the existence of a superlinear speedup. An important observation is to return to the speedup definition,

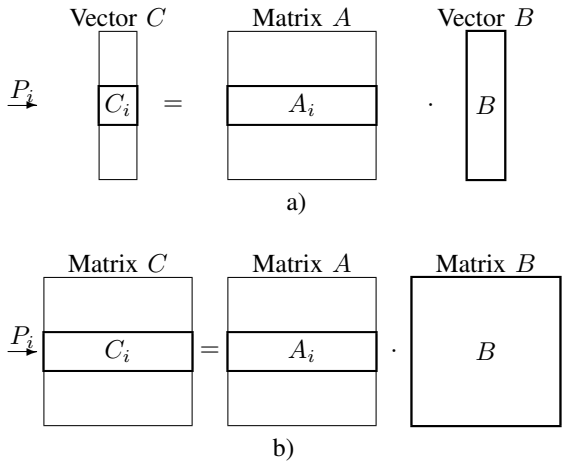


Fig. 11. Example of parallel implementations of matrix-vector and matrix-matrix multiplication.
a) Processor P_i uses chunk block A_i of matrix $A_{N \cdot N}$ and the whole shared vector $B_{N \cdot 1}$
b) Processor P_i uses chunk block A_i of matrix $A_{N \cdot N}$ and the whole shared matrix $B_{N \cdot N}$

i.e. to the benefits of parallelization that should compensate its overhead.

Also, another condition is to use cache-intensive algorithms, or to reuse of data; otherwise, having cache memory could be useless since, in both executions, each access will generate a cache miss. Even more, the superlinearity appears for some range of the number of processors, or for some problem size, or for both.

B. A new challenge: How to scale?

Cache-intensive granular algorithms, whose data reuse complexity is proportional to the problem size, will benefit from bigger cache. Many Intel's multiprocessors use a marketing trick with a huge L3 smart cache. However, one can easily check that it is not shared among all cores, but only among part of them. For example, 6MB of total 12MB cache is shared between each group of two cores. In this case, vertical scaling will utilize more the last level cache. AMD multiprocessors usually have smaller L3 cache, but it is shared among all cores of the multiprocessor. Therefore, depending on the algorithm, appropriate processor and scaling type should be chosen in order to achieve the best speedup, potentially superlinear.

On the other hand, today's cloud elastic resources can also be scaled in different ways: horizontally, vertically or diagonally, each of which can offer various performance and possibility for achieving superlinear speedup [35]. Vertical scaling provides a better speedup, but horizontal offers more flexible scaling of resources, which can minimize the cost.

C. Is the superlinear speedup always our target?

Achieving superlinear speedup does not necessarily mean that customers will obtain the maximum achievement. For example, the cache associativity in CPUs and GPUs [36] can provide a huge performance drawbacks for a specific memory pattern reading, and achieving the superlinear speedup for those inputs is not enough, but other techniques, such as padding, should be used. In the workflow executions in parallel and distributed systems, customers usually use bi-objective optimizations to minimize the makespan and cost. These two parameters are opposite one to another. Minimizing the makespan produces greater cost and vice verse.

Cloud computing customers can set a deadline for the execution requiring minimal cost, rather than minimal makespan [37]. In these cases, budget constraints and reducing the race for the speedup can yield the reduced cost for the execution. For example, although superlinear speedup is achieved in Windows Azure cloud for matrix multiplication when virtual machine instances with Windows operating system are used, Linux virtual machine instances achieved better performance cost trade-off because they are cheaper.

On the other side, there is a risk of cloud resources performance variation and instance failure during the time. Increasing the budget by duplicating the tasks on more than one instance could mitigate those risks, in order to meet the deadline [38]. Sometimes, using a bigger instance executes the task faster, rather than waiting several minutes for the deployment time to start another smaller, but an appropriate instance, which reduces the turnaround time of an activity [39].

VII. Conclusion and Future Work

Since the race in processor's frequency (Gigahertz) was abandoned a decade ago, which in the meantime has been migrated into the TOP500 race [40] for hunting ExaFLOPS, this overview of superlinearity could have an impact in the supercomputers' architecture and design, since the goal of each parallelization is to achieve the maximal speedup, which is superlinear.

The defined taxonomy for various scalings and definitions of superlinearity can open new ways for parallel and distributed systems. Defining how much to scale the resources is insufficient. One needs to define how to scale. Algorithms that can benefit from greater cache memory should scale vertically, while those that need to finish more work in a given time, should scale horizontally.

This paper overviews many reasons and presents practical cases to achieve a superlinear speedup when an algorithm is executed using various scaling. The analysis can help to maximize the utilization of the parallel and distributed hardware [41].

This work summarizes and discusses several cases for the appearance of superlinearity. Superlinear speedup in non-persistent algorithms appears due to a smaller number of executed operations. Mainly the superlinear speedup performance in persistent algorithms occurs due to the increased cache resources in the parallel computer architectures, the prefetching of shared variables in shared memory organization, or better scheduling in heterogeneous environments. The effects of the shared memory architectures also impact the performance behavior of the granular and scalable algorithms. All these analyses will guide the developers of parallel implementation

not only how to parallelize a given problem, but to choose the most appropriate environment and scaling type in order to achieve the maximal speedup.

Additionally, this analysis opens many challenges, such as finding a correlation among parallel hardware's architecture and organization, a certain form of a parallelized algorithm, a parallelization technique, the server load and input problem size, and other possible factors that impact the existence of a superlinear speedup.

Further focus will be towards modeling the speedup by considering all these factors, as well as to determine an analytical relation of a complex computer system that will enable the conditions for superlinearity. Additionally, our challenge is to model the multidimensional space of superlinearity, which will determine the value of superlinearity by considering the problem size region and the region of the number of processors. Since this paper focuses on granular high performance algorithms, we would analyze and define the taxonomy for *scalable* algorithms, in which many tasks that are coming, are scattered among parallel processing units.

ACKNOWLEDGMENT

This work is supported by the European Union's Horizon 2020 research and innovation programme under the grant agreement 644179 ENTICE: dEcentralized repositories for traNsparent and efficienT vIrtual maChine opErations.

The authors would like to acknowledge networking support by the COST programme Action IC1305, Network for Sustainable Ultrascale Computing (NESUS).

REFERENCES

[1] E. Alba, G. Luque, and S. Nesmachnow, "Parallel metaheuristics: recent advances and new trends," *International Transactions in Operational Research*, vol. 20, no. 1, pp. 1–48, 2013. doi: 10.1111/j.1475-3995.2012.00862.x

[2] J. L. Gustafson, "Reevaluating Amdahl's law," *Communication of ACM*, vol. 31, no. 5, pp. 532–533, May 1988. doi: 10.1145/42411.42415. [Online]. Available: http://doi.acm.org/10.1145/42411.42415

[3] X. Ye, W. Dong, P. Li, and S. Nassif, "Hierarchical multialgorithm parallel circuit simulation," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 30, no. 1, pp. 45–58, Jan 2011. doi: 10.1109/TCAD.2010.2067870

[4] J. Rufino, A. I. Pereira, and J. Pidanic, "copssa - constrained parallel stretched simulated annealing," in *Radioelektronika (RADIOELEKTRONIKA), 2015 25th International Conference*, April 2015. doi: 10.1109/RADIOELEK.2015.7129044 pp. 435–439.

[5] T. Ciamulski and M. Sypniewski, "Linear and superlinear speedup in parallel fdtd processing," in *2007 IEEE Antennas and Propagation Society International Symposium*, June 2007. doi: 10.1109/APS.2007.4396642. ISSN 1522-3965 pp. 4897–4900.

[6] A. Gupta and D. Milojicic, "Evaluation of hpc applications on cloud," in *Open Cirrus Summit (OCS), 2011 Sixth*, Oct 2011. doi: 10.1109/OCS.2011.10 pp. 22–26.

[7] S. A. Tsaftaris, "A scientist's guide to cloud computing," *Computing in Science Engineering*, vol. 16, no. 1, pp. 70–76, Jan 2014. doi: 10.1109/MCSE.2014.12

[8] L. Liu, M. Zhang, Y. Lin, and L. Qin, "A survey on workflow management and scheduling in cloud computing," in *Cluster, Cloud and Grid Computing (CCGrid), 2014 14th IEEE/ACM International Symposium on*, May 2014. doi: 10.1109/CCGrid.2014.83 pp. 837–846.

[9] M. Gusev and S. Ristov, "Superlinear speedup in Windows Azure cloud," in *Cloud Networking (IEEE CLOUDNET), 2012 IEEE 1st International Conference on*, Paris, France, 2012, pp. 173–175.

[10] G. M. Amdahl, "Validity of the single-processor approach to achieving large scale computing capabilities," in *AFIPS Conference Proceedings*, vol. 30. AFIPS Press, Reston. Va., Atlantic City, N.J., Apr. 18-20 1967. doi: 10.1145/1465482.1465560 pp. 483–485. [Online]. Available: http://doi.acm.org/10.1145/1465482.1465560

[11] A. H. Karp and H. P. Flatt, "Measuring parallel processor performance," *Commun. ACM*, vol. 33, no. 5, pp. 539–543, May 1990. doi: 10.1145/78607.78614. [Online]. Available: http://doi.acm.org/10.1145/78607.78614

[12] Y. Shi, "Reevaluating amdahl's law and gustafson's law," Computer Sciences Department, Temple University, Tech. Rep. MS:38-24, Oct. 1996.

[13] M. Gusev and S. Ristov, "A superlinear speedup region for matrix multiplication," *Concurrency and Computation: Practice and Experience*, vol. 26, no. 11, pp. 1847–1868, 2013. doi: 10.1002/cpe.3102. [Online]. Available: http://dx.doi.org/10.1002/cpe.3102

[14] ——, "Resource scaling performance for cache intensive algorithms in Windows Azure," in *Intelligent Distributed Computing VII*, ser. SCI, F. Zavoral, J. J. Jung, and C. Badica, Eds. Springer International Publishing, 2014, vol. 511, pp. 77–86. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-01571-2_10

[15] J. L. Hennessy and D. A. Patterson, "Computer Architecture, Fifth Edition: A Quantitative Approach," MA, USA, 2012.

[16] M. Otte and N. Correll, "C-forest: Parallel shortest path planning with superlinear speedup," *IEEE Transactions on Robotics*, vol. 29, no. 3, pp. 798–806, June 2013. doi: 10.1109/TRO.2013.2240176

[17] A. F. P. Camargos, R. M. S. Batalha, C. A. P. S. Martins, E. J. Silva, and G. L. Soares, "Superlinear speedup in a 3-d parallel conjugate gradient solver," *IEEE Transactions on Magnetics*, vol. 45, no. 3, pp. 1602–1605, March 2009. doi: 10.1109/TMAG.2009.2012753

[18] G. Velkoski, S. Ristov, and M. Gusev, "Loosely or tightly coupled affinity for matrix - vector multiplication," in *Information Communication Technology Electronics Microelectronics (MIPRO), 2013 36th International Convention on*. Opatija, Croatia: IEEE, May 2013. ISBN 978-953-233-076-2 pp. 228–233.

[19] L. Djinevski, S. Ristov, and M. Gusev, "Superlinear speedup for matrix multiplication in GPU devices," in *ICT Innovations 2012*, ser. Advances in Intelligent Systems and Computing, S. Markovski and M. Gusev, Eds. Springer Berlin Heidelberg, 2013, vol. 207, pp. 285–294. ISBN 978-3-642-37168-4. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-37169-1_28

[20] N. Anchev, M. Gusev, S. Ristov, and B. Atanasovski, "Intel vs AMD: Matrix multiplication performance," in *Information Communication Technology Electronics Microelectronics (MIPRO), 2013 36th International Convention on*. Opatija, Croatia: IEEE, May 2013. ISBN 978-953-233-076-2 pp. 182–187.

[21] C. Augonnet, S. Thibault, R. Namyst, and P.-A. Wacrenier, "Starpu: A unified platform for task scheduling on heterogeneous multicore architectures," *Concurr. Comput. : Pract. Exper.*, vol. 23, no. 2, pp. 187–198, Feb. 2011. doi: 10.1002/cpe.1631. [Online]. Available: http://dx.doi.org/10.1002/cpe.1631

[22] G. Kosec and M. Depolli, "Superlinear speedup in OpenMP parallelization of a local PDE solver," in *MIPRO, 2012 Proceedings of the 35th International Convention*, 2012, pp. 389–394.

[23] J. L. Gustafson, "Fixed time, tiered memory, and superlinear speedup," in *Distributed Memory Computing Conference, 1990., Proceedings of the Fifth*, vol. 2, Apr 1990. doi: 10.1109/DMCC.1990.556383 pp. 1255–1260.

[24] I. A.-S. Ibrahim, H.-W. Loidl, and P. Trinder, "High-performance cloud computing for symbolic computation domain," *Journal of Computations & Modelling*, vol. 6, no. 1, pp. 107–133, 2016. [Online]. Available: http://www.scienpress.com/journal_focus.asp?main_id=58&Sub_id=IV&Issue=1771

[25] N. Theera-Ampornpunt, S. G. Kim, A. Ghoshal, S. Bagchi, A. Grama, and S. Chaterji, "Fast training on large genomics data using distributed support vector machines," in *2016 8th International Conference on Communication Systems and Networks (COMSNETS)*, Jan 2016. doi: 10.1109/COMSNETS.2016.7439943 pp. 1–8.

[26] P. E. McKenney, "Retrofitted parallelism considered grossly sub-optimal," in *Proceedings of the 4th USENIX Conference on Hot Topics in Parallelism*, ser. HotPar'12. Berkeley, CA, USA: USENIX Association, 2012, pp. 13–13. [Online]. Available: http://dl.acm.org/citation.cfm?id=2342788.2342801

[27] J. Ichnowski and R. Alterovitz, "Scalable multicore motion planning using lock-free concurrency," *IEEE Transactions on Robotics*, vol. 30, no. 5, pp. 1123–1136, Oct 2014. doi: 10.1109/TRO.2014.2331091

[28] S. Priyadarshi, C. S. Saunders, N. M. Kriplani, H. Demircioglu, W. R. Davis, P. D. Franzon, and M. B. Steer, "Parallel transient simulation of multiphysics circuits using delay-based partitioning," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 31, no. 10, pp. 1522–1535, Oct 2012. doi: 10.1109/TCAD.2012.2201156

[29] M. Monagan and R. Pearce, "Parallel sparse polynomial division using heaps," in *Proceedings of the 4th International Workshop on Parallel and Symbolic Computation*, ser. PASCO '10. New York, NY, USA: ACM, 2010. doi: 10.1145/1837210.1837227. ISBN 978-1-4503-0067-4 pp. 105–111.

[30] R. D. Phillips, L. T. Watson, and R. H. Wynne, "An smp soft classification algorithm for remote sensing," in *Proceedings of the 19th High Performance Computing Symposia*, ser. HPC '11. San Diego, CA, USA: Society for Computer Simulation International, 2011, pp. 104–110. [Online]. Available: http://dl.acm.org/citation.cfm?id=2048577.2048591

[31] P. Peschlow, A. Voss, and P. Martini, "Good news for parallel wireless network simulations," in *Proceedings of the 12th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, ser. MSWiM '09. New York, NY, USA: ACM, 2009. doi: 10.1145/1641804.1641828. ISBN 978-1-60558-616-8 pp. 134–142. [Online]. Available: http://doi.acm.org/10.1145/1641804.1641828

[32] M. Gusev and S. Ristov, "The optimal resource allocation among virtual machines in cloud computing," in *Proceedings of The 3rd International Conference on Cloud Computing, GRIDs, and Virtualization (CLOUD COMPUTING 2012)*, Nice, France, 2012, pp. 36–42.

[33] S. S. Bokhari, "Parallel solution of the subset-sum problem: An empirical study," *Concurr. Comput.: Pract. Exper.*, vol. 24, no. 18, pp. 2241–2254, Dec. 2012. doi: 10.1002/cpe.2800

[34] J. Maqbool, S. Oh, and G. C. Fox, "Evaluating arm hpc clusters for scientific workloads," *Concurrency and Computation: Practice and Experience*, vol. 27, no. 17, pp. 5390–5410, 2015. doi: 10.1002/cpe.3602 CPE-14-0161.R2. [Online]. Available: http://dx.doi.org/10.1002/cpe.3602

[35] S. Ristov, K. Cvetkov, and M. Gusev, "Implementation of a scalable l3b balancer," *Scalable Computing: Practice and Experience*, vol. 17, no. 2, pp. 79–90, 2016. doi: 10.1109/TE.2014.2327007

[36] S. Ristov, M. Gusev, L. Djinevski, and S. Arsenovski, "Performance impact of reconfigurable L1 cache on GPU devices," in *Computer Science and Information Systems (FedCSIS 2013), Federated Conference on*, Krakow, Poland, Sep. 2013, pp. 507 – 510.

[37] M. A. Rodriguez and R. Buyya, "Deadline based resource provisioningand scheduling algorithm for scientific workflows on clouds," *IEEE Transactions on Cloud Computing*, vol. 2, no. 2, pp. 222–235, April 2014. doi: 10.1109/TCC.2014.2314655

[38] R. N. Calheiros and R. Buyya, "Meeting deadlines of scientific workflows in public clouds with tasks replication," *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 7, pp. 1787–1796, July 2014. doi: 10.1109/TPDS.2013.238

[39] M. Mao and M. Humphrey, "Scaling and scheduling to maximize application performance within budget constraints in cloud workflows," in *Parallel Distributed Processing (IPDPS), 2013 IEEE 27th International Symposium on*, May 2013. doi: 10.1109/IPDPS.2013.61. ISSN 1530-2075 pp. 67–78.

[40] Supercomputers, "Top500," [retrieved: April, 2015]. [Online]. Available: http://www.top500.org/

[41] X. Ye, W. Dong, P. Li, and S. Nassif, "Maps: Multi-algorithm parallel circuit simulation," in *Proceedings of the 2008 IEEE/ACM International Conference on Computer-Aided Design*, ser. ICCAD '08, 2008. doi: 10.1109/ICCAD.2008.4681554. ISBN 978-1-4244-2820-5 pp. 73–78.

Education, Curricula & Research Methods

ECRM is a FedCSIS conference area aiming at interchange of information, ideas, new viewpoints and research undertakings related to university education and curricula as well as recommended methods of doing research in all computing disciplines, i.e. computer science, computer engineering, software engineering, information technology, and information systems. This area spans typical FedCSIS events (conferences, workshops, etc.) with rigorous paper submissions and review processes as well as panels, PhD and research consortia, summer schools, etc. Events that constitute ECRM are:

- IEES'16 - 1st International E-education Symposium—Education of the Future
- DS-RAIT'16—3rd Doctoral Symposium on Recent Advances in Information Technology

1st International E-education Symposium—Education of the Future

THE idea of the conference is a meeting of scientists, researchers and business sector in order to determine the development of the changes which are taking place in modern education. Solutions like cloud computing, virtualization or high performance computing are changing the nature and extent of education. Disappearance of barriers and boundaries between regions and sectors mean that education must adapt to the modern world. The conference participants will be able to discuss existing solutions, focus on the problems of their implementation and the risks and benefits that may involve.

TOPICS

- E-learning
- Distance education
- Education and new technologies
- Blended education
- Education through business practices
- Education and people with disabilities
- Innovative teaching and learning methodologies
- Information Technologies Supporting Learning
- Strategies for effective teaching

EVENT CHAIRS

- **Kaliński, Marcin,** Action Education Center, Poland
- **Karski, Michak,** Action Education Center, Poland
- **Siemek, Tomasz,** Action Education Center, Poland
- **Suszek, Halszka,** Polish-Japanese Academy of Information Technology, Poland

PROGRAM COMMITTEE

- **Banachowski, Lech,** Polish-Japanese Academy of Information Technology
- **Cieciora, Małgorzata,** Polish-Japanese Academy of Information Technology
- **Dąbrowski, Włodzimierz,** Warsaw University of Technology
- **Hälinen, Raimo,** Hame University of Applied Science, Finland
- **Järvinen, Pertti,** University of Tampere, Finland
- **Koponen, Erkki,** University of Tampere
- **Morańska, Danuta,** University of Dąbrowa Górnica
- **Pedaste, Margus,** University of Tartu
- **Wrycza, Stanisław,** University of Gdansk, Poland

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 903–912
DOI: 10.15439/2016F432
ACSIS, Vol. 8. ISSN 2300-5963

A blended learning model for practical sessions

Nuno Barreiro, Carlos Matos
Department of Computer Science,
Royal Holloway, University of London,
Egham Hill,
TW20 0EX, UK
Email: {Nuno.Barreiro, Carlos.Matos}@rhul.ac.uk

***Abstract*—We describe an education model that was developed and put in place to improve student engagement and attainment in a first year undergraduate programming course.**

The work is founded in a checkpoint-based formative assessment experiment undertaken for two years, the success of which is analysed in this document. The results provide evidence leading to a move towards a blended model of education, which requires the design of a software application to support the system. We present the main features of that application, covering aspects that range from traditional approaches and established delivery methods, to e-learning and MOOCs with, for instance, gamification.

This blended model of education fosters the development of a teaching practice that adapts to student diversity through informed teaching.

I. Introduction

FIRST year higher education undergraduate teaching is normally challenging for most areas: students face issues when transitioning from the school system; the class as a whole presents an inhomogeneous skill diversity. This is certainly the case in computer science, an area that attracts candidates with many different backgrounds. A particularly challenging topic within the discipline is the induction of students to programming [1], [2]: students taking a first year undergraduate course in that subject range from those who have never seen a line of code, to those who have completed software projects at school.

This document describes and evaluates a formative assessment experiment undertaken for two years in such a first-year course: the practical sessions were radically transformed by the introduction of a *checkpoint system*. The main goal of this new approach was to improve student engagement and attainment, but other objectives were also taken into consideration when designing the system.

Lab sheets with strategically placed low granularity checkpoints drive the practical sessions, which are supported by teaching assistants (TAs). Their role is to validate the checkpoints, provide help and feedback to the students, and gather data. Updated on a weekly basis, a checkpoints map measures how students progress through the course, and is used to inform teaching: early failure triggers quick remedial actions; high achievers are provided with challenging material; the contents of the sessions are adapted weekly to the overall pace and performance of the class. The experience of running this model for two years proved successful, with high student engagement and very positive feedback both from academic staff and students.

The analysis of the experiment has led to the development of a blended learning model to support practical sessions. Blended learning [3], [4], [5] uses both paradigms of traditional brick-and-mortar teaching, and online digital technologies, taking stock of techniques proven successful in e-learning and MOOCs, for example in the context of primary school mathematics education [6].

The course already uses Moodle to publish all its material, namely the weekly lab sheets used in practical sessions. The approach we propose includes extensions based on gamification [7] and a semi-automation of the checkpoint validating process. This is achieved through a software application, which design is described in this document.

The remainder of this paper is organised as follows: Section II provides context to both model and experiment, which are described in Section III. Section IV details the qualitative analysis of the experiment, and leads into a discussion on moving towards a blended model of education, presented in Section V. The checkpoint system is supported by an application described in Section VI. Section VII discusses related work, and Section VIII presents the conclusions.

II. Context

A. Teaching first year students

Formative assessment plays a major role in teaching computer science at our department, in particular when it comes to programming. Students start the degree with a great diversity of individual skills, which adds to the challenge of getting all of them to the end of first year with a similar body of knowledge. Treating that range of ability as an homogeneous body raises lack of engagement at both ends of the spectrum, which has been observed during many years of teaching experience in the department. As such, we need tools that help us both address failure at a very early stage, and encourage students who have a high level of familiarity with the subject to move on to more challenging material.

Each course delivery takes that diversity into account, to some extent. However, when it comes to first-year practical sessions, this aspect plays a major role: those sessions are, for many students, the first contact with written computer code. Students with different backgrounds proceed at different paces,

and a one-size-fits-all approach will inevitably fail to engage at least two kinds of students:

- those that are further dragging behind after each session, skipping most practical exercises and writing very little code;
- those that already know how to write some code and would appreciate a higher degree of complexity in the presented material.

An example of a first-year course with a strong practical component is CS1801 – Object-oriented programming. It spans over the two teaching terms and is taken by all single and joint-honours undergraduate computer science degrees. During the course, students learn the concepts of object-oriented programming that will be used through the entire degree and in their future profession. Acquiring those essential skills is critical to their success during the following years.

B. The CS1801 course

The course structure follows a traditional approach. Every week, there are two one-hour lectures, delivered to all the cohort in one go, and one three-hour practical session, organised in groups of approximately 30 students each. The uneven level of the students prevents the lecturer, during a practical session, from addressing the attendants as a whole. That problem is addressed by the presence of TAs for one-to-one interactions with students that have questions or are lost.

Each practical session is attended by three TAs, one per 10 students, which amounts to an interaction average time of 18 minutes per student per session. This can seem a lot of time, but in a three-hour session it corresponds to 6 minutes per hour – too short a time for detailed feedback.

Students are given a lab sheet that they are supposed to complete by the end of the session, but can carry from one session to the following ones, proceeding at their own pace. The lab sheets consist of several programming exercises, which the students have to code, compile and run. The exercises cover material from the lectures, and also include revisions from previous practical sessions. Completing the lab sheets plays an essential role in acquiring the skills of a proficient programmer: any developer knows that going from the pseudo-code written in a piece of paper to the real thing running on a computer requires a great deal of craftsmanship [2]. This learn-by-doing process of trial and error is regarded as one of the central aspects of education, and is not specific to computer science. For instance, In his 1984 book on *experiential learning* [8], Kolb emphasises the role of discovery and experience as sources of learning and development.

What we observe, however, is that many students fail to engage with the lab sheets: they do not tackle most of the exercises, leaving the sessions with important gaps in essential skills. This leads to a gradual block in their progress (most tasks require material from previous checkpoints). Furthermore, given that the acquisition of craftsmanship is a slow process requiring repetition, the more they drag behind, the less a chance they have to recover: it becomes a slow road to failing the course and, most probably, the first year.

The lack of engagement of some other students happens for different reasons: the presented material is trivial for their level; attending practical sessions is viewed as a mandatory boring activity. To address that issue, there are exercises, towards the end of the lab sheet, specifically aimed at highly-skilled students. They present interesting challenges requiring both problem solving and code proficiency. These are exercises that benefit from fruitful one-to-one interactions, which would need to be longer that the 18 minutes allocated to each student.

Overall, regardless of the motives, we observe a general tendency among students: even if the TAs are available and willing to help, most students will choose skipping exercises over calling the TAs.

III. An experiment in the practical sessions

A. Checkpoint system

During the last two academic years, with the main goal of overcoming the lack of engagement, but also in response to feedback provided by student and staff on the teaching of programming, we have conducted a formative assessment experiment in the CS1801 practical sessions: in each session, students have several checkpoints that help us (and them) keep track of their progress.

The process is coordinated by a lecturer, and follows the steps detailed below. We would like to note, as an early motivation of a blended model, that all these steps were conducted without any tool support, being therefore onerous in terms of time and dedication.

1) Lab sheet: A highly structured lab sheet is published on Moodle at the start of each practical session – this contains not only exercises, which are separated by well-identified (and strategically placed) checkpoints, but also examples and hints on how to approach the different tasks.

2) Practical session: During the session, each TA has a list with the checkpoints that every student has completed so far. When a student reaches a checkpoint and calls a TA, the checkpoint is verified – which consists in checking if the tasks were correctly completed, and the checkpoints list is updated accordingly.

3) Processing the information: Before the next practical session, the lecturer merges the information from the individual sheets, and publishes the updated list on Moodle. The history of those weekly updates is kept for reference – it shows the individual learning curves.

4) Publishing the results: At strategic points during the term, the lecturer produces a report with every student's achievements. This is used for internal monitoring of progression, and for deciding on any remedial actions.

B. Verifying the checkpoints

Every student has a different way of approaching checkpoints: some will regularly call the TAs at each checkpoint, others will prefer to have them verified in bulk. Nonetheless, whatever the approach, we have identified three stages at which students choose to have their checkpoints verified.

1) Complete solutions: Some students make sure that they have gone as far as they can, and present a program that not only works, but often will have different solutions for the same task. They expect the TA to give them advice on which solution is better, to comment on the code, and to encourage them to dig deeper.

2) Good attempts at a solution: Most students will consider a checkpoint ready to be verified as soon as they have written a snippet of code that compiles and presents a reasonable behaviour. Those programs will often miss particular cases and clever solutions, and the role of the TA is to inform the student about those alternatives.

3) Insufficient work: There are a few students that call the helper as soon as they have something that resembles the desired outcome. They expect to get major help from the TA in order to complete the task.

C. Acting on the results

One of the main aspects of using the checkpoint system is the ability to act on the gathered information. This is done at two levels: the information concerning a student allows for quick remedial actions; the information about the progression of the class as a whole provides a measure of the overall pace, informing the design of the lab sheets from session to session. If the entire class is lagging behind on a specific lab sheet, the following session can be lighter, allowing the students to catch up and relieve a possible frustration.

IV. Analysis of the experiment

A. Student engagement

The results of the experiment are presented in Table I, which only covers the first term of each year. There are significant differences in the approach to course delivery within both terms, which led us to focus mainly on the first term.

1) Differences between the two terms: In general, the level of student engagement drops during the second term, which is partially due to an increase in the number of checkpoints from one term to the other. Both terms have eleven practical sessions, but the last four sessions of term one are dedicated to the students that still have checkpoints to verify. All the other students no longer have to attend the sessions, dedicating this extra free time to other courses where their new skills will thrive in concrete applications (like, for example, games or robotics). For the students that are required to attend the remaining CS1801 sessions, there is no introduction of new material, and given the reduced number of students, the average support time from TAs is much higher.

We have observed that, during the second term of both years, the increase of 50% in the number of checkpoints (from 26 to 44 in 2014/15, and from 28 to 40 in 2015/15) leads to a decrease of approximately 75% in the overall rate of completion, when compared to the corresponding first term. The material studied in the practical sessions is also more complex during the second term, and the number of exercises is higher. All these contribute to the lack of engagement during the second term, and one of the reasons we went for the design of a blended model is to achieve, during the entire year, the success rate of term one.

TABLE I
Student engagement during the first term

	2014/15	**2015/16**
Number of checkpoints	24	28
Number of students	83	113
Students with more than 50% of the tasks verified	93%	88%
Students with more than 75% of the tasks verified	87%	77%

TABLE II
CS1801 results for non-repeating students

	2012/13 – 2013/14	**2014/15 – 2015/16**
Grade average	59.4%	66.2%
Failing students	19.6%	13.6%

2) The success of term one: In term one, as Table I shows, approximately 90% of the students complete more than half of the tasks. Given that this corresponds to the least amount of work expected from an average student, the percentage shows a very good level of overall engagement. Moreover, approximately 80% of the students complete more than three quarters of the tasks, which shows a high level of participation.

These numbers are not directly comparable to any measures from previous years (no control was done during practical sessions), but a measurable outcome is the level of success in the mid-year CS1801 test, which is a piece of formative assessment conducted on a weekly basis, from weeks 7 to 11 of term one. The students that achieve a grade above 85% are released from the lectures for the rest of the term. Usually, by the end of term one, most students will have succeeded in passing the test, but we have observed that, with the introduction of the checkpoint system, students are reaching the passing grade earlier.

3) Summative assessment outcomes: Summative assessment for CS1801 consists of several small pieces of coursework (10% of the final grade) and an exam (90% of the final grade).

The checkpoint system was introduced in 2014/15 and has been running for the last two academic years. Table II presents the outcomes of those two years, comparing them to the two previous years. In order to better measure the effect of the experiment, we have only considered students taking the exam for the first time – repeating students may have undertaken more programming courses, which gives them a clear advantage. The results indicate a significant increase in the success of summative assessment: higher grades and fewer failing students. However, these first observations still need to validated by a proper statistical analysis, for which we are gathering further data.

B. Further improvements

When looking at the numbers, and whilst the experiment can be considered successful, there are still 20% of the students

that are missing one quarter of the checkpoints, which is a significant number. To identify these students, we need a closer analysis of the process. The students have different levels of interaction with the TAs; the nature of those fundamental and time-consuming interactions reflects the level of the student, and raises different types of frustration.

1) Skilled students: Students with a prior knowledge of programming are very confident of their skills, and fail to realise that, although they have some programming abilities, there are specific requirements (for instance the style of the code and its readability by other programmers), which they are not familiar with. They regard the checkpoint system as a boring formality, and will skip it altogether. However, most of those students, when forced to check their work by an engaging TA, come to the conclusion that having skilled programmers reviewing the code is always a fruitful source of knowledge.

Skilled students tend to persist on recurrent mistakes and, in general, overestimate their abilities. It is not rare to find, in solutions presented by those more experienced students, many fragments of *bad code* that reveal misunderstanding and confusion. One such example is the following Java snippet (while being fully functional, it certainly leaves a bad impression on any trained programmer):

```
public static boolean negate(boolean a) {
  if (a == true) return false;
  else return true;
}
```

The student that produces such a piece of code is certain of having reached the goal of the task, but is missing important notions of programming. Those gaps in the acquired knowledge are mended by the interaction with TAs.

2) Struggling students: Students that are having a great deal of difficulty with programming feel uncomfortable asking for help, in particular when they look around and observe colleagues progressing at a faster pace than they are. When talking to those students, the variety of entry requirements becomes apparent: since neither mathematics nor programming are required, students that have taken those courses during their undergraduate studies are at an advantage. We also insist regularly on the fact that the learning outcomes should be measured by the end of the year, and that they should keep on trying. However, this does not prevent their frustration and/or impression of underachievement.

3) Limited resources: There is also the overhead of checking the exercises. The resources are limited, and sometimes students have to wait a considerable amount of time before getting checked. As they have to call a TA as soon one becomes free, this may end up being tiresome. Some students simply abandon the process midway.

C. Student satisfaction

There are several factors that affect the way students perceive the effectiveness of a course delivery. In the context of our experiment, we could identify two main perceptions among students: checkpoints are viewed either as a reward system, or as a remedial plan to address failure. Those different perceptions entail different ways of interacting with the system, leading to variable outcomes.

1) Feedback from students: Every term, students fill in a feedback questionnaire provided by Royal Holloway. Those questionnaires are highly detailed and present an overall picture of the course status. In the feedback forms for 2013/14 – term 1, 10 comments (in a total of 18) complain about the practical sessions. In the corresponding feedback forms from 2014/15, when the checkpoint system was introduced, 29 comments (in a total of 56) specifically mention good points about the practical sessions. The same tendencies were observed for the second term, although with less significant numbers. Anonymous comments included:

> "Compared to last year, the new format of the lab is much better."
>
> "The checkpoints in the labs were a good way of tracking progress."
>
> "Tasks on the lab sessions were chosen very well."

Besides those standard course feedback questionnaires, we foster further discussion with both staff and students to identify possible improvement opportunities. The yearly feedback provided by the students' committee for 2013/14 (the year preceding the introduction of checkpoints) mentioned that the practical sessions were delivered at a "too fast pace for people with little experience". During the next academic year, in the mid-term one-to-one meetings with their advisors, students were directly asked about the impact of checkpoints on their learning experience. The response was unanimous in acknowledging the system as extremely helpful. This impression was reiterated in the general feedback from the students' committee for 2014/15.

2) A rewarding system: The checkpoint system is perceived by students as a reward for their effort. Most students view this as an incentive, others as a way of monitoring their learning, and some even view the system as a competition. In general, they agree that the checkpoints really help them progress.

The responsible for the Student Experience in our Department has provided feedback about the system [9]: "The checkpoint system provides immediate formative feedback to the students. Students can identify problems and then in the lab discuss with the TAs what their problems are."

Given that a satisfied student is a better student, the positive feedback from students partially validates the effectiveness of the process. Furthermore, it also informs the approach, enabling a continuous fine-tuning of details in response to input from students.

3) Personal development: There is a growing awareness of the need to develop personal skills such as the ability to communicate, to present oneself with confidence, and to tackle unfamiliar problems. More generally, students' personal development should be understood to cover aspects of educational development (for example, the ability to use feedback to improve performance or to make use of educational resources such as Moodle or the library) and of career development

(for example, understand professional issues associated with careers in IT or the ability to work effectively in teams).

Questionnaires give evidence of students' perception of the quality of support that they received. However, to act on such responses and improve our support, one also needs to:

- understand the expectations that students have in relation to their personal development;
- ensure that students understand how, through the received feedback, they are given the opportunity to develop those personal skills.

That is, students need to take ownership of their development, and they cannot do so on a vague understanding of what support they are provided with.

Our Head of Department has provided feedback about the system in this context [9]: "The checkpoint system is an excellent contribution to allowing students take more ownership of their personal development. It allows them to progress at their own pace by getting quick feedback on their performance and understand what they need to improve."

D. Informed teaching

At first, the system was introduced to promote students' engagement, by having them asking for help and interacting with the TAs. However, it ended up covering many other aspects of the student's learning process that deserve more attention and development. The system can, namely:

- maximise success and minimise failure in practical courses by continuously giving feedback to students on their progression;
- measure the pace at which each student progresses, allowing for early actions to be taken on the learning difficulties that are detected.

1) Acting on failure: The weekly feedback provided by checkpoints allows academics to take action early on.

A co-responsible for the CS1801 lectures has provided feedback about the system in this regard [9]: "The checkpoint system provides me with much needed timely feedback on student performance in the weekly labs. I can quickly grasp how well students keep up with the course material and can spot particularly challenging topics. This is invaluable especially in the first year, where many students do not easily come forward after lectures or during office hours to ask for clarification or help. In addition, it allows to spot students who seem to not be engaging with the course; their personal advisors can then focus their attention on these students during the tutorial sessions for the course."

Together with his feedback mentioned in Section IV-C2, the system has been deemed by the responsible for Student Experience, with respect to providing a continuous overall picture of students' progression, to provide staff members with feedback that is "immensely useful for the department particularly with respect to quickly identifying progress in CS1801 for the progressions committee."

2) Promoting success: During the first year of the experiment, we have regularly provided optional exercises in the lab sheets, aimed at the students that welcome harder challenges. However, we have observed that most students would simply ignore those extra tasks. We were even surprised by students complaining about not feeling challenged in the practical sessions, and confessing at the same time that they had never attempted to solve any of those extra tasks.

Under the hypothesis that this phenomenon was mainly due to the fact that those extra tasks did not correspond to an extra checkpoint, we have made small change to the lab sheets published during the second year of the experiment: for each practical session, we have moved those optional exercises into a *platinum* checkpoint. And, in every weekly status update, we have published the results of the platinum checkpoints along all the other ones. The result was surprising: this simple alteration has triggered a much higher student participation in those extra tasks.

Observing the success of platinum checkpoints, we are led to consider the system as a tool enabling a fine analysis of the students' behaviour, both when addressing failure and when promoting a higher level of engagement for highly skilled students.

V. Towards a blended model of education

The success of the experiment lies in both the achieved results and, more importantly, the room for improvement that the process seems to present with respect to automation. Indeed, there are several aspects of the checkpoint system that would greatly benefit from automated or semi-automated mechanisms involving online techniques.

A. Assessment and feedback

The checkpoint system provides continuous feedback on each student's progress. With an average of four checkpoints per practical session, students reach the end of term with more than 40 checkpoints.

Since the results are made available every week, both students and academics have a detailed perception of individual learning curves throughout the term. We aim to combine those qualitative appreciations with a score (1-3 stars) and a badge-reward system (similar to the ones implemented in games and other scenarios [10]). Platinum checkpoints are a first simple example of what special badges can achieve, namely when it comes to encouraging students to complete the most challenging exercises.

B. Bringing the MOOCs to the classroom

MOOCs have been, in the past two years, gaining traction as a model for massively teaching students off campus. Although the results are mixed and debatable, some of their characteristics are remarkable: active learning (self-pacing and instant feedback) and gamification (with, for instance, badge-awarding systems).

What we intend is to bring those successful aspects of MOOCs to the classroom. According to Anant Agarwal, CEO

of edX, this process is about "taking ... the technologies we are developing in the large and applying them in the small to create a blended model of education" [11].

We propose to bring teaching, assessment and gamification techniques from MOOCs to the classroom, namely:

- students are able to follow the sessions at their own pace;
- progression to the next level only occurs when the previous levels have been completed, understood and verified;
- badges encourage students to perform better in the achievement of specific goals;
- the visualisation of progress bars and learning curves provides students with an overall perception of their performance.

C. Advantages of an automated system

We have identified the following possible improvements.

1) A first response line: If the code written by students is submitted to an IT system instead of being directly present to a TA, a few automatic tests will provide feedback to students on how to correct common mistakes. Several systems for automatic assessment of programming assignments exist, with different levels of support and feedback, as discussed in [12]. In our case the complexity of the tests can vary, but they constitute an automated line of action that will rely on human assistance only when necessary: the TAs will be called fewer times and their overall availability will increase. Furthermore, the automation trivially orders TA requests by submission time, which frees the student to start working on the next checkpoint after the submission of the code, instead of recurrently looking for an available TA.

2) Speeding up feedback: The automated submission also provides different levels of feedback, going beyond the typical interaction with the TAs. After passing the code through several tests, the student may get specific feedback on particular mistakes without interacting with a TA. Nonetheless, the TAs will always be available to provide further detail, if necessary. When the system is not able to provide automated feedback, a TA is called. Whatever the reason to call for human help, the responding TA will have a device with available information about the student's checkpoint history, the current checkpoint, and the code issues. This will speed up the process of interaction, during which the TAs can spend a considerable amount of time going through the checkpoints and looking for problems in the code before being able to help the student.

3) Increased promotion of different paces: The lab sheets are published on a weekly basis, rather then in bulk. The rationale behind this procedure is to allow the progressive unfold of a story during the term, for which the sessions sequence is essential. However, this approach has two main drawbacks:

- the faster students will finish the lab session in less than two hours, even when completing some optional harder tasks, which leaves out an hour that could be used to progress to the next tasks;
- some students will get stuck on a particular task and, although we insist that students may move on and come back later to that particular problem, some future tasks may depend on the left-behind task, and students get sometimes confused about what they know, and what they still have to acquire.

An automated system would overcome these issues by using a graph of task precedences that would allow all the tasks to be published at the same time, disabling those that require previous tasks to be completed. As soon as a student completes one checkpoint, the system will display any new available checkpoints. Also, any disabled checkpoint will display the tasks that are required for its unlocking. This also doubles as a knowledge map, where students can easily find paths leading to the acquisition of a specific skill.

D. Further aspects of an automated system

Beyond the advantages described in Section V, the checkpoint system automation would also provide a realtime source of information about the students' progress, both individually and as a whole. Different ways of looking at the data provide different insights. An automated system could easily provide charts with realtime statistical analysis, that can include the progression curve of one student or of the class, a classification of students according to several filter options (number of completed checkpoints, total score, ranking), comparison between several years and sessions, etc.

VI. Designing the application

In order to fulfil the mentioned goals and provide a better service to both students and academics, we have designed a web-based application that automates the checkpoints process. Running on the browser of any computer (desktop, laptop, tablet or phone), the application may be used by lecturers and students alike. The basic functionality of the checkpoints mechanism will be supported by a core module, and an API will provide the ability to extend the application with additional useful features that can enhance the learning experience of students, individually and as a whole. A proof-of-concept prototype has been developed and is functional, covering some of the essential features. The final application is under development since early 2016.

The outcome of this development is a software system that will not be specific to computer science, and that can be used by any discipline to provide formative assessment in practical sessions. The overall goals of the system are to:

- monitor student achievement and performance;
- reward the student with a score corresponding to those achievements and performances;
- measure the difficulty of exercises by looking at the overall performance of the class;
- adapt at runtime the delivery of the practical sessions to the level of the class, and propose course revisions for the next years;
- give tailored content to students that have specific difficulties or a higher level of achievement;
- provide a methodology for exercise-based sessions that can be used in a systematic way.

A. Overall description

During a typical session, students are working on computers where they can submit their checkpoints. TAs will have portable devices (tablets) that they use to identify students in need of help, and to verify the checkpoints.

When the users log into the system they have access to a general view of their account where they can either edit personal information, or select a course. Inside the course, each type of user has different options, as described below.

1) Students: Students are able to view rankings, individual achievements, and an overall picture of the checkpoints. In that view, students access the graph of checkpoints for the course, as well as the checkpoints that are completed and those that are unlocked, given the current score. There is also a graph with the learning curve of the student with respect to that course.

2) Academics: Academics are able to edit the graph of checkpoints, add new checkpoints, and monitor all the students' activities. They are also able to create rewards awarded to students when they reach certain points of the graph, or certain scores.

3) Administrators: Administrators have a different kind of access, allowing them to create courses and to perform functionalities such as management, monitoring and maintenance.

B. Definition of checkpoint

A checkpoint is a set of tasks that students have to perform. Usually that set is small, as fragmentation is an essential aspect of checkpoints: overcoming several small challenges is more rewarding than dealing with a complex task that will take long to complete [13].

The contents of a checkpoint can have several declinations: tutorial with simple tasks, complex tasks requiring problem solving, quizzes, etc.

The students' general view of all the checkpoints in the application highlights those that are accessible. Regardless of accessibility, each checkpoint displays a description of its content. This allows the student to understand, with the help of the graph, which checkpoints need to be completed before acquiring a specific knowledge.

The students' work on a checkpoint is independent of the software application, which only manages checkpoints and deals with the submitted material. This allows different kinds of disciplines to use the system. In the case of CS1801, for example, students work on a Linux server using an editor of their choice, and a Java compiler. Once the student considers that the tasks are completed, they upload the relevant files to the application and wait for the verification process.

C. Verifying a checkpoint

The solutions to the checkpoint tasks are verified by an automated system and/or a TA. The verification by an automated system requires the solution of the tasks to be submittable to an electronic system. In the absence of that feature, the system will always rely on a TA to verify a checkpoint. In this document, we will consider that the students are able to submit a file to get the checkpoint verified by an automated testing mechanism.

The verification process may have several outcomes, including the need to go back to the tasks in order to correct critical mistakes. One crucial aspect of a checkpoint is that is can be attempted several times: each attempt can result in a better outcome, which may also be awarded a higher score. This characteristic is aligned with literature on gamification [14], and with the article by Karpicke and Blunt [15] where they conclude:

> "Research on retrieval practice suggests a view of how the human mind works that differs from everyday intuition. Retrieval is not merely a readout of the knowledge stored in one's mind; the act of reconstructing knowledge itself enhances learning."

Hence, the students overcome mistakes by trial and error, which is a procedure they are familiar with [8].

Figure 1 shows the flow involved in the verification of a checkpoint. Each step of the process is detailed below.

1) Working on a checkpoint: The student is working on the tasks of the checkpoint. Some help from the TAs may be required, but students are encouraged to work on their own and submit solutions they deem adequate.

2) Submitting the tasks: When a student finishes a set of tasks, they upload the relevant files to the application, or fill in a form with the outcomes of the task (depending on the nature of the task). In the specific case of CS1801, the student submits a Java source file. As soon as the submission is done, the checkpoint becomes unavailable until either the verification is completed, or the process is cancelled by the student.

While waiting for the verification, the student may start working on any other unlocked checkpoint. The student may also choose not to submit the solutions to the tasks, but this means that no new checkpoints will ever be unlocked, and the student will eventually get stuck.

3) Testing the submission: A series of automated tests are performed. This first line of action looks for common mistakes that are easily detectable, and for which some feedback can be provided without any human intervention. Sophisticated tests can also be performed, if such a test suite is available for the discipline that is using the system. In the case of CS1801, a series of automated tests allow a complete verification of the checkpoint (see Section VI-D).

4) Getting feedback: If the tests are not passed, the system will try to provide some automated feedback to the student. If that feedback cannot be produced, the application puts the submission in the TAs' queue: a TA will be notified and the student will eventually be approached to get some oral feedback, or to have the checkpoint validated. This may happen when, for instance, the student has an easy-to-fix error in the code.

5) Validating a checkpoint: The validation of a checkpoint can be automatically completed by the application, or require the intervention of a TA. In the latter case, the TAs use their tablets to capture a QR code on the computer of the student (uniquely identifying the student-checkpoint pair).

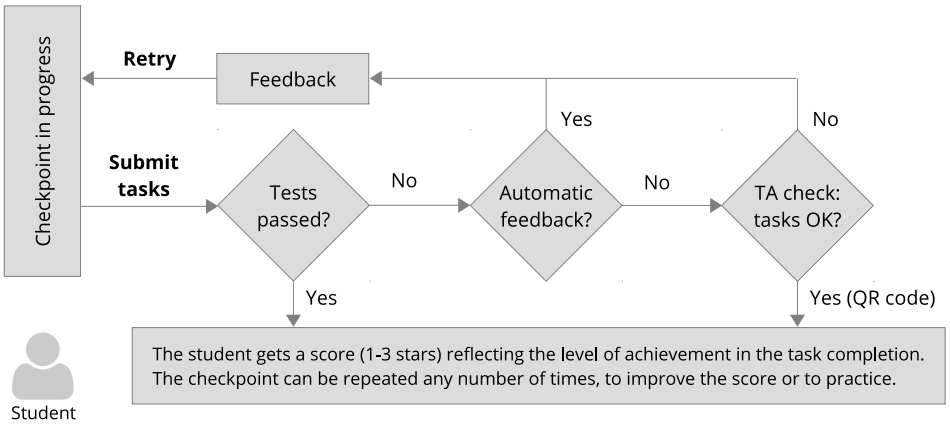


Fig. 1. Verifying a checkpoint with the application

This will bring up an interface that allows the TA to score the checkpoint. When the checkpoint gets validated, some new checkpoints may become unlocked.

6) Score: When the checkpoint is validated, the student gets a score (1-3 stars) based on several criteria. This score can be automatically provided by the system, according to the outcomes of the test suite, or it can be attributed by the TA.

D. The submission assessment for CS1801

The submission of a checkpoint consists of uploading the file with the code to the server. The file is then submitted to a series of tests that check if the code corresponds to what was asked in the set of tasks. The tests are sequential and follow the steps described below, shown in Figure 2.

1) Compilation: The first requirement to complete a checkpoint is that the submitted code compiles. If the compilation issues warnings, the student may be given some automated feedback that will suggest improvements on the code.

2) Unit testing: The exercises have specific outcomes that allow unit testing. For example, a program that defines a class can be tested by creating an object of that class and calling its methods. There are several tests for each exercise; those tests can be weighted in order to assess the student accordingly. The number of stars the student is awarded for the checkpoint will increase with the number of passed tests.

3) Checking the style: Students often write code that does the required job, but presents stylistics flaws that, even if they are optimised by the compiler, show some gaps in the acquired knowledge. One of the goals of CS1801 is to get students to write programs that are readable and free from *bad code*.

E. Prototype

We have developed an early prototype that covers the following functionality:

- log in with testing accounts;
- set different groups of checkpoints (emulating a course);
- organise, in each group, a linear sequence of checkpoints;
- view the stages of completion of each group of checkpoints;
- view the unlocked checkpoints;
- view the locked checkpoints, and the precedences that unlock them;
- verify a checkpoint using a QR code.

F. Extensions to the prototype

The application extends the prototype with the following additional features, most of them via an API.

1) Remote authentication: The users can use their usual credentials, which are securely fetched from a remote server.

2) Timed checkpoints: When a student starts a set of tasks, a timer is activated to measure the time spent on that checkpoint.

3) Graphical representations of data: Several views of the students' progress, individual and by groups, will help both students and teachers have snapshots of the students' status and the learning curves.

4) Graph of checkpoints: Students can progress according to their actual skills, in a non-linear fashion, instead of having to follow a sequence of checkpoints that may not be the most adequate to their learning curve.

5) Automated verification process: The submissions are subjected to an automated suite of tests.

6) Enhanced gamification: Student's achievements will be awarded trophies, rewarding effort – not just success; those trophies are viewed by other students and establish peer motivation.

7) Students' feedback on the exercises: Academic staff will receive a constant measurement of how students perceive the exercises, allowing them take early action and adapt the exercises to the needs of the students.

8) Look and feel: The new features will require a revision of the application user interface.

VII. Related Work

A. Checkpoints

The notion of checkpoints applied to support teaching of programming has been used for some time. The concrete goals have ranged from formal assessment [16] to testing the rate of student progress and improve the process [17]. The

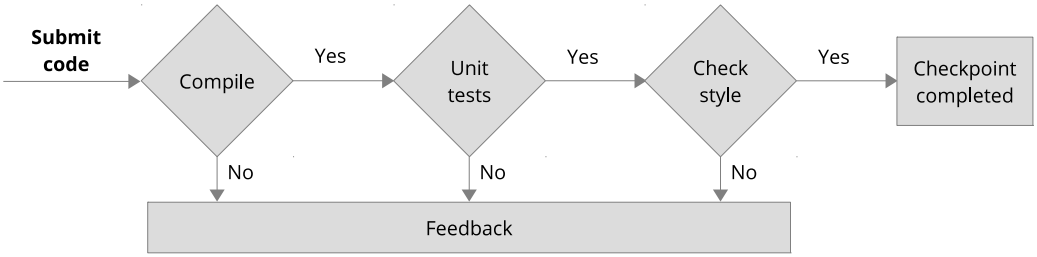


Fig. 2. Automated test sequence for CS1801

granularity at which these are used also varies significantly. In some models, checkpoints are set after small exercises are completed [18], whilst others set these only when more considerable tasks are finalised [19]. In the case of our approach, the checkpoints are designed at a low level of granularity to allow both a detailed view of student progress, and a good understanding of the difficulty in each set of tasks. However, the application is also suited to different approaches, like for instance long sessions focusing on one single difficult task.

The way the completion of these checkpoints is taken into account also varies. The results are logged in various fashions, ranging from simple manual accounting to free [20] and commercial tool support [21], [22]. The model we present includes the design of a toolchain that takes into account the required granularity of the specific checkpoints, fosters a good user experience for both assessors and students (e.g. the use of QR code recognition so that manual input is minimised), allows for a graph of dependencies, and is integrated with some gamification concepts.

Checkpoints are also used either as a means to assist in supporting students in traditional programming laboratory settings [21] or as part of automated assessment [20]. Our approach still includes traditional sessions, where students are presented with exercises and benefit from the support of teaching assistants. But it also takes advantage of modern techniques, such as those widely used in Continuous Integration, to reduce the amount of required intervention and to give immediate feedback to students.

B. Serious games and gamification

Serious games were introduced in order to facilitate purposes other than simply entertainment. These ranged from games in a general sense as described by Abt in 1970 [23], to digital games as presented more recently by Sawyer and Rejeski [24] and Michael and Chen [25]. Whilst sharing some concepts with serious games, *gamification* [7] focuses on using game elements to enhance several activities, rather than having users specifically play a game. These activies include learning [26], [27], business [28], and even self-help [29]. The model we describe uses the concept of gamification to improve student engagement in practical sessions (and with the goal of improving attainment), whilst resulting in better learner analytics.

C. Blended learning

Blended learning [3], [4], corresponds to models where the Internet and digital media are combined with traditional classroom settings. In his 2012 report [30], Friesen presents the following definition:

> "Blended learning" designates the range of possibilities presented by combining Internet and digital media with established classroom forms that require the physical co-presence of teacher and students.

This technique was used in the context of our approach as a means to reap benefits from both worlds. Traditional practical sessions allow close support and quick and helpful feedback, whilst the digital content and automatic checks accelerate the administrative process and allow students to focus in learning how to think computationally.

D. Automatic checks

Automatically checking student programming work has been researched for many years [31], [32], and is still an active area [12], [33]. Whilst it is possible to use these techniques to address, for instance, issues of scale, the level of feedback and assistance still lags behind what can be achieved with direct support from experienced teaching assistants. Notwithstanding this gap, it is a useful approach that can be integrated with the traditional one. In the model we describe, this has been achieved by including automated checks to address a first line of issues, followed by human support when required.

VIII. Conclusion

We have presented a blended model of education motivated by a successful formative assessment experiment conducted in the practical sessions of a first-year undergraduate programming course. That model will benefit from the support of an application, which design was informed by building and testing a functional prototype. The application is currently under development, and will be deployed in 2016/17.

One of the main improvements in the final application, compared to the prototype, is the strong gamification component. We have observed how the introduction of platinum checkpoints has motivated some students to engage with harder tasks. We expect, similarly, the star-based scores and the

achievement badges to stimulate a higher level of overall engagement. Moreover, given the current trends in gamification research, this is an aspect with much room for improvement: the API will easily accommodate third-party awards like, for instance, Mozilla Open Badges [10].

We will evaluate and monitor the initiative mainly through: student feedback (advisor tutorials, one-to-one meetings and student feedback questionnaires); results of further formative and summative tests, including the final CS1801 examination.

This will allow us to consider any necessary improvements and evolve the model accordingly, before moving to a wider dissemination of the application that, by design, is not specifically targeted at teaching programming. Nonetheless, the system has the potential of creating a baseline of tools and techniques for other courses in our department that include exercise-based practical sessions, providing a clear added value: it allows the identification of students needing additional support, and measures the effectiveness/difficulty of particular tasks. By increasingly adopting this technique, the Department of Computer Science, as a whole, can have a better and up-to-date perception of student achievement and, hence, proactively make changes as necessary.

We also plan, in the future, to introduce a mechanism through which students can give a simple quantitative feedback on the quality and perceived difficulty of each set of tasks. Based on those pieces of information – as measured through performance and as perceived by students – academics will move yet another step towards taking more informed pedagogical decisions.

REFERENCES

[1] M. McCracken, V. Almstrum, D. Diaz, M. Guzdial, D. Hagan, Y. B.-D. Kolikant, C. Laxer, L. Thomas, I. Utting, and T. Wilusz, "A multi-national, multi-institutional study of assessment of programming skills of first-year cs students," in *Working Group Reports from ITiCSE on Innovation and Technology in Computer Science Education*, ser. ITiCSE-WGR '01. New York, NY, USA: ACM, 2001. doi: 10.1145/572133.572137 pp. 125–180.

[2] T. Jenkins, "On the difficulty of learning to program," in *Proceedings of the 3rd Annual Conference of the LTSN Centre for Information and Computer Sciences*, vol. 4, 2002, pp. 53–58.

[3] D. R. Garrison and H. Kanuka, "Blended learning: Uncovering its transformative potential in higher education," *The internet and higher education*, vol. 7, no. 2, pp. 95–105, 2004. doi: 10.1016/j.iheduc.2004.02.001

[4] C. J. Bonk and C. R. Graham, *The handbook of blended learning: Global perspectives, local designs*. John Wiley & Sons, 2012. ISBN 9781118429570

[5] F. A. Marco, V. M. R. Penichet, and J. A. G. Lázaro, "Drawer: an innovative teaching method for blended learning," in *Proceedings of the 2013 Federated Conference on Computer Science and Information Systems*, M. P. M. Ganzha, L. Maciaszek, Ed. IEEE, 2013, pp. 727–734.

[6] R. Murphy, L. Gallagher, A. E. Krumm, J. Mislevy, and A. Hafter, "Research on the use of Khan Academy in schools: Research brief," *SRI International*, 2014.

[7] S. Deterding, D. Dixon, R. Khaled, and L. Nacke, "From game design elements to gamefulness: Defining "gamification"," in *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments*, ser. MindTrek '11. New York, NY, USA: ACM, 2011. doi: 10.1145/2181037.2181040 pp. 9–15.

[8] D. A. Kolb, *Experiential Learning: experience as the source of learning and development*. Prentice Hall, 1984. ISBN 9780132952613

[9] N. Barreiro and C. Matos, "Checkpoint system documentation," Department of Computer Science, Royal Holloway, University of London, Tech. Rep., 2016, internal document.

[10] "Mozilla Open Badges," accessed: 2016-05-09. [Online]. Available: http://openbadges.org/

[11] A. Agarwal, "Why MOOCs (still) matter," 2013, TED talk by Anant Agarwal, CEO of edX, Accessed: 2016-05-06. [Online]. Available: http://bit.ly/1kZwi9Y

[12] P. Ihantola, T. Ahoniemi, V. Karavirta, and O. Seppälä, "Review of recent systems for automatic assessment of programming assignments," in *Proceedings of the 10th Koli Calling International Conference on Computing Education Research*, ser. Koli Calling '10. New York, NY, USA: ACM, 2010. doi: 10.1145/1930464.1930480. ISBN 978-1-4503-0520-4 pp. 86–93.

[13] T. Amabile and S. Kramer, *The progress principle: Using small wins to ignite joy, engagement, and creativity at work*. Harvard Business Press, 2011. ISBN 9781422198575

[14] J. McGonigal, *Reality Is Broken: Why Games Make Us Better and How They Can Change the World*. Penguin Group , The, 2011. ISBN 9780143120612

[15] J. D. Karpicke and J. R. Blunt, "Retrieval practice produces more learning than elaborative studying with concept mapping," *Science*, vol. 331, no. 6018, pp. 772–775, 2011. doi: 10.1126/science.1199327

[16] J. Bennedsen and M. E. Caspersen, "Assessing process and product: a practical lab exam for an introductory programming course," *Innovation in Teaching and Learning in Information and Computer Sciences*, vol. 6, no. 4, pp. 183–202, 2007. doi: 10.11120/ital.2007.06040183

[17] S. Cvetkovic, R. Seebold, K. Bateson, and V. Okretic, "CAL programs developed in advanced programming environments for teaching electrical engineering," *IEEE Transactions on Education*, vol. 37, no. 2, pp. 221–227, 1994. doi: 10.1109/13.284998

[18] University of Edinburgh, "Physics 2A Scientific Programming in JAVA," 2009, course notes, Accessed: 2016-05-07. [Online]. Available: http://www2.ph.ed.ac.uk/~aturner/teaching/Scientific-Programming/documentation/booklet.pdf

[19] D. Parsons and P. Haden, "Programming osmosis: Knowledge transfer from imperative to visual programming environments," in *Conference of the National Advisory Committee on Computing Qualifications*, 2007.

[20] S. Zhigang, S. Xiaohong, Z. Ning, and C. Yanyu, "Moodle plugins for highly efficient programming courses," in *1st Moodle Research Conference*, 2012.

[21] E. Seung, "Examining the development of knowledge for teaching a novel introductory physics curriculum," Ph.D. dissertation, Purdue University, 2007.

[22] "WebAssign," accessed: 2016-05-07. [Online]. Available: http://webassign.net/

[23] C. C. Abt, *Serious games*. Viking Press, 1970. ISBN 9780670003136

[24] B. Sawyer and D. Rejeski, "Serious games: Improving public policy through game-based learning and simulation," 2002.

[25] D. R. Michael and S. L. Chen, *Serious Games: Games That Educate, Train, and Inform*. Muska & Lipman/Premier-Trade, 2005.

[26] K. M. Kapp, *The gamification of learning and instruction: game-based methods and strategies for training and education*. John Wiley & Sons, 2012. ISBN 9781118096345

[27] B. Kumar and P. Khurana, "Gamification in education-learn computer programming with fun," *International Journal of Computers and Distributed Systems*, vol. 2, no. 1, pp. 46–53, 2012.

[28] B. Burke, "Gamification: Engagement strategies for business and IT," Gartner Inc, Tech. Rep. G00245563, 2012.

[29] A. M. Roepke, S. R. Jaffee, O. M. Riffle, J. McGonigal, R. Broome, and B. Maxwell, "Randomized controlled trial of SuperBetter, a smartphone-based/internet-based self-help tool to reduce depressive symptoms," *Games for health journal*, vol. 4, no. 3, pp. 235–246, 2015. doi: 10.1089/g4h.2014.0046

[30] N. Friesen, "Report: Defining blended learning," 2012, accessed: 2016-05-08. [Online]. Available: http://learningspaces.org/papers/Defining_Blended_Learning_NF.pdf

[31] J. Hollingsworth, "Automatic graders for programming classes," *Communications of the ACM*, vol. 3, no. 10, pp. 528–529, Oct. 1960. doi: 10.1145/367415.367422

[32] J. B. Hext and J. W. Winings, "An automatic grading scheme for simple programming exercises," *Communications of the ACM*, vol. 12, no. 5, pp. 272–275, May 1969. doi: 10.1145/362946.362981

[33] N. A. Rashid, L. W. Lim, O. S. Eng, T. H. Ping, Z. Zainol, and O. Majid, *Advanced Computer and Communication Engineering Technology: Proceedings of ICOCOE 2015*. Springer International Publishing, 2016, ch. A Framework of an Automatic Assessment System for Learning Programming, pp. 967–977.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 913–920
DOI: 10.15439/2016F356
ACSIS, Vol. 8. ISSN 2300-5963

Pitfalls of E-education: from multimedia to digital dementia?

R. Robert Gajewski
Warsaw University of Technology
Al. Armii Ludowej 16, 00-637
Warszawa, Poland
Email: rg@il.pw.edu.pl

***Abstract*—This paper presents lessons learned from nearly 25 years long experiences with different forms of E-education. All experiences are definitely positive but during conducted research many pitfalls and traps were recognized and observed. Widely used multimedia materials do not motivate weak students to learn. Instead of learning they do prefer to watch materials in a passive way. Mobile learning in which all materials are available also for smartphones increased this attitude to learning. All quizzes and tests even very sophisticated cannot replace a real exam. Knowledge of the answers on hundreds of questions is not equal to the real knowledge of a certain field. Flipped classroom paradigm forcing to learn at home was not accepted by students. Moreover, E-education creates chances for e-cheating. All these pitfalls and traps lead to the conclusion that E-education is not a straightforward remedy for all current education problems.**

I. Introduction

IN 450 B.C Confucius aid: "tell me and I will forget, show me and I may remember, involve me and I will understand." As outlined by many researchers individuals remember much more details and information as well as for longer if they are more involved in the learning process. In 1946 Dale published his famous Cone of Experience [1]. Dale stated that the cone device can be a visual metaphor of learning experiences, where the different kinds of audio-visual materials are arranged in the order of increasing abstractness as one proceeds from direct experiences (see Fig. 1). One of the later extensions of this idea is a common opinion that individuals generally remember: 10% of what they read, 20% of what they see, 50% of what they see and hear, 70% of what they say and write and 90% of what they say as they perform a task. Moreover, the entire process of learning is split into two parts: passive learning and active learning.

Blooms Taxonomy proposed in 1956 [2] by a panel of educators chaired by Benjamin Bloom is a categorization of learning objectives as well as activities split up into three areas: cognitive (mental skills, knowledge), affective (feelings, emotional areas and attitude) and psychomotor (manual and physical skills). The cognitive domain most significant in higher education requires mental abilities and also knowledge. Within this domain one can find six major categories outlined from the most straightforward: knowledge, comprehension, application, analysis, synthesis and finally evaluation.

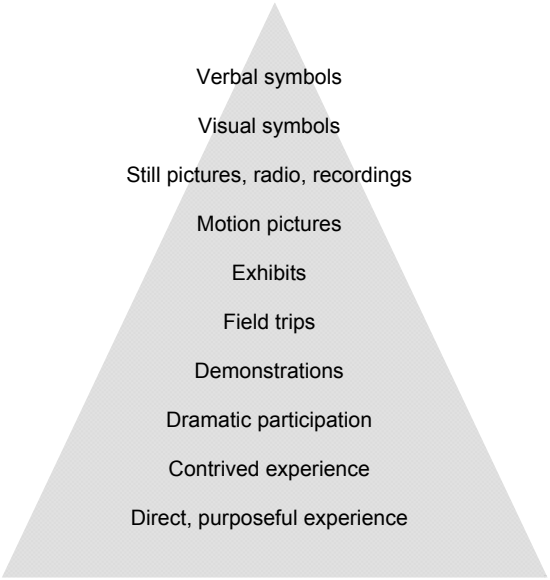


Fig. 1. Dale Cone of Experience

In the middle of 1990's the cognitive domain has been modified. Titles associated with different types have been transformed from nouns to verbs. Moreover, their order has been somewhat changed. Bloom's Revised Taxonomy [3] demonstrates to a greater extent the active way of thinking and also consists of six different categories: remembering, understanding, applying, analyzing, evaluating and finally creating. This taxonomy much better accounts for completely new behaviors and multimedia technology innovations (see Fig. 2).

E-education enables to take into account different learning styles [4]. Such approach increases costs of education but also increases its efficiency. In the class "one type of delivery" should satisfy all participants. E-education enables addressing different materials for different learning styles [5], [6] even for Computer Science and Informatics Courses.

This work was supported by 504/01921/1088/40 grant.

II. MULTIMEDIA

Multimedia materials were prepared for all subjects taught by the Division of Information Technologies (DoIT), namely Information Technologies, Fundamentals of Computing and Computational Methods in Civil Engineering in the form of podcasts – personal on demand broadcasts. First podcasts prepared by DoIT had the form of screencasts – "digital recordings of computer screen output often containing audio narration". Screencasts contain software animations helping students to learn how to use software. The second kind of podcasts are slidecasts – "audio podcasts combined with slideshow". Slidecasts have the form of knowledge clips – short explanatory presentations of a particular problem and its solution. The last kind of multimedia materials prepared by DoIT are webcasts – "media presentations distributed over the Internet using streaming media technology to many simultaneous viewers". In fact webcasts were lecture captures which were recorded and later distributed as podcasts. Tenths of hours of podcasts stored on an educational portal helped a lot during classes but did not have an expected impact on quality of learning process measured in terms of grades obtained by students.

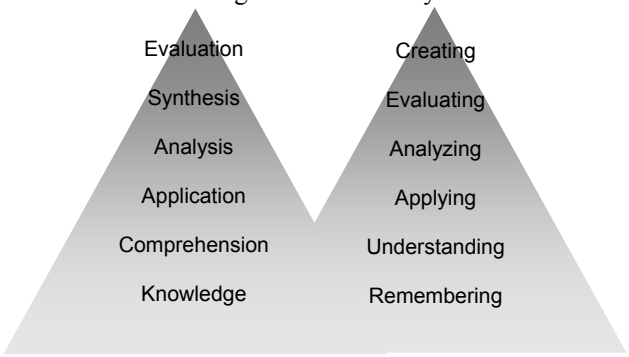


Fig. 2. Bloom's Revised Taxonomy

First podcasts were prepared in a Flash format. Nowadays this format is not available for iOS and the newest version of Android so for current mobile devices. All Flash podcasts were converted to the MP4 format or created from scratch. Now on the YouTube channel there are more than 200 clips. Their total length exceeds 50 hours.

Questionnaires performed in the academic year 2011/12 showed that having a full range of podcasts not all of students were fully satisfied by them. They were pleased by the quality and ease of use as well as by their availability in the mode 24/7. Moreover, they stressed positively that such an approach addressed different learning styles. However, in additional field of questionnaire reserved for remarks some of the students complained that the part of computer laboratories was boring for them because they repeated what was recorded in screencasts. All podcasts were designed as an additional, supplementary and auxiliary tool and all teaching and learning activities were conducted in a traditional way. Students were "taught" at the university how to use software and they were supposed to solve individual problems at home. In many cases solving problems was too difficult for them.

Starting from the academic year 2012-2013 in some of the groups podcasts were used in a different way. Students were asked to watch podcasts at home. During classes they should be prepared to use software without any problems and to solve particular problems using it. First results of this experiment were to some extend promising - students gained better scores in this mode, but they were not very keen to spend time at home watching podcasts. Students do prefer to "be taught" during classes. This problem can be easily solved by adding a simple point to subject regulations – students should be prepared to computer laboratories and this fact is checked by means of a test before the class. In fact, according to European Credit Transfer System (ECTS) an average student should spend learning at home the same amount of time as at the university. It is much more effective to watch passive by nature screencasts at home and solve problems with a tutor in the class than the other way round.

III. FLIPPED CLASSROOM

Idea of inverting education is already nearly fifteen years old. One of the first papers in that field was published in 2000 [7]. This paper describes two parts of subject taught at Miami University while using the inverted classroom concept and analyzes the outcomes. Numerous technologies offered completely new possibilities for students to learn away from the classroom, while a school period was used to perform collaborative experiments and worksheets. Authors of the paper concluded that the idea of inverted classroom offers alternatives for various learning styles and report that students favor that strategy. A different outline and evaluation of flipped education within a huge, primarily based on lectures, computer science course was published in 2002 in [8]. In this project new multimedia and video streaming application eTech was employed to change a course. In-class lectures were substituted by recorded lectures and auxiliary materials which could be viewed by students in the Internet independently. This make it possible to utilize the live period in the class for team problem solving facilitated by tutors. Another interesting paper in that field was published one year later in 2003 [9]. Within a series of five experiments hundreds of students from two different universities supervised by three different professors and six different teaching assistants took one semester long course in the field of casual and statistical reasoning in both traditional or online format. Within the frame of this project pre and post test results were compared. Features of the online experience which were helpful and which were not helpful were identified as well as most and least effective student learning strategies. Three years later a paper evaluating a web lecture intervention in a human–computer interaction course was published [10]. By utilizing lectures available in the Web before class more in-class period was used engaging students with hands-on

tasks. Class time was spent on learning by doing rather than learning by listening. In 2007 Gannod presented his work in progress on how to use podcasts in an inverted classroom [11]. One year later Helmick presented integrated online courseware for computer science courses [12]. Last but not least in 2008 a paper describing how to use the inverted classroom to teach software engineering was published [13]. Idea of flipped classroom was fully described in three books recently published by Bergmann and Sams [14], Bretzmann [15] and Walsh [16].

The research concerning students' satisfaction with flipped classroom was conducted in academic year 2013/2014 on a group of 222 students studying in Polish (PL) and a group of 51 students studying in English (EN). Out of 222 PL students the questionnaire was filled by 211 students which makes 95%. Similar data are for students studying in English. Questionnaire was filled by 49 out of 51 students. One third of students studying in English were foreigners.

Questionnaire used in this survey consists of fifteen closed form questions and 6 opened form questions. Due to the nature of answers all questions were divided into three groups. In order to compare the results of survey with other outcomes some of the questions were based on similar surveys: first one conducted in Canada [17] and second one described in blog Flipping with Kirch conducted by Mary Kirch from United States.

Scale of answers for all first five questions is from "strongly agree" to "strongly disagree". Results for Polish language and English language students were compared with surveys from Canada. The first of the asked questions was about the level of engagement in traditional classroom instructions and flipped classroom (see Fig. 3). The second question from that group was about potential recommendation of a flipped classroom to a friend (see Fig. 4).

40% of students studying in Polish language strongly disagree or disagree with the statement what is in accordance with the observation, that nearly half of the students was not interested in traditional classes. Answers of students studying in English language are closer to the answers from survey conducted in Canada.

For this question answers of students studying in Polish and English languages are similar but they definitely differ from the results of survey conducted in Canada. Nearly six times more students studying in Polish language in comparison to Canadian agree or strongly agree with the statement that they would not recommend flipped classroom to a friend.

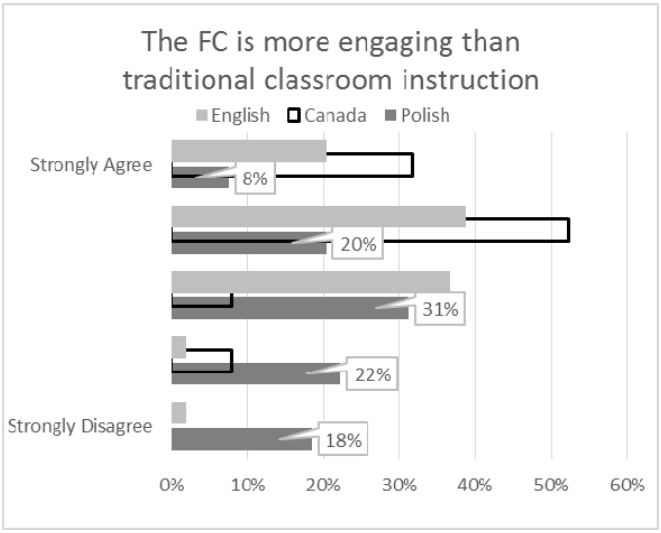


Fig. 3. Answers on question 1.1

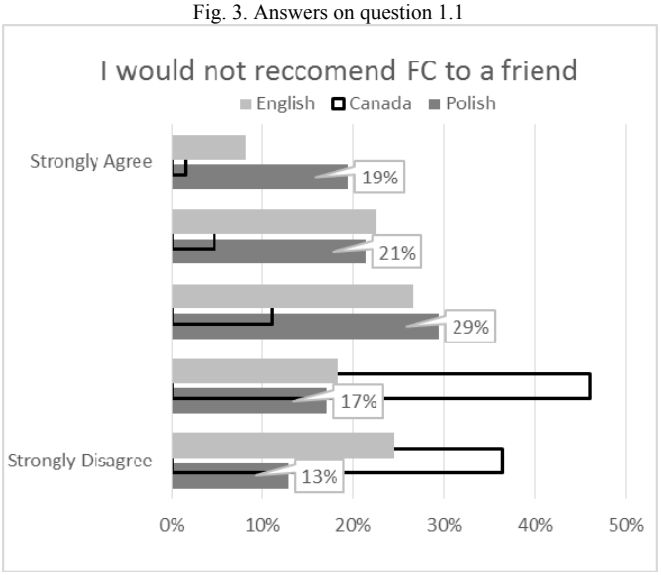


Fig. 4. Answers on question 1.2

Next question (statement) was very simple – I like watching lessons on video (see Fig. 5). In this case answers for all three groups were very similar.

Fourth question in this group of questions was about better motivation to learn in the flipped classroom mode (see Fig. 6). In the case of this question answers of students studying in Polish language differ from the answers of two other groups. Nearly 40% of them strongly disagree or disagree with that statement that they are more motivated to learn in a flipped classroom mode.

The last question in this group is about improvement of learning in the flipped classroom mode (see Fig. 7).

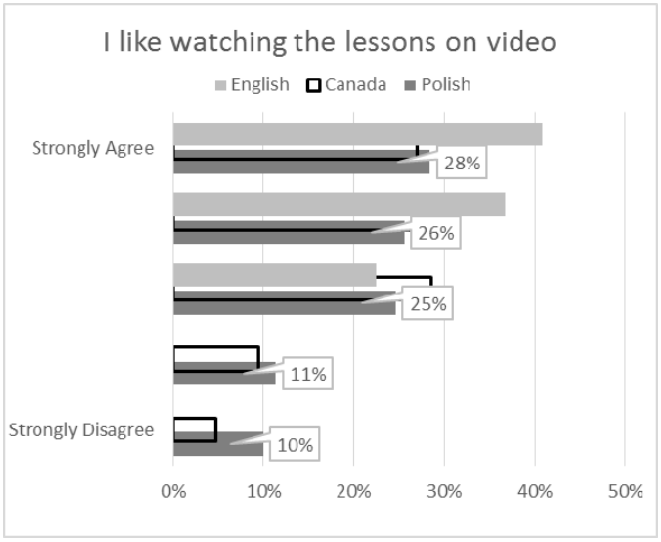


Fig. 5. Answers on question 1.3

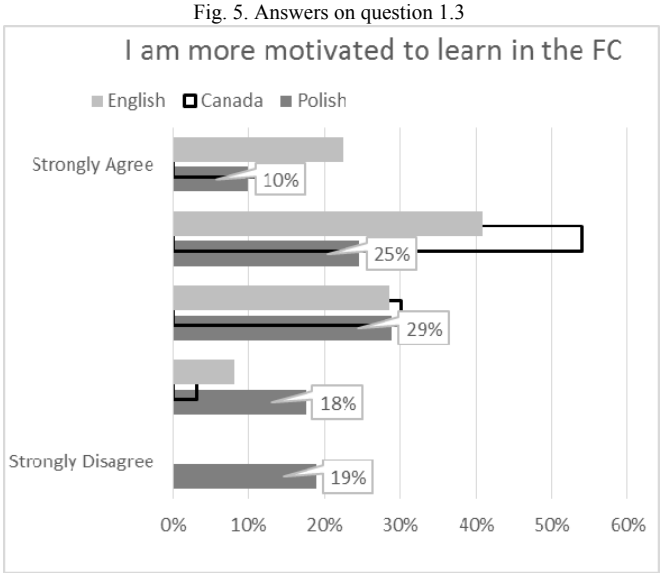


Fig. 6. Answers on question 1.4

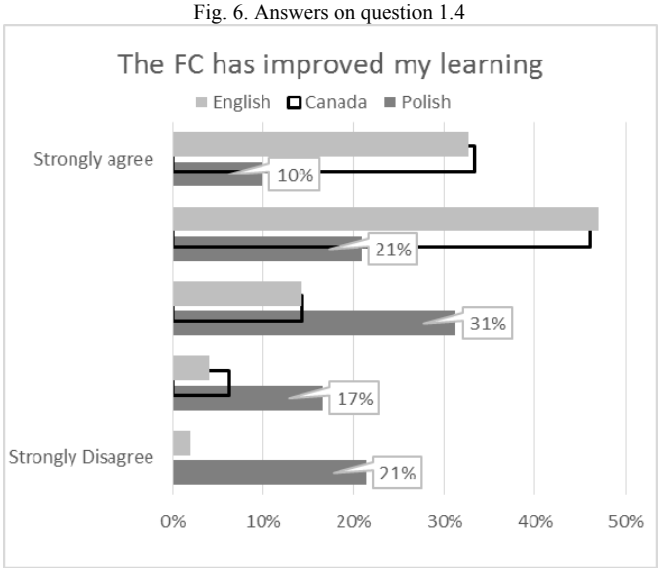


Fig. 7. Answers on question 1.5

IV. Tests and Quizzes

From the very beginning different forms of tests and quizzes were used mainly due to increasing number of students. It was not possible to check their knowledge in a classical way. Databases for subjects taught by DoIT consist of hundreds of different types of questions available on the Moodle platform like: calculated, simple calculated, calculated multi choice, description, matching, multiple choice, short-answer, numerical and True/False. After twenty years' experiences are rather sad. Students are trying to memorize answers on the questions rather than to understand the appropriate part of material. Quizzes were also used in a flipped classroom experiment. There were quick tests consisting of up to ten questions checking knowledge gained before the class at home from podcasts. Nowadays quick tests are placed at the end of classes and they force students to make notes during the class.

In order to help students to prepare for tests flashcards were used (see Fig. 8). This tool invented by Sebastian Leitner [18] can support learning treated as memorizing but rather not as understanding.

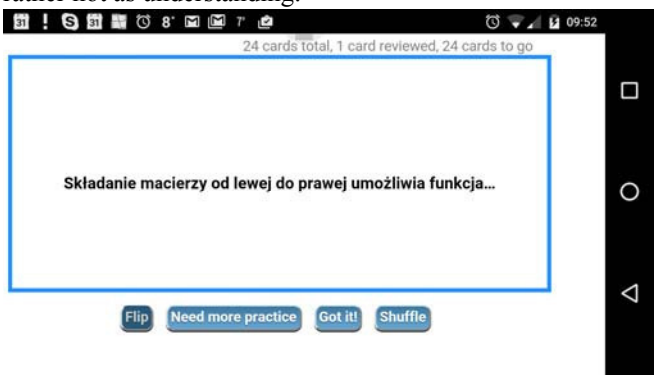


Fig. 8. Sample flashcard for mobile device

During the present academic year for traditional lectures conducted in a lecture theater at the university clickers were used. Instead of special hardware devices specialized software and smartphones were utilized (see Fig. 9).

V. Mobile Learning

Shift from instructional design to e-Learning [19] was the first step of educational revolution. The next step will be devoted to transforming the system of delivery of education and training [20]. The best historical overview of m-Leaning [21] can be found in Handbook of m-Learning [22]. In the last decades there were many shifts in learning and learner-centered pedagogies and theories. Mobile courses from the field of Computer Sciences or Engineering require the usage of new and effective design strategies [23] and implementation of appropriate learning theories [24]. From a technical point of view instead of producing different applications for different mobile operating systems used on various mobile devices it is more efficient to create courses available through web browsers also on mobile devices (see Fig. 10).

Fig. 9. Sample Poll Everywhere question

Changes in the course of engineering calculations and their programming which took place in the last few years show differences and similarities between traditional learning, e-Learning and m-Learning [25]. Regarding time traditional learning is frequently constrained by school hours, e-Learning by a time of access to computer while in the case of m-Learning in fact there are no time constraints. Learning can occur anywhere where access to the network is possible. Traditional learning is rather not personalized which contrasts with personalized e- and m-Learning. Traditional learning is definitely formal while m-Learning is rather informal – e-Learning can be formal and informal. Last but not least traditional learning is not spontaneous while m-Learning is highly spontaneous.

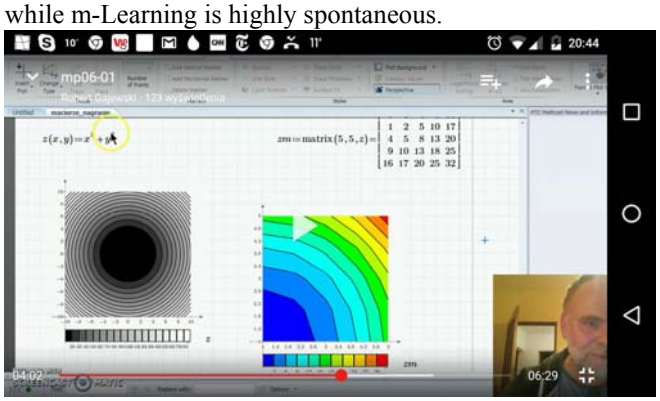


Fig. 10. Sample mobile podcast

VI. Cheating

Cheating is perhaps as old as education. Mavis [26] wrote about college cheating as a function of subject and situational variables in 1962 and Haines [27] about college cheating as effect of immaturity, lack of commitment, and the neutralizing attitude in 1986 and also ten years later [28]. But nowadays due to the information and communication technology it is much easier to cheat so it starts to be a crucial problem. One can say that e-Learning caused e-Cheating as presented by Jones in [29] and in [30]. Cyber cheating is a crucial problem in an information technology age [31]. There were tenths of papers written on this subject. Their review can be found in [32]. The answer on the important philosophical question why cheating is so wrong is given in [33]. More information about this subject can also be found in [34].

In order to learn what is the attitude towards cheating among students two surveys were conducted. In order to learn what are the cultural differences between different countries first survey was based on survey from Gettysburg in USA and second was based on survey conducted in Monash University in Australia. Similar comparative analysis on students' perception and attitudes towards academic dishonesty between the students in China and United Stated was done by Zhou and Lan in [35]. Research on cheating was also conducted in Dubai [36] and in Philippines [37], [38].

The first survey was conducted during the first week of classes in October 2015 and was based on the test from Gettysburg. Total number of responses was 203. Number of students registered for the subject was 221. Total number of the questions in this survey was 24. Answers on the question "have you ever reported another student you suspected of cheating" are rather similar (see Fig. 11).

Answers for two next questions ("have you ever interrupted a student who was cheating" and "did you ever cheat during high school") show definitely bigger differences in attitude to cheating (see Fig.12 and Fig. 13.).

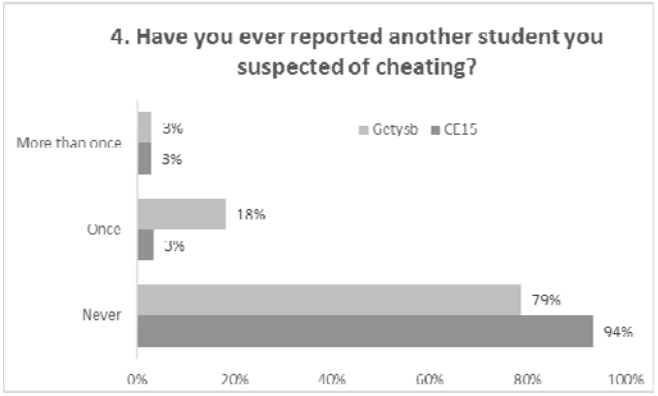


Fig. 11 Comparison of the answers on question 4 from Gettysburg survey

The second survey was conducted during the last week of classes in January 2016. Total number of responses was 179. Number of students attending classes was 201. Questionnaire of this survey is fully based on the questionnaire used in 2000 during the survey conducted in Australia at Monash University and at Swanbourne University which results were published in [39]. The same survey was conducted ten years later and results were compared in [40]. The most important part of both surveys

consists of 18 scenarios. For each of them answers are given using Likert's scale [41] with the answers ranging from 1 – acceptable to 5 – not acceptable.

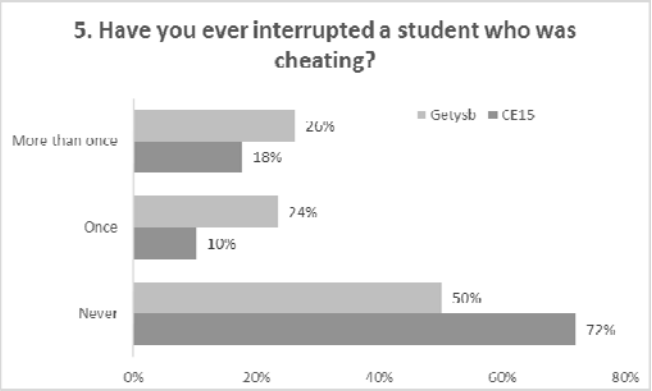


Fig. 12. Comparison of the answers on question 5 from Gettysburg survey

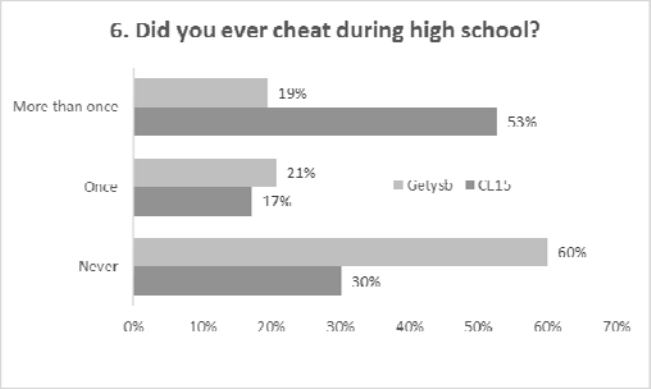


Fig. 13. Comparison of the answers on question 5 from Gettysburg survey

What can be easily learned from Table 1 is that in Australia a positive change has occurred among students over the decade with regard to cheating awareness, acceptability and practice. Results of survey conducted in Poland are much worse than Australian from the year 2000. Especially big differences are in the case of the three following scenarios:

- Copying material from the book or from the Internet,
- Swapping assignments with another person,
- Using a hidden sheet of paper with important facts during an exam.

The first problem can be generally solved by means of plagiarism checkers on the university level. The second one should be mainly solved by instructor manually. The third one which can be generally called as using unauthorized materials during exams can be solved by means of specialized IT tools. For the last mentioned scenario – using unauthorized materials – worth mentioning is a relatively small difference in mean values of acceptability (4.59, 4.64 and 4.32) and a very big difference in practice (4%, 2% and 53%).

Perhaps cheating is as old as education, and the only change is in methods (see Fig. 14).

Fig. 14. No cheating
http://i.dailymail.co.uk/i/pix/2014/05/03/article-2619387-1D89BFD000000578-597_634x397.jpg

VII. Discussion

"Digital Dementia", a term coined by the top German neuroscientist Manfred Spitzer in his 2012 book with the same title [42], is a term used to describe how overuse of digital technology is resulting in the breakdown of cognitive abilities in a way that is more commonly seen in people who have suffered a head injury or psychiatric illness. Spitzer proposes that short-term memory pathways will start to deteriorate from underuse if we overuse technology.

Nowadays it is getting more and more important to know how does the brain work [43]. In many situations treating Internet as a natural source of information is replaced by FoMO – Fear of Missing Out [44]. Multimedia which ten years ago were a very promising educational tool nowadays make students unwilling to learn. They do prefer to watch video in a passive way rather than to learn in an active way.

Tests widely used in E-education to control knowledge are also a kind of pitfall. What is really controlled by tests this is an art of passing tests and knowledge of the answers on numerous questions. The best databases updated continuously cannot replace the teacher asking questions. Knowledge and understanding of a certain field is not equal to the knowledge how to answer on numerous questions.

Last but not least E-cheating seems to be the biggest problem in E-education. Different IT tools can make cheating more difficult but will not fully stop it.

Lessons learned from twenty five years long research in the field of e-Learning show that cultural differences should be taken into account while introducing new solutions like in the case of flipped classroom. Fifteen years after the question "how to change the unchanging" [45] was raised many things has changed. Despite all pitfalls and traps there is the only one answer on the question to be e- or not to be in the field of education. But we should be e- in a more rational way. There is still need for deep look into all three dimensions of learning [46] or how professionals learn in practice [47].

Acknowledgment

The author wishes to thank Marcin Jaczewski for his support during flipped classroom experiment and all surveys. Thanks are also due to Tomasz Dubilis and Tomasz

TABLE I.
COMPARISON OF RESULTS OF SURVEYS

Scenario	Acceptability			Practice		
	2000 mean	2010 mean	2015 mean	2000 %	2010 %	2015 %
Showing assignment work to a lecturer for guidance	2.07	2.21	1.68	37	42	67
Posting to an Internet newsgroup for assistance	2.07	2.28	1.77	27	34	68
Two students collaborating on an assignment meant to be completed individually	2.54	3.20	2.65	44	36	54
Resubmitting an assignment from a previous subject in a new subject	2.82	2.99	2.49	27	17	41
Submitting a friend's assignment from a past running of the subject	2.86	3.46	3.03	34	20	32
Copying the majority of an assignment from a friend's assignment, but doing a fair bit of work yourself	2.98	3.37	2.98	31	21	39
Not informing the tutor that an assignment has been given too high a mark	3.08	3.29	2.82	17	16	39
Being given the answer to a tutorial exercise worth 5% by a class mate if the computer you used has problems	3.76	4.29	3.03	7	3	9
Copying material for an essay from a text book	3.81	4.19	3.71	22	10	34
Copying material for an essay from the Internet	3.85	4.28	4.29	23	10	33
Obtaining a medical certificate from a doctor to get an extension when you are not sick	3.94	4.02	3.35	12	3	9
Swapping assignments with a friend, so that each does one assignment, instead of doing both	3.96	4.45	3.37	9	3	37
Copying another student's assignment from their computer without their knowledge and submitting	4.18	4.62	3.94	7	3	7
Copying all of an assignment given to you by a friend	4.30	4.62	3.92	10	3	10
Hiring a person to write your assignment for you	4.51	4.62	3.97	3	1	6
Using a hidden sheet of paper with important facts during an exam	4.59	4.64	4.32	4	2	53
Hiring someone to sit an exam for you	4.65	4.69	4.32	3	0	5
Taking a student's assignment from a lecturer's pigeonhole and copying it	4.72	4.72	4.29	4	2	3

Warda from for their IT assistance. Last but not least the author wishes to thank all students who filled many very long questionnaires in the beginning and at the end of semester.

REFERENCES

[1] E. Dale, *Audio-visual methods in teaching*. The Dryden Press, 1946.
[2] B. S. Bloom, *Taxonomy of Educational Objectives Book 1: Cognitive Domain*, 2nd edition. Addison Wesley Publishing Company, 1984.
[3] L. W. Anderson, D. R. Krathwohl, P. W. Airasian, K. A. Cruikshank, R. E. Mayer, P. R. Pintrich, J. Raths, and M. C. Wittrock, *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives, Abridged Edition*, 2nd ed. Pearson, 2000.
[4] R. R. Gajewski, "O stylach uczenia sie i I-edukacji," *E-Mentor*, vol. 3, no. 4[11], pp. 28–35, 2005.
[5] O. Mironova, T. Rüütmann, I. Amitan, J. Vilipõld, and M. Saar, "Computer Science E-Courses for Students with Different Learning Styles," in *Proceedings of the 2013 Federated Conference on Computer Science and Information Systems*, 2013, pp. 735–738.
[6] O. Mironova, I. Amitan, J. Vendelin, M. Saar, and T. Rüütmann, "Strategies for the Individualization of an Informatics Course," in *Proceedings of the 2014 Federated Conference on Computer Science and Information Systems*, 2014, vol. 2, p. pages 835–840., DOI: 10.15439/2014F259.
[7] M. J. Lage, G. J. Platt, and M. Treglia, "Inverting the Classroom: A Gateway to Creating an Inclusive Learning Environment," *J. Econ. Educ.*, vol. 31, no. 1, pp. 30–43, 2000, DOI: 10.1080/00220480009596752.
[8] J. Foertsch, G. Moses, J. Strikwerda, and M. Litzkow, "Reversing the Lecture/Homework Paradigm Using eTEACH® Web-based Streaming Video Software," *J. Eng. Educ.*, vol. 91, no. 3, pp. 267–274, Jul. 2002, DOI: 10.1002/j.2168-9830.2002.tb00703.x.
[9] R. Scheines, J. Smith, G. Leinhardt, and K. Cho, "Replacing lecture with Web-based course materials," *J. Educ. Comput. Res.*, vol. 32, pp. 1–26, 2003.
[10] J. A. Day and J. D. Foley, "Evaluating a Web Lecture Intervention in a Human-Computer Interaction Course," *IEEE Trans Educ*, vol. 49, no. 4, pp. 420–431, Nov. 2006, DOI: 10.1109/TE.2006.879792.
[11] G. C. Gannod, "Work in progress; Using podcasting in an inverted classroom," in *Frontiers In Education Conference - Global Engineering: Knowledge Without Borders, Opportunities Without Passports, 2007. FIE '07. 37th Annual*, 2007, p. S3J–1–S3J–2, DOI: 10.1109/FIE.2007.4418119.
[12] M. T. Helmick, "Integrated Online Courseware for Computer Science Courses," in *Proceedings of the 12th Annual SIGCSE Conference on Innovation and Technology in Computer Science Education*, New York, NY, USA, 2007, pp. 146–150, DOI: 10.1145/1268784.1268828.
[13] G. C. Gannod, J. E. Burge, and M. T. Helmick, "Using the Inverted Classroom to Teach Software Engineering," in *Proceedings of the 30th International Conference on Software Engineering*, New York, NY, USA, 2008, pp. 777–786, DOI: 10.1145/1368088.1368198.
[14] J. Bergmann, A. Sams, *Flip your classroom: reach every student in every class every day*. Eugene, Or: International Society for Technology in Education, 2012.
[15] J. Bretzmann, *Flipping 2.0*. Bretzmann Group LLC, 2013.
[16] K. Walsh and P. J. Walsh, *Flipped Classroom Workshop in a Book*, 1 edition. Kelly Walsh, 2013.

[17] G. B. Johnson, “Student Perceptions on the Flipped Classroom.” The University of British Columbia, 2013.
[18] D. Bryson, “Using Flashcards to Support Your Learning,” J. Vis. Commun. Med., vol. 35, no. 1, pp. 25–29, Mar. 2012, doi: 10.3109/17453054.2012.655720.
[19] M. W. Allen, Designing Successful e-Learning: Forget What You Know About Instructional Design and Do Something Interesting. Pfeiffer, 2007.
[20] M. Ally, Ed., Mobile Learning: Transforming the Delivery of Education and Training. Athabasca University Press, 2009.
[21] H. Crompton, “A Historical Overview of m-Learning,” in Handbook of Mobile Learning, 1 edition., Z. L. Berge and L. Muilenburg, Eds. New York: Routledge, 2013, pp. 3–14.
[22] Z. L. Berge and L. Muilenburg, Eds., Handbook of Mobile Learning, 1 edition. New York: Routledge, 2013.
[23] I. A. Alshalabi and K. Elleithy, “Effective M-learning design Strategies for computer science and Engineering courses,” Int. J. Mob. Netw. Commun. Telemat., vol. 2, no. 1, pp. 1–11, 2012, doi: 10.5121/ijmnct.2012.2101.
[24] I. A. Alshalabi, Sa. Hamada, and K. Elleithy, “Research Learning Theories that Entail M-Learning Education Related to Computer Science and Engineering Courses,” Int. J. Eng. Sci., vol. 2, no. 3, pp. 88–95, 2013.
[25] H. Crompton, “Mobile Learning: New Approach, New Theory,” in Handbook of Mobile Learning, 1 edition., Z. L. Berge and L. Muilenburg, Eds. New York: Routledge, 2013, pp. 47–58.
[26] E. Mavis and S. E. Feldman, “College cheating as a function of subject and situational variables,” J. Educ. Psychol., vol. 55, no. 4, pp. 212–218, 1964, DOI: 10.1037/h0045337.
[27] V. J. Haines, G. M. Diekhoff, E. E. LaBeff, and R. E. Clark, “College cheating: Immaturity, lack of commitment, and the neutralizing attitude,” Res. High. Educ., vol. 25, no. 4, pp. 342–354, 1986, DOI: 10.1007/BF00992130.
[28] G. M. Diekhoff, E. E. LaBeff, R. E. Clark, L. E. Williams, B. Francis, and V. J. Haines, “College cheating: Ten years later,” Res. High. Educ., vol. 37, no. 4, pp. 487–502, 1996, DOI: 10.1007/BF01730111.
[29] K. O. Jones, J. Reid, and R. Bartlett, “E-learning and E-cheating,” presented at the 3rd E-Learning Conference, Coimbra, Portugal, 2006, pp. 45–48.
[30] K. O. Jones, J. Reid, and R. Bartlett, “E-learning and E-cheating,” Commun. Cogn. Monogr., vol. 41, no. 1–2, pp. 61–70, 2008.
[31] K. O. Jones, J. Reid, and R. Bartlett, “Cyber Cheating in an Information Technology Age,” Digithum, no. 10, pp. 19–28, 2008, DOI: 10.7238/d.v0i10.508.
[32] Z. Ercegovac and J. V. Richardson, “Academic Dishonesty, Plagiarism Included, in the Digital Age: A Literature Review,” Coll. Res. Libr., vol. 65, no. 4, pp. 301–318, Jul. 2004, DOI: 10.5860/crl.65.4.301.
[33] M. Bouville, “Why is Cheating Wrong?,” Stud. Philos. Educ., vol. 29, no. 1, pp. 67–76, Aug. 2009, doi:10.1007/s11217-009-9148-0.
[34] R. R. Gajewski, “IT in Educational Management: Can it Support Solution of e-Cheating Problem?,” presented at the SAITE 2016, Minho, 2016.
[35] H. Zhou and S. S. Lan, “A Comparative Analysis On Students’ Perceptions And Attitudes Towards Academic Dishonesty Between Students In China And In The United States,” in Proceedings of the Spring 2007 American Society for Engineering Education Illinois-Indiana Section Conference, 2007.
[36] Z. Khan and S. Subramanian, “Students go click, flick and cheat... e-cheating, technologies and more,” J. Acad. Bus. Ethics, vol. 6, pp. 1–26, 2012.
[37] G. G. Ravasco, “Technology-Aided Cheating in Open and Distance e-Learning,” Asian J. Distance Educ., vol. 10, no. 2, pp. 71–77, 2012.
[38] G. G. Ravasco, “Technology-Aided Cheating in ODeL: What Else Do We Need to Know?,” in Creating Spaces and Possibilities, 2012.
[39] J. Sheard, M. Dick, S. Markham, I. Macdonald, and M. Walsh, “Cheating and plagiarism: perceptions and practices of first year IT students,” ACM SIGCSE Bull., vol. 34, no. 3, pp. 183–187, 2002, DOI: 10.1145/544465.544468.
[40] J. Sheard and M. Dick, “Computing student practices of cheating and plagiarism: a decade of change.,” in The 16th Annual SIGCSE Conference on Innovation and Technology in Computer Science Education, ITiCSE 2011, Darmstadt, Germany, June 27-29, 2011, 2011, pp. 233–237, DOI: 10.1145/1999747.1999813.
[41] R. Likert, “A Technique for the Measurement of Attitudes,” Arch. Psychol., no. 140, pp. 1–55, 1932.
[42] M. Spitzer, Digitale Demenz. Droemer Knaur, 2012.
[43] M. Spitzer, The Mind within the Net: Models of Learning, Thinking, and Acting, 1st edition. The MIT Press, 1999.
[44] L. Dossey, “FOMO, digital dementia, and our dangerous experiment,” Explore N. Y. N, vol. 10, no. 2, pp. 69–73, Apr. 2014, DOI: 10.1016/j.explore.2013.12.008.
[45] R. R. Gajewski, “How to change the unchanging? Restructuring Polish universities for the XXI century,” in TelE-Learning, D. Passey and M. Kendall, Eds. Springer US, 2002, pp. 297–300.
[46] K. Illeris, Three Dimensions of Learning: Contemporary Learning Theory in the Tension Field Between the Cognitive, the Emotional and the Social. Malabar, Florida: Krieger Publishing Company, 2004.
[47] G. Cheetham and G. Chivers, “How professionals learn in practice: an investigation of informal learning amongst people working in professions,” J. Eur. Ind. Train., vol. 25, no. 5, pp. 247–292, 2001, DOI: 10.1108/03090590110395870.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 921–925
DOI: 10.15439/2016F75
ACSIS, Vol. 8. ISSN 2300-5963

Semiotic Training for Brain-Computer Interfaces

Mariya Timofeeva
Sobolev Institute of Mathematics SB RAS,
4 Acad. Koptyug Avenue,
Novosibirsk State University,
2 Pirogova Str. 630090 Novosibirsk,
Russian Federation
Email: timof@math.nsc.ru

***Abstract*—With time education becomes more personified. New categories of learners join in educational processes and new areas of education appear. Brain-computer interfaces have good perspective to contribute to these tendencies. This technology may allow disabled people to participate in social life, including education, and may let healthy people to develop the skill of controlling brain waves. Training the skill is the object of investigation and researchers recommend taking into account human factors: general principles of learning, motivations, personal patterns and abilities (among them, spatial). Education may be a source of highly motivated training tasks. The paper treats brain-computer interface as a type of communication and argues for semiotic training that is a variant of training spatial abilities. Semiotic training have proved effectiveness in other areas of education. Theoretical background and preliminary empirical comments of the approach are considered.**

I. Introduction

THE aim of the paper consists in attracting attention to educational perspectives connected with future development of brain-computer interfaces (BCI) and proposing theoretical rationale for semiotic training as a possible way of developing the skill of controlling brain waves. BCI is considered as a sort of communication.

I. Structure of the Paper

Consequent reasoning characterizes the trends in education and introduces the aspects of BCI that are important for the issue under consideration (Section II). Section III presents general information about BCI. Section IV proposes semiotic view on BCI, argues for methodology of field linguistics as the perspective way of learning BCI skill; discusses basic lines of semiotic training and its benefits. Section V proposes the illustrative model by considering semiotic regularities of human pictorial imagination appearing in comics, the other products of human intentional imagination. The model analysis provides data that may be useful for strategies of semiotic training. Section VI resumes the considered issue.

II. Trends in Education

Three trends of contemporary education are especially important for the current consideration: individualization of educational trajectories, widening the scope of educational activity, and involving new categories of learners.

The scope of educational activity have substantially widened when its virtual forms appeared and allowed new categories of learners to participate in e-learning. Besides, education tends to become more individually fitted and polyvariant. Strategies of education come closer to self-education. Subjectively estimated, elaborated and regulated ways of developing knowledge and experience gain their significance for the subjects of education. Not only general solutions concerning perspectives of education in whole are significant, specific educational technologies designed for concrete category of learners are also valuable.

Developing BCI seems to have a strong potential of contributing to these trends. Communication in whole is basic for education, thus training BCI as the specific means of communicating with external world for mastering internal skills can involve certain categories of people (disabled or healthy) into educational processes [1]–[4]. Besides, according to [5], reproducing physical actions (with the help of movement interactive devices) trains memory and motor abilities. This result supports the hypothesis that implementing physical actions by the force of imagination may give the same benefits. The proposed semiotic training procedure may give double benefit by affecting both semiotic and brain control skills. The former is valuable for dealing with semiotic systems (for instance, natural languages), the latter – for mastering BCI.

III. Brain-Computer Interfaces

I. General Characteristics

BCI allows a person to perform physical actions (on the screen or really) by the force of intended imagination, or more precisely, by the force of another side of imagination – brain waves – that are fixed and translated into characteristics of the physical action. The intentions often are not fully arbitrary, and should follow a task that is given to a BCI user. The tasks may differ in complexity and content. Some BCI designs presuppose a free training session. During this session, a BCI user chooses imagery tasks by his / her own decision [6].

Among the current discussions about BCI, two items are especially crucial for prospective education.

I am very grateful to *Sobolev Institute of Mathematics* and *Novosibirsk State University* for supporting my research

First, training procedures for mastering this sort of interaction, particularly mental tasks and feedbacks. Currently used training protocols are discussed, for instance, in [6]. Feedback provides perceivable data for accessing effectiveness of brain control. There are different variants of feedback, for instance, neurofeedback [7], the motor imagery training system proposed in [8]. The present paper discusses only the variant of training BCI skill and does not concern the details of designing feedbacks.

Second, perspectives of BCI that go beyond solely medical purposes and cover other sorts of activities intended for disabled or healthy people [1]–[2].

BCI is substantially interdisciplinary undertaking and, as the authors of [9] notice, cooperation across disciplines has good potential to improve situation. The present paper proposes linguistic view on the situation.

This view rests on considering BCI as a sort of communication subject to semiotic interpretation. The stance suggests the idea of semiotic training based on the methodology of field linguistics. In case of BCI, the semiotic training will develop spatial abilities. This is important because according to the hypothesis [10]–[11] spatial abilities may efficiently contribute to the development of BCI skill.

II. Training Procedures

Efficiency of BCI may depend on ability of a human to control brain activity, and this skill is not inherent by nature, many people (between 15 and 30 % [10]) cannot use it at all. It is crucial for certain types of BCI, particularly for spontaneous BCI [6]. Lack of the capacity to control brain waves is called “BCI illiteracy” or “BCI deficiency” [10].

Thus, thinking over strategies of preliminary training and perfecting the capacity is often marked as actual. Two approaches are noteworthy in the context of the present paper. The first [6] recommends to generalize the key features of efficient training in different areas of knowledge and to infer relatively skill-independent recommendations. The second [10] suggests personifying the process by using individually designed strategies of mastering the skill. Since we consider BCI as a sort of communication, the former approach gives reason for generalizing linguistic experience of learning unknown languages by reconstructing the semiotic structure of a language on the base of the speech data. According to the latter approach individual semiotic training (as a skill of reconstructing semiotic structures from communicative data) seems to be promising.

IV. Semiotic View on Brain-Computer Interface

I. Brain-Computer Interface as a Semiotic System

BCI is interpretable as a sort of translation that transforms information along the following line: intention – (visual or kinesthetic) image of a desired physical action – brain waves – specification of the action (recognized by special equipment and appropriate algorithms) – physical action. A user of BCI cannot monitor directly brain waves; nevertheless, (due to perceivable feedback) he / she may regulate it indirectly, through modifying intentions and therefore images. Thus for a user of BCI the aforementioned line of translation is shorter: intention – image – physical action. The shortened line admits semiotic representation.

The notion of sign varies in different semiotic theories. “Dyadic” [12] and “triadic” [13] models of sign are the most common. The visual form of the latter is “semiotic triangle”. There is no full agreement about the terms marking the interrelated angles of the triangle. We will use the following terms: sign (signifier, a linguistic form) – meaning (mental entity, concept) – reference (physical or abstract entity indicated by the sign in a real act of communication). A language user can directly control the usage of a sign and voluntary modify it if necessary.

In BCI, image plays the role of a signifier. A BCI user can intentionally modify it. Resulting physical action plays the role of the referent Meaning is individual intrinsic skill of building appropriate images, i.e. images that effectively initiate desired actions.

An image may be a simple sign or a compound sign. The components of a compound sign correlate with the features of a referent. Compound signs appear on rather developed level of communicative skill.

In linguistics, natural language signs are usually considered as linear, one-dimensional. In BCI, signs are non-linear. BCI as communication is closer to sign languages, (used by deaf people): sign languages localize in space, not in line.

II. Extractability of Semiotic System

Semiotic systems can vary in their substance broadly. They are not limited in modalities of realization, in number of dimensions, or nature of elementary and complex signs. At the same time, there are general rules governing their internal structure and functioning, and these two aspects are closely connected: a structure becomes apparent during functioning and thus is extractable from the instances of functioning. Validity of the result depends on quantity and quality of communicative data.

Sign languages give the appropriate illustration. These languages are spatial: a “speaker” creates signs by different sorts of gestures, including body movements and mimics. Thus, competence in a sign language requires sufficiently developed spatial abilities.

The history of sign languages is rather instructive. Many years had passed before linguists recognized semiotic nature of sign languages and began to consider them not as a mere pantomime. This happened due to the pioneer work of William Stokoe [14]; in 1960th sign languages became the objects of contemporary linguistics. Possibility of the crucial turn was stipulated by semiotic analysis fulfilled by Stokoe: he extracted constituents of semiotic system (“cheremes”) from continuum of raw communicative data. This process is similar to deciphering a language. Corresponding linguistic methodology is typical for field linguistics.

Viewing the gesture space as a medium obeying semiotic regularities permits to segment gestures into elementary features of signs and to imitate (to a certain extent) communication based on a sign language automatically in the systems of machine translation [15].

Theoretically, an intended image in BCI can also be semiotically complex and consist of several more simple parts

of signs. Presumably, these constituents of the intended image correspond to certain constituents of the desired resulting movement; and a set of such constituents is individual and specific for each person.

Special area of linguistics develops methodology for extracting semiotic systems from raw communicative data. This is "field linguistics" [16].

III. Field Linguistics

Field linguists are learning and studying the languages (usually exotic) that are unwritten, familiar only to their native speakers who do not know any other language besides their own. Moreover, the circumstances of life and phonetic features of the target language are also exotic thus limiting linguists in using analogies.

Field linguist works in specific investigational situation when the researcher has no linguistic competence in the target language, and the native speaker constitutes the only source of information about the language [16]. Similarly, a BCI user fulfills the role of a "linguist", which is analyzing perceptive data in order to extract the underlying semiotic patterns.

Both types of communicators, intentionally and experimentally, by trial and error, are trying to detect (rationally or only perceptually) the proper abstract knowledge. Difference is in the nature of this knowledge: a BCI speaker develops visual or kinesthetic skill, a linguist – articulatory skill that also can be auditory or kinesthetic. In both cases kinesthetic skill is said to be more basic ([6], [16]).

At the very beginning of a scientific research (within so called "zero cycle"), a field linguist knows nothing about semiotic structure of the target language, he / she should discover the signs and meanings of the language on the base of analyzing language use.

Undoubtedly, BCI speaker usually is not a linguist; these two types of people have different interests and goals. Nevertheless, the linguistic character of BCI training is substantially similar to that of a zero cycle in field linguistics. Both are akin to deciphering.

IV. Semiotic Training

If a BCI user wants to develop the skill of controlling brain waves, he / she tries to discover why the image of desirable movement appeared to be insufficiently strong. Actually, the process means that the BCI user tries to change hypothesis about the proper constituents of the images (signs) and to elaborate the more effective ones. Developing BCI skill actually means the process of refining the hypothesis. In fact, a BCI user intentionally trains his / her linguistic competence in BCI communication. The field linguist fulfills the similar task.

The learning procedures for developing semiotic skill should train general subject independent schemas of semiotic analysis; at the same time, the object of the analysis is individual (individual perceptive data). In contrast, a foreign language learner usually learn the concrete semiotic system with previously stated signs.

Semiotic training involves series of iterative semiotic analysis. The latter consists in searching for combinatorial regularities in perceptive data. General regularities are common and sufficiently strict; [17] presents basic variant of the procedures used in field linguistics. The whole scope of investigation in field linguistics embraces all language levels and thus may seem not easy for a BCI user if he / she is not a linguist. Nevertheless, semiotic analysis is flexible and does not require building a multilevel structure (as for natural languages); the analysis may not go beyond one level thus becoming well understandable.

The proposed procedure of semiotic analysis is a sort of mental experiment. A user of BCI, or an "experimenter", may realize it by fulfilling the following types of operations: 1) to fix some part of an image of the desired physical action; 2) to vary the remained part / parts and simultaneously trace the results (check if the resulting physical action is appropriate); 3) to repeat (if necessary) the experiment with different variants of the division; 4) to search for interchangeable parts that may replace each other without changing the context (the sets of such parts hypothetically correspond to abstract entities).

The experimenter may repeat these operations cyclically and then summarize the results of several experiments. The obtained system of signs will be individual and specific for each person.

It is worth noting that each semiotic cycle exercises spatial abilities of the person.

Semiotic training may give double benefit by affecting both brain control skill and semiotic skill.

The latter is substantial for dealing with other semiotic systems, particularly with natural languages, and this idea has the empirical confirmation. The Traditional Linguistics Olympiads successfully use the linguistic variant of supposed semiotic tasks (in the spirit of field linguistics) for searching and training linguistically gifted children. For solving the tasks of the Olympiad, no prior knowledge of linguistics or languages is required: logical ability and the will are sufficient. Information about the Olympiads and the collections of tasks are available at http://www.ioling.org/, http://www.lingling.ru/olymps/mos_olymp/. Several decades (46 years) of fruitful practicing this type of semiotic tasks give empirical support for efficiency of training procedures based on the methods of field linguistics.

V. Semiotic Analysis

I. Parameters of Images

Semiotic analysis briefly depicted in the previous section does not provide deterministic procedure. The analysis has no less than four degrees of freedom: 1) the way of segmentation and granularity of the static perceived data, 2) the way of segmentation and granularity of the dynamic perceived data, 3) the set of simple / elementary actions, and 4) the goals of actions. For instance, a BCI user fulfilling the task of grasping an apple may imagine only a rather undetermined and undivided movement beginning from the initial location of the hand and ending at the location of the apple. In other case, the user may imagine the same process in details including, for instance, the images of tensed muscles, trajectory of hand movements, positions of fingers on the apple, rotation of the apple. The whole task may correspond to one goal or to the

set of interconnected goals, for instance: to bring the hand nearer to an apple, to open the fingers, to touch the apple by the palm, to clasp the fingers around the apple.

Thus, semiotic analysis may be characterized by certain numeric parameters, among them: number of discrete parts detected in the visual continuum; number of modified parts; number of steps implementing the modifications; power of the set of elementary static / dynamic constituents. A person may determine these parameters of imagination by his / her own choice.

Nevertheless, semiotic analysis depend not only on individual decisions; it may reflect some general regularities: limitations, peculiarities and habits of people. Semiotic analysis made by linguists cannot give an appropriate example for learning these regularities because linguists usually have overtrained semiotic skill; a BCI user, on the contrary, may have a minimal experience in semiotic analysis. Besides, linguists as a rule do not analyze dynamic imagination. According to the survey of linguistic approaches to comics [18], the works in this area usually use comics for supporting the already existing linguistic theories elaborated for ordinary human languages. For our aims, we need to understand how the mind transfers meanings from imaginary to visual modality.

It would be useful to have some preliminary considerations about the parameters. These considerations may help to specify initial point for launching a semiotic cycle and to propose some guidelines for its developing.

It is not easy to undertake wide-ranging semiotic investigation of BCI data. Contemporary BCI is still barely used outside laboratories [10]–[11] and thus does not allow to observe diversified and complex physical movements implemented by brain force. Therefore, it is productive to make preliminary observations concerning the abovementioned substantial features of human intentional imagination on the base of other data that are more easily acceptable and do not need sophisticated or expensive equipment. We suppose to consider pictorial imagination for illustrative model analysis.

II. Model Analysis

The main question for consideration: How do a person split imaginary actions into parts when he / she reproduces the copies of the actions in the physical world?

The purposes of model analysis are: 1) to refine the list of parameters substantial for visualizing imaginary actions; 2) to provide data useful for operating with the parameters. Primarily we should choose a type of pictorial imagination appropriate for the purposes of model analysis.

As [19] shows, the way of performing a spatial task depends on the medium where it is performed. If one performs the task in virtual world, on the screen, the operations should be precise, discrete, well planned, governed by the goal in mind. The details of performing operations should be rather explicit than implicit. On the contrary, in the real world, one performs the same task intuitively, operations are continuous, and do not require high degree of specification. The image of a future action often appears in consciousness only vaguely. The authors of [19] suggest that spatial thinking skills required for creating representations in virtual and real worlds may be different. Actions performed due to BCI, like actions in virtual world, require higher degree of refinement.

In this respect, painting gives a similar example because for depicting an imagined scene, a painter should refine the image (initially viewed as integral and continuous) and may divide it into discernible well-defined parts. A picture alone is static, yet a sequence of pictures may depict dynamic stages of an action. This is typical for comics, thus comics give appropriate illustrative model for considering regularities in visual representations of imagined actions.

A person implementing a complex action by the force of imagination divides the action into parts. The painter of a comics also transforms a continuous action into discrete parts, i.e. determines, among other things, the static and dynamic components of the image, the elementary movements, the goals and sub-goals of actions; the abstract characteristics like speed, chronological order, importance, intensity. In other words, the painter makes decisions similar to these that are crucial for implementation of a physical action within BCI.

Comics may vary broadly in their genres. For our purposes, we need an example with a realistic plot, which depicts coherent realizable situations and processes. Logicomix [20] is the suitable variant for our model analysis.

The analysis showed that the way of splitting actions into parts depends rather on distribution of attention than on type of action. Attention may focus on the stages of an action and / or on its characteristics (particularly, abstract). For instance, intention to attract attention to quickness of an action often results in dividing the action into stages, i.e. fixing the successive phases of the action. This mode of emphasizing the speed / intensity may be used with reference to physical or to internal actions (for instance, mental).

Thus, the main recommendation for developing semiotic abilities may consist in training the skill of distributing degrees of attention between actions, their stages and characteristics. The skill presupposes elaborating the subjective patterns of simple actions (stages of actions) typical for essential behavior and training these patterns separately, as the isolated signs; stages are especially actual for intensive / high-speed actions. It may be useful to ask a person who wants to develop semiotic skill to make a drawing of the stages of a desired action. However, for disabled people this variant may be impossible. Concrete material for preliminary designing patterns and splitting actions is extractable from the typology of actions.

The analyzed material induces the ternary typology of actions that include active states, detailed actions, and undetailed actions.

Active states include: a) monotonic repeated movements, b) random movements specific for a situation under consideration, c) panoramic view of co-located motions. One active state usually corresponds to one panel.

For instance, several elementary movements on one panel may correspond to a continuous monotonous process like walking round and round a flowerbed during a course of cogitation. The panel in this case shows several positions of the person located on the garden path around the flowerbed. Monotonous repeated processes like dangling a foot give

another example of elementary movements representable on one panel; in this case, the panel shows several positions of the foot simultaneously. Both cases give evidences for relatively small importance of elementary movements; the movements from one panel have common purpose.

The detailed action has refinements that may indicate stages, chronological order, highlighted constituent parts, speed, and degree of importance or intensity. Indicators of these characteristics include assortment of colors (the past is less colored), focus distance (may reflect degree of importance, intensity, speed); rotations (may serve as a means of highlighting the actual constituents). Little difference between successive panels may indicate intensity of an action (predominantly for mental actions). Visual indicators of changing attention may include; moving away from a depicted scene or closer to it (making an image, not frame, smaller / larger); modifying the angle of the field of vision; modifying the environment, elimination of the background (thus emphasizing the focus of attention).

In addition to the parameters listed above the annotation of the panels includes information about the hierarchy and the types of physical actions.

The whole slot of the comics breaks up to sub-slots (scenes); a sub-slot depicts a sequence of collocated actions. In Logicomix, the volume a sub-slot may reach several dozens of panels.

Realization of a complex action may include several simple actions (the stages). A simple action may include several elementary movements; a panel presents all of them simultaneously. Thus, we have the following hierarchy: slot – sub-slots – complex actions – simple actions – elementary movements. The hierarchy of actions corresponds to the hierarchy of goals.

The database provides characteristics actual for semiotic training (calibration, launching and developing semiotic analysis); proposes the ways of visualizing complex actions and their characteristics (particularly, abstract) for designing BCI and other types of visualized human-computer interfaces.

VI. Conclusion

BCI and the skill of controlling brain waves are prospective areas of future education. This position, expressed, for instance, in [3] and [4], is the starting point of the paper.

BCI is a sort of communication and allows semiotic analysis. The proposed semiotic training procedure bases on the linguistic experience elaborated for dealing with unknown semiotic systems. Semiotic training (in its linguistic version) has already proved its effectiveness in linguistic education.

The realized model analysis considers pictorial intentional imagination as a source of preliminary data useful for developing semiotic analysis.

The proposed variant of semiotic training is consistent with general recommendations that the specialists in BCI suggest: it takes into account human factors, individual capacities, and trains spatial abilities of a person.

Semiotic training may have different implementations; its potential seems to be promising and going beyond the area of BCI or training the skill of controlling brain waves.

References

[1] Jan B.F. Van Erp, F. Lotte, M. Tangermann, “Brain-computer interfaces: beyond medical applications”, *Computer -IEEE Computer Society-*, IEEE, 2012, 45 (4), pp.26-34. http://doi.ieeecomputersociety.org/10.1109/MC.2012.107.

[2] C. Ring, A. Cooke, M. Kaussanu, D. Mclntyre, R. Masters, “Investigating the efficiency of neurofeedback training for expediting expertise and excellence in sport”, in *Psychology of sport and exercise*, vol. 16, part 1, January 2015, pp. 118-127. DOI:10.1016/j.psychsport.2014.08.005.

[3] B. Sabitzer, “Neurodidactics – a new stimulus in ICT and computer science education”, *INTED2011 Proceedings*, 2011, pp. 5881-5889.

[4] M. Ferrari and H. McBride, “Mind, Brain, and Education: The Birth of a New Science”, in *LEARNing Landscapes*, vol. 5, No. 1, autumn 2011, pp. 85-100.

[5] J. E. Garrido, V. M. R. Penichet, M. D. Lozano and L. A. Sánchez, "Mobility and memory training through movement interaction," *Computer Science and Information Systems (FedCSIS), 2012 Federated Conference on*, Wroclaw, 2012, pp. 883-889.

[6] F. Lotte, F. Larrue, Ch. Mühl, “Flaws in current human training protocols for spontaneous brain-computer interfaces: lessons learned from instructional design”, in *Frontiers in Human Neuroscience*, 2013, vol. 7, N 568 (11 p.). DOI: 10.3389/fnhum.2013.00568.

[7] C. Neuper and G. Pfurtscheller, “Neurofeedback training for BCI control”, in *Brain-computer interfaces*, Eds. B. Graimann, G. Pfurtscheller and B. Allison, London: Springer, 2010, pp. 65–78. DOI: 10.1007/978-3-642-02091-9_4.

[8] H.-J. Hwang, K. Kwon, and C.-H. Im, “Neurofeedback-based motor imagery training for brain-computer interface (BCI)”, in *Journal of neuroscience methods*, vol. 179, issue 1, April 2009. Pp. 150-156. DOI: 10.1016/j.jneumeth.2009.01.015.

[9] *Brain-computer interfaces. Revolutionizing human-computer interaction*. The Frontiers Collection. Eds. B. Graimann, G. Pfurtscheller and B. Allison, London: Springer, 2010, 393 p. DOI: 10.1007/978-3-642-02091-9.

[10] C. Jeunet, B. N’Kaoua, S. Subramanian, M. Hacher, and F. Lotte, “Predicting mental imagery-based BCI performance from personality, cognitive profile and neurophysiological patterns”, *PLOS ONE*, vol. 10 (12): Dec. 2015. DOI: http://dx.doi.org/10.1371/journal.pone.0143962.

[11] F. Lotte, C. Jeunet, “Towards improved BCI based on human learning principles”, *3rd International Winter Conference on Brain-Computer Interfaces*, High1 Resort, South Korea, Jan. 2015, <hal-01111843>. DOI: 10.1109/IWW-BCI.2015.7073024.

[12] F. de Saussure, *Course in general linguistics*, trans. Roy Harris, [1916] London: Duckworth, 1983.

[13] C. S. Peirce, *Collected writings* (8 Vols.), Ed. Charles Hartshorne, Paul Weiss and Arthur W. Burks, Cambridge, MA: Harvard University Press, 1931-58.

[14] W. C. Stokoe, “Sign Language structure: an outline of the visual communication systems of the American deaf”, in *J. of Deaf Studies and Deaf Education*. 2005; 10(1), pp. 3-37. DOI: 10.1093/deafed/eni001.

[15] M. Huenerfauth, “A survey and critique of American Sign Language natural language generation and machine translation systems”, Technical Report MS-CIS-03-32, Computer and Information Science, University of Pennsylvania, 2003, 36 p.

[16] A. Kibrik, *The methodology of field investigations in linguistics (setting up the problem)*, Janua Linguarum. Series Minor, 142, The Hague/Paris: Mouton, 1977, 130 p.

[17] H. Gleason, *An introduction to descriptive linguistics*, rev. ed., New York: Holt. Rinehart & Winston, 1961, 503 p.

[18] N. Cohn, “Comics, linguistics, and visual language: the past and the future of a field”, in F. Bramlett (Ed.), *Linguistics and the study of comics*, New York: Palgrave MacMillan, 2012, pp. 92-118. DOI: 10.1057/9781137004109_5.

[19] H.Erhan, B. Yousuf, and B. Berry, ‘Teaching Spatial Thinking in Design Computation Contexts: Challenges and Opportunities”, in N. Gu, & X. Wang (Eds.) *Computational Design Methods and Technologies: Applications in CAD, CAM and CAE Education*, 2012, pp. 365-389 (chapter 21). Hershey, PA: Information Science Reference. DOI: 10.4018/978-1-61350-180-1.ch021.

[20] A. Doxiadis, C. H. Papadimitriou, *Logicomix: An Epic Search for Truth*, London: Bloomsbury Publishing PLC, 2009, 352 p.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 927–931
DOI: 10.15439/2016F159
ACSIS, Vol. 8. ISSN 2300-5963

The synthesis of a unified pedagogy for the design and evaluation of e-learning software for high-school computing

Peter Yiatrou
University of Central Lancashire. Fylde Road, Preston, Lancashire PR1 2HE, United Kingdom
Email: pyiatrou@uclan.ac.uk

Irene Polycarpou
University of Central Lancashire Cyprus. 12 - 14 University Avenue Pyla 7080 Larnaka, Cyprus
Email: ipolycarpou@uclan.ac.uk

Janet C Read
University of Central Lancashire. Fylde Road, Preston, Lancashire PR1 2HE, United Kingdom
Email: jcread@uclan.ac.uk

Maria Zeniou
University of Central Lancashire Cyprus. 12 - 14 University Avenue Pyla 7080 Larnaka, Cyprus
Email: mzeniou1@uclan.ac.uk

***Abstract*— This study develops a unified pedagogy for the design and evaluation of e-learning software for high-school Computer Science. In accordance with the pedagogy, prototype e-learning software was developed for use in student instruction and independent learning. The pedagogy was iteratively refined based on the evaluation of teachers and education experts and the resulting e-learning software was developed considering student feedback. The problem domain focuses on the UK's recent shift in educational emphasis towards Computer Science GCSEs; however, the findings are broadly transferable to other developed nations. The pedagogy synthesizes the following learning theories: Constructivism, Social Constructivism, Connectivism, Cognitive Load, ARCS and VARK learning styles, these were in turn distilled into 31 heuristics. The research is broken into three phases, the first two phases are discussed in this paper; Phase 1 is the Initial Pedagogical Strategy and Prototype, Phase 2 is the Elaboration via Action Research.**

I. Introduction

THERE is a well-publicized body of inquiry, consisting of various reports, analysis and political rhetoric that assert that computing education in the UK has been struggling [9], [11], [22]. This concern led to the programme of study for Information and Communication Technology (ICT) being temporarily dis-applied while new initiatives introduced an arguably more academically rigorous Computer Science GCSE [8].

The prevalence and ubiquitous nature of ICT in developed countries and its impact on recent generations is well documented [13]. Digital technology is shown to be a fundamental part of the fabric of society in developed countries such as the UK [9], [15], [18]. These findings arguably make the need for computing education all the more important, and in support of computing education it is postulated that e-learning software can offer learning benefits in the form of a media rich interactive environment that is engaging and can promote active learning [7].

The objective of the research presented in this paper is to study and synthesize leading learning theories into a single unified e-learning pedagogy that will support high school computing, and in particular, the new Computer Science GCSEs. This pedagogy will be embodied in an e-learning software prototype and both will be evaluated to identify their impact on student learning and engagement.

This work was not supported by any organization

II. Quality E-Learning Software

Although e-learning software has become mainstream, one of the main concerns still remains that what is delivered often falls short [1], [6]. Content quality, pedagogical usability, instructional design and a lack of alignment with education needs and standards remain a concern in existing e-learning software [6], [12], [20].

This research aims to support increased use of e-learning in high-school Computer Science, and to safeguard the pedagogical quality of the e-learning software, by defining a comprehensive set of pedagogical heuristics to guide teachers in designing and/or evaluating e-learning software for use in teaching. Therefore, in the context of this research, quality focuses on the standard and degree of excellence of the pedagogy underpinning the learning process.

III. Learning Theories

One approach to ensure the pedagogical quality of e-learning software is to ground it in established learning theories. There is a significant body of research into learning theories, e-learning and Science, Technology, Engineering and Mathematics (STEM) education, but this is somewhat overwhelming; there are complementary and competing learning theories and varied perspectives on how to best implement these theories in technology [14], [25].

Illeris [14] proposes that since learning is so complicated, any "*analyses, programmes and discussions of learning must consider the whole field if they are to be adequate and reliable*"(p.18). It is for this reason that this research synthesizes this overwhelming body of knowledge into an accessible set of pedagogical heuristics. The learning theories considered include Constructivism [4], Social Constructivism [19], [24], Connectivism [3], Cognitive Load [7], [23], ARCS [16], [17] and VARK learning styles classification [10]. These theories were selected primarily

due to their maturity and the availability of research that discusses them, their effective implementation, and the evidence of their positive affect on learning and motivation. A final consideration was to include theories that relate to technology and our current digital society.

IV. Research Methodology

The research study that is presented in part in this paper is divided into three phases (Refer to Fig. 1) and each phase uses a mixed methods approach, but with a different qualitative-quantitative mix. Overall, the phased approach utilizes an exploratory mixed methods design in which Phase 1 and 2 use primarily qualitative data from students, teachers and education experts to synthesize a set of pedagogical heuristics. The rationale for this research design is to initially work in depth using significant literature review and iterative Action Research cycles to gradually refine the heuristics and e-learning test tool.

Phase 3, although not discussed in this paper, will have a quantitative priority and a larger student sample in order to generalize the findings to the wider student population. The aim will be to measure whether there are improved assessment results using e-learning software that adheres to the pedagogical heuristics synthesized in this research.

A. Phase 1: Initial Pedagogical Strategy and Prototype

The primary objective of Phase 1 was to set a strong foundation for the research study; this was achieved in terms of:

1. Piloting the research methods and protocol,
2. Developing the first draft of the e-learning pedagogical heuristics,
3. Developing a working e-learning software prototype for the topics of Algorithms and Computational Thinking that was in turn used for evaluation purposes.

This supported the early identification of shortcomings in the research methods and protocol, and in the design and development processes; but most importantly, it allowed early feedback on the pedagogy and e-learning software.

The first step was to undertake a comprehensive literature review resulting in the first draft of the e-learning pedagogical heuristics for GCSE Computing.

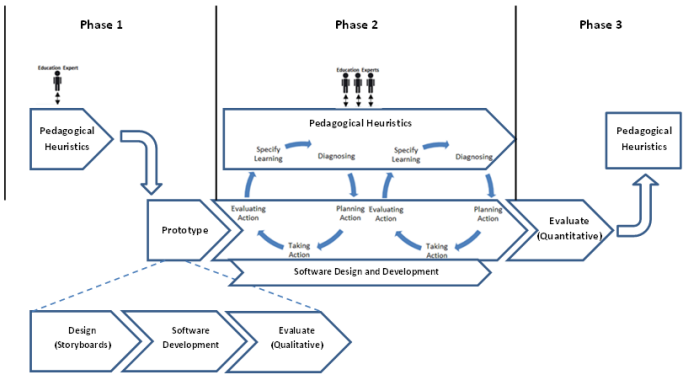


Fig. 1 The three phases of research.

An experienced GCSE teacher and five education experts then evaluated the pedagogy and provided constructive feedback and validation of the appropriateness of the heuristics for Computer Science for the GCSE age group. This feedback was analyzed and incorporated into the research prior to Phase 2-Cycle 1.

As a proof of concept, a prototype e-learning software was developed according to the e-learning pedagogical heuristics. Eight GCSE ICT students were recruited from a local high school and observed using the e-learning software. Prior to the observation study, an online VARK questionnaire was administered to participants to collect their learning style preferences. Additionally, after the observation study an online survey was administered to collect participants' feedback on their experience of using the e-learning software. This survey also included John Keller's Instructional Materials Motivation Survey (IMMS), which measures student motivation according to the ARCS motivation model [16]. Subsequently, a focus group was held in which the participants had a facilitated discussion where they elaborated on the key themes identified in the observation study and the survey results.

The findings from the observation study, VARK questionnaire, student survey and focus group were examined holistically to understand whether the different research strands converged or diverged, to identify areas requiring further investigation, and to inform the development of the pedagogical heuristics and the e-learning software for Phase 2-Cycle 1. The findings are predominantly qualitative in nature; have a small sample size and are a collection of open and ordinal data (Likert and Ranking). Descriptive statistics gave basic quantification of results and are followed by textual descriptions that link to open responses from either the questionnaire or focus group and interpret the result findings.

B. Phase 2: Elaboration via Action Research

The purpose of Phase 2 was to further refine and elaborate the e-learning pedagogical heuristics and software via an Action Research methodology. An Action Research approach was chosen since it links theory and practice, achieving both practical and research objectives. The practical focus lies in the iterative development of the e-learning software and the research focus on the elaboration, evaluation and validation of the e-learning pedagogical heuristics. Susman and Evered [21] detail a five phase cyclical process, which forms the basis of the action research cycle used in this research. The five phases are diagnosing, action planning, action taking, evaluating and specify learning; these are outlined in Fig. 2.

Two cycles of evaluation and update were included in Phase 2. Phase 2-Cycle 1 study participants were six year 5 high school students (2nd year of the GCSE) and three year 4 high school students (1st year of the GCSE). The year 5 students also participated in Phase 1 and have worked with

the previous version of the e-learning software. With each cycle of Action Research, the e-learning pedagogical heuristics were updated based on the evaluation of teachers, education experts and ongoing literature review. Then, aspects of the pedagogy were represented in the e-learning software for evaluation purposes.

As with Phase 1, the student feedback on the e-learning software was collected via a combination of direct observation of software usage, online questionnaire and associated focus groups. Again, aligned with Phase 1, these findings were represented with descriptive statistics and contextualized with text description and links to open responses. These findings are examined holistically and merged with the feedback from education experts in order to inform the next cycle of action research.

V. Development of Pedagogical Heuristics

The pedagogical heuristics originally developed in Phase 1 have been iteratively refined and evaluated by two teachers and five education experts in Phase 1 and Phase 2-Cycle 1. Both teachers teach GCSE Computer Science or ICT; one teacher had five years of experience at the time of the study, the other less than one year of experience. The education experts were university academics in the fields of Child Computing Interaction, Computer Science, Education and Educational Media.

In Phase 1-Cycle 1 there were 31 e-learning heuristics defined. The evaluation findings in Phase 2-Cycle 1 indicate that the heuristics have a comprehensive pedagogical coverage, are appropriate for Computer Science, and are overall appropriate for the GCSE age group (15 and 16 year olds).Whilst the heuristics themselves have received positive feedback, the pedagogy document requires further work in Phase 2-Cycle 2 to reduce the document size, rationalize the heuristics and make the pedagogy more appropriate for its intended usage and audience.

VI. Results

Research findings from Phase 1 and Phase 2-Cycle 1 have informed the e-learning pedagogical heuristics and the e-learning prototype; however, this section discusses only the most significant findings, such as multimodal learning, active learning, authenticity vs. cognitive load, moderation in the heuristics, collaborative learning and learner motivation. Phase 2-Cycle 1 gave some initial encouraging findings since all nine participants agreed or strongly agreed that the e-learning software was easy to use. All participants either agreed or strongly agreed that the learning content in the e-learning software was represented in a clear and understandable way. Furthermore, none of the eight respondents reported that they supplemented, or needed to supplement, the learning material in the e-learning software with further textbook reading. This is self-reported feedback and does not equate with learning, but it does give a positive indicator of the value of the pedagogical heuristics.

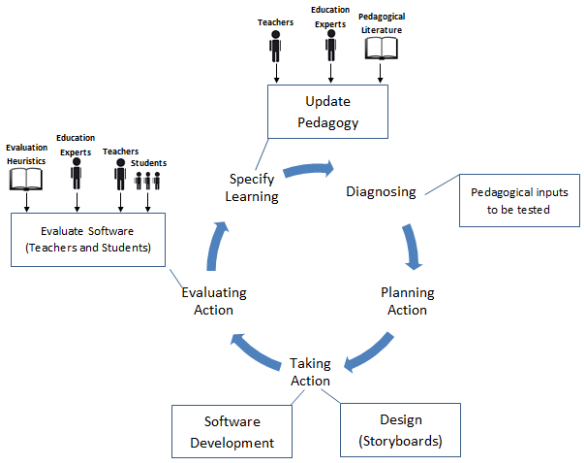


Fig. 2 Action Research Cycle

A. Multimodal Learning

Multi-modal learning was implemented in the software by representing educational material through combinations of text, audio, video, diagrams, pictures, animations and activities. The findings from Phase 1 and Phase 2-Cycle 1 support the multi-modal strategy outlined in the pedagogy. The combined VARK modal preferences of the student participants in both phases show a relative balance across the four modalities. This is further supported by the feedback in Phase 2, where four students agreed and four students strongly agreed with the statement "*The approach of using varying methods to represent the same educational concepts helped my understanding.*" Speaking on this point one student summarized that: "*I really liked the short videos, I think the pictures were actually really good, the text was also decent…"* [Student 19]

The multimodal approach is also given additional credibility since in Phase 1 the instructional designer for the e-learning software had a dominant Read/Write modal preference, which was left unmanaged and resulted in e-learning software with a heavy text bias. This led to a clear negative reaction in the Phase 1 student feedback. In Phase 2-Cycle 1, knowing this bias, the instructional designer took appropriate mitigation steps. This led to an improved response from the student participants in Phase 2-Cycle 1.

B. Active Learning

In support of the constructivist principle of active learning, the pedagogy proposes a significant focus on activities, problem solving and kinesthetic learning. This form of learning was well received by students; all nine students agreed or strongly agreed that the activity-based parts (problem solving, games, simulations, assessments and quizzes) of the e-learning software helped them to understand the subject matter. In addition, all nine students agreed or strongly agreed that the activity-based parts of the e-learning software were engaging.

C. Authentic learning vs reducing cognitive load

In schools, there is a tension between preparing students for assessments and the responsibility to develop well-rounded individuals who can think beyond exams. This pedagogical tension is also reflected in the e-learning pedagogy. For example, in order to reduce cognitive load, all learning material that does not directly contribute to learning outcomes should be reduced or removed; but this in turn, can weaken the authentic learning also proposed in the pedagogy. The tension between authenticity and cognitive load was discussed in both focus groups; the majority of students felt that authentic learning material that may not be directly examinable should not be sacrificed in favor of reducing cognitive load. Speaking on this point one student summarized that: "*It might not be in the exam but it really broadens your perspective, especially in giving you examples of real life applications. "* [Student 23]

The important factor was that each student needed to use their judgment to decide whether they would focus on such supplementary learning content. In the focus groups, it was agreed that in order to support this decision making process, such learning material would be clearly marked with a visual marker to reflect that it is supplementary information.

D. Moderation and balance in Heuristics

Taken at their extremes, learning theories such as Constructivism and Connectivism postulate some radical positions; these include learners being free to identify their own learning needs as they materialize, being free to support those learning needs by choosing their own learning (nodes) from the learning network and being free to construct their own understanding. In contrast, the preliminary findings of this research support a moderate balanced set of heuristics. The findings from Phase 1 and Phase 2-Cycle 1 show that the students value a degree of freedom, especially in learning activities, but that freedom needs to be bounded within a structured environment.

In Phase 2 Cycle 1, five students agreed and four students strongly agreed with the statement "*It is easy to use the navigation and program controls of the e-learning software.*" This strengthened the Phase 1 results and gave positive support for heuristics related to restricted navigational control. During both focus groups, the students clearly voiced support for the restricted navigation approach.

Furthermore, the results of the survey instrument suggest that the students are fully aware of the value of using the Web to support their subject learning, but the majority still prefer the guidance of the e-learning software to support their learning. This aligns with the moderate Connectivist approach suggested in the pedagogy in which the e-learning software becomes the hub that suggests (links to) other learning resources that are known to be reliable. As in Phase 1, the e-learning software was perceived to be comprehensive and the students preferred the structure of having one place to learn from, avoiding the wasted time in searching the Web and evaluating whether the information they find is correct. Almost all student feedback was aligned with the following comment: "*The E-Learning software has activities and features that are much more engaging than looking up information on the Web. The Web sometimes contains sites with wrong information, so the E-Learning software would be a much more reliable and easy source of information.*" [Student 14]

The Phase 2-Cycle 1 focus groups also revealed some unexpected feedback. Although, the e-learning software had an implicit structure and grouping of learning content, a significant portion of the students expressed the opinion that they would like explicit sections to be introduced. These sections were suggested to mimic a traditional chapter format, with the chapter learning objectives, learning material, review questions and finally a learning summary.

E. Collaboration

The research study has shown conflicting findings in relation to collaborative learning. Based on a student ranking of ten learning object types collaborative learning is ranked lowly in eighth place. Furthermore, only three from eight students agreed or strongly agreed that collaborative activities helped them understand the subject matter. However, a follow up question in the survey instrument and further discussion in the focus groups offered a different context. The students were asked to describe briefly, whether the collaborative activities gave them any learning or motivational benefits and what those benefits were. "*The collaborative activities gave me and my partner the opportunity to help each other understand the questions we had. One could answer the questions of the other, which was really helpful in order to complete our task.*"[Student 14]

The positive responses where reiterated and elaborated during the focus groups; however, one critical factor was also expressed. The respondents clarified that technology enhanced collaborative activities are artificial in a classroom context and much more suited for homework. In the context of this study, the majority of time spent using the e-learning software and the collaborative learning environment (CLE) was in class hence did not feel natural to the students. What was tentatively postulated in Phase 1 was more clearly established in Phase 2-Cycle 1; the experiment design had influenced the research findings. To naturally reflect a collaborative learning context requires a longer duration in which the students have weeks to review, respond and interact. However, the abiding conclusion from Phase 2-Cycle 1 was that despite the mixed feedback, overall, the students saw good potential for collaborative learning in spite of an initial learning curve.

F. Motivation

One of the key objectives of the e-learning pedagogy is to improve learning motivation; a number of heuristics specifically focus on student motivation and others offer supplementary motivational benefits. Phase 2-Cycle 1

findings are therefore broadly positive that six from nine respondents reported that the e-learning software increased their overall enthusiasm and interest in computing.

In relation to the ARCS motivation model, which focuses on Attention, Relevance, Confidence and Satisfaction, the results from the IMMS survey gave further positive indicators for student motivation. The results range between 3.52 and 3.76 on a continuum between Moderately True (3) to Mostly True (4), the maximum on this scale being Very True (5).

VII. Conclusion and Further Work

In this paper we have presented early research that suggests that e-learning can offer educational benefit to high school Computer Science. This educational benefit is realized and maximized by ensuring that the e-learning is underpinned by an appropriate set of e-learning pedagogical heuristics. The qualitative research described here offers support for a number of the heuristics developed in the pedagogy and offer further direction on areas to further refine in Phase 2-Cycle 2.

Although the focus of this research is primarily the UK there are various international comparisons [2], [5], [22] that show that the concerns and challenges outlined in the UK are common to a number of countries. The US, New Zealand, Israel, Germany and India are all in varying stage of similar initiatives to give greater prominence to the high-school computing curriculum. It follows that the findings of this research will be broadly transferable to such initiative in other nations.

The progress so far in Phase 1 and Phase 2 of the presented research, give a sound indication that the developed heuristics positively affect the pedagogical quality of e-learning software. The planned final research phase (Phase 3) of this study is to establish whether the pedagogy influences learning performance and motivation as theorized, it will further validate the findings of the first two phases by confirming them using quantitative methods and attempting to generalize them to a wider population.

References

[1] Alonso, F., López, G., Manrique, D., & Vines, J. (2005). An instructional model for web-based e-learning education with a blended learning process approach. British Journal of Educational Technology, 36(2), 217–235. http://dx.doi.org/10.1111/j.1467-8535.2005.00454.x

[2] Bell, T., Andreae, P. and Lambert, L., 2010, January. Computer science in New Zealand high schools. In *Proceedings of the Twelfth Australasian Conference on Computing Education-Volume 103* (pp. 15-22). Australian Computer Society, Inc..

[3] Bessenyei, I. (2008). Learning and teaching in the information society. Elearning 2.0 and connectivism. Information Society, R.Pinter (Ed), Ed.Gondolat, 2008(9), 1–14.

[4] Brooks, J. G., & Brooks, M. G. (1999). In Search of Understanding: The Case for Constructivist Classrooms. Alexandria, VA, USA: Association for Supervision & Curriculum Development (ASCD).

[5] CAS. (2011). Computing at School International Comparisons. Retrieved May 27, 2015, from http://www.computingatschool.org.uk/data/uploads/internationalcomparisons-v5.pdf

[6] Chan, C. H., & Robbins, L. I. (2006). E-learning systems: Promises and pitfalls. Academic Psychiatry, 30(6), 491–497. doi:10.1176/appi.ap.30.6.491

[7] Clark, R. C., & Mayer, R. E. (2011). e-Learning and the Science of Instruction: Proven Guidelines for Consumers and Designers of Multimedia Learning, 3rd Edition: John Wiley & Sons.

[8] Computing at Schools Working. (2012). Computer Science: A curriculum for schools. Retrieved May 27, 2015, from http://www.computingatschool.org.uk/data/uploads/ComputingCurric.pdf

[9] e-skills UK. (2012). Technology Insight 2012. Retrieved May 27, 2015, from http://www.e-skills.com/research/research-publications/insights-reports-and-videos/technology-insights-2012/

[10] Fleming, N. D., & Mills, C. (1992). Not another inventory, rather a catalyst for reflection. To Improve the Academy, 11, 137 – 155.

[11] Gove, M. (2012). Michael Gove gives a speech at the BETT Show 2012 on ICT in the National Curriculum. Retrieved May 27, 2015, from https://www.gov.uk/government/speeches/michael-gove-speech-at-the-bett-show-2012

[12] Hadjerrouit, S. (2010). A conceptual framework for using and evaluating web-based learning resources in school education. Journal of Information Technology Education: Research, 9(1), 53–79.

[13] Halse, M. . M., Mallinson, B., & Mallison, B. J. (2009). Investigating popular Internet applications as supporting e-learning technologies for teaching and learning with Generation Y. International Journal of Education and Development Using ICT, 5(5), 58–71.

[14] Illeris, K. (Ed.). (2009). Contemporary Theories of Learning: Learning Theorists In Their Own Words. New York: Routledge - Taylor & Francis Group. http://dx.doi.org/10.4324/9780203870426

[15] ITU, I. T. U. (2013). Measuring the Information Society. Retrieved from http://www.itu.int/en/ITU-D/Statistics/Pages/publications/mis2013.aspx

[16] Keller, J. (2006). Development of two measures of learner motivation. Unpublished Manuscript in Progress. Florida State ….

[17] Keller, J. M. (1987). Development and use of the ARCS model of instructional design. Journal of Instructional Development, 10(3), 2–10. doi:10.1007/BF02905780

[18] Office of National Statistics. (2013). Internet Access - Households and Individuals, 2013. Retrieved May 27, 2015, from http://www.ons.gov.uk/ons/rel/rdit2/internet-access---households-and-individuals/2013/index.html

[19] Pritchard, A. (2009). Ways of learning learning theories and learning styles in the classroom (2nd ed.). Abingdon, Oxon; New York, NY: Routledge. http://dx.doi.org/10.4324/9780203887240

[20] Sun, P.-C., Tsai, R. J., Finger, G., Chen, Y.-Y., & Yeh, D. (2008). What drives a successful e-Learning? An empirical investigation of the critical factors influencing learner satisfaction. Computers & Education, 50(4), 1183–1202. doi:10.1016/j.compedu.2006.11.007

[21] Susman, G., & Evered, R. (1978). An assessment of the scientific merits of action research. Administrative Science Quarterly, 23. http://dx.doi.org/10.2307/2392581

[22] The Royal Society. (2012). Shut down or restart? The way forward for computing in UK schools. Technology. Retrieved May 27, 2015, from https://royalsociety.org/education/policy/computing-in-schools/report/

[23] Vogel-Walcutt, J. J., Gebrim, J. B., Bowers, C., Carper, T. M., & Nicholson, D. (2011). Cognitive load theory vs. constructivist approaches: which best leads to efficient, deep learning? Journal of Computer Assisted Learning, 27(2), 133–145. doi:10.1111/j.1365-2729.2010.00381.x

[24] Vygotsky, L., 1978. *Interaction between learning and development*, Available at: https://scholar.google.com/scholar?q=Constructivism+and+Learning.+cobb+1994&btnG=&hl=en&as_sdt=0%2C5#1 [Accessed February 3, 2015].

[25] Reigeluth, C.M. ed., 2013. Instructional Design Theories and Models: An Overview of Their Current Status, Routledge. http://dx.doi.org/10.4324/9780203872130

3rd Doctoral Symposium on Recent Advances in Information Technology

THE third international Doctoral Symposium on Recent Advances in Information Technology (DS-RAIT 2016) will be held as a satellite event of the Federated Conference on Computer Science and Information Systems (FedCSIS 2016) and Education, Curricula & Research Methods (ECRM 2016) conference.

The aim of this meeting is to provide a platform for exchange of ideas between early-stage researchers, in Computer Science, PhD students in particular. Furthermore, the symposium will provide all participants an opportunity to get feedback on their studies from experienced members of the IT research community invited to chair all DS-RAIT thematic sessions. Therefore, submission of research proposals with limited preliminary results is strongly encouraged.

Besides receiving specific advice for their contributions all participants will be invited to attend plenary lectures on conducting high-quality research studies, excellence in scientific writing and issues related to intellectual property in IT research. Authors of the two most outstanding submissions will have a possibility to present their papers in a form of short plenary lecture.

TOPICS

- Automatic Control and Robotics
- Bioinformatics
- Cloud, GPU and Parallel Computing
- Cognitive Science
- Computer Networks
- Computational Intelligence
- Cryptography
- Data Mining and Data Visualization
- Database Management Systems
- Expert Systems
- Image Processing and Computer Animation
- Information Theory
- Machine Learning
- Natural Language Processing
- Numerical Analysis
- Operating Systems
- Pattern Recognition
- Scientific Computing
- Software Engineering

EVENT CHAIRS

- **Kowalski, Piotr Andrzej,** Systems Research Institute, Polish Academy of Sciences; AGH University of Science and Technology, Poland
- **Lukasik, Szymon,** Systems Research Institute, Polish Academy of Sciences, AGH University of Science and Technology, Poland

PROGRAM COMMITTEE

- **Arabas, Jaroslaw,** Warsaw University of Technology, Poland
- **Atanassov, Krassimir T.,** Bulgarian Academy of Sciences, Bulgaria
- **Balazs, Krisztian,** Budapest University of Technology and Economics, Hungary
- **Bronselaer, Antoon,** Department of Telecommunications and Information at Ghent University, Belgium
- **Castrillon-Santana, Modesto,** University of Las Palmas de Gran Canaria, Spain
- **Charytanowicz, Malgorzata,** Catholic University of Lublin, Poland
- **Corpetti, Thomas,** University of Rennes, France
- **Courty, Nicolas,** University of Bretagne Sud, France
- **De Tré, Guy,** Faculty of Engineering and Architecture at Ghent University, Belgium
- **Fonseca, José Manuel**
- **Fournier-Viger, Philippe,** University of Moncton, Canada
- **Gil, David,** University of Alicante, Spain
- **Herrera Viedma, Enrique,** University of Granada, Spain
- **Hu, Bao-Gang,** Institute of Automation, Chinese Academy of Sciences, China
- **Koczy, Laszlo,** Szechenyi Istvan University, Hungary
- **Kokosinski, Zbigniew,** Cracow University of Technology, Poland
- **Krawiec, Krzysztof,** Poznan University of Technology, Poland
- **Kulczycki, Piotr,** Systems Research Institute, Polish Academy of Sciences, Poland
- **Kusy, Maciej,** Rzeszow University of Technology, Poland
- **Lilik, Ferenc,** Szechenyi Istvan University, Hungary
- **Lovassy, Rita,** Obuda University, Hungary
- **Malecki, Piotr,** Institute of Nuclear Physics PAN, Poland
- **Mesiar, Radko,** Slovak University of Technology, Slovakia
- **Mora, André Damas**
- **Noguera i Clofent, Carles,** Institute of Information Theory and Automation (UTIA), Academy of Sciences of the Czech Republic, Czech Republic
- **Pamin, Jerzy,** Institute for Computational Civil Engineering, Cracow University of Technology, Poland

- **Petrik, Milan,** Masaryk University, Czech Republic
- **Ribeiro, Rita A.**
- **Sachenko, Anatoly,** Ternopil State Economic University, Ukraine
- **Samotyj, Volodymyr,** Lviv Polytechnic National University, Ukraine
- **Szafran, Bartlomiej,** Faculty of Physics and Applied Computer Science, AGH University of Science and Technology, Poland
- **Tormasi, Alex,** Szechenyi Istvan University, Hungary
- **Wei, Wei,** School of Computer science and engineering, Xi'an University of Technology, China
- **Wysocki, Marian,** Rzeszow University of Technology, Poland
- **Yang, Yujiu,** Tsinghua University, China
- **Zadrozny, Slawomir,** Systems Research Institute, Poland
- **Zajac, Mieczyslaw,** Cracow University of Technology, Poland

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 935–938
DOI: 10.15439/2016F403
ACSIS, Vol. 8. ISSN 2300-5963

Improving precision and accuracy of DTI experiments with the simplified BSD calibration – computer simulations

Karol Borkowski
Faculty of Physics and Applied Computer Science,
AGH University of Science and Technology
30-059 Cracow, Poland
e-mail: Karol.Borkowski@fis.agh.edu.pl

Artur Krzyżak
Faculty of Geology, Geophysics and Environmental Protection,
AGH University of Science and Technology
30-059 Cracow, Poland

Abstract— The *b*-matrix spatial distribution (BSD) is an effective method of improving accuracy of diffusion tensor imaging (DTI) measurements. It is based on a calibration with an anisotropic phantom measured in six positions inside an MRI scanner. However, if the contribution of non-diffusion gradients to the *b*-matrix can be neglected, simplified form of calibration with only 3 positions of the phantom could be sufficient. We called this approach simplified BSD-DTI (sBSD-DTI).

In this paper we introduce the above-mentioned technique and present the results of the computer simulations of BSD-DTI experiments and compare them with standard DTI. The complete BSD method, sBSD as well as BSD with assumption of phantom uniformity (uBSD) were simulated.

The simulations revealed that both simplified methods are less accurate than the complete BSD-DTI. Nevertheless, the calibration procedures and the algorithms are streamlined.

I. Introduction

Diffusion tensor imaging (DTI) is a powerful MRI technique with multiple clinical applications, including presurgical planning and intraoperative guidance in regions adjacent to critical neural tracts, evaluation and treatment of neurodegenerative diseases like epilepsy, multiple sclerosis, Alzheimer's or ischemic stroke [1]. Its accuracy depends strictly on correct identification of the *b*-matrix for a given imaging sequence [2][3]. Due to the complex character of the *b*-matrix and the fact that it is not constant across the volume [2], its analytical derivation is a very tedious and imprecise process [4][5]. Therefore, in the majority of MRI systems, the *b*-matrix is calculated relying only on the diffusion gradient vectors.

Several ways of improving the precision of derivation of the correct form of the *b*-matrix have been reported. Some of them take into account the cross-terms between the diffusion and imaging gradients. That includes refocusing each imaging gradient before turning on the diffusion gradient and refocusing each diffusion gradient before the imaging gradient is applied [6] or acquiring data twice; once with the given diffusion gradient and once with the opposite polarity [7][8]. However, these methods involve a prolonged acquisition time. The cross-terms effect can be also minimized by numerical calculations of the *b*-matrix [9][3], using calibration scans of an isotropic phantom [10] or by establishing the optimal diffusion gradient scheme [11].

In all of the above-mentioned techniques only diffusion, imaging and other known gradients are taken into account, while the background noises and further unknown factors, which may have an impact on the *b*-matrix, are neglected. Moreover, all of these practices assume the same form of the b-matrix throughout the imaging space, what is not true in general.

The *b*-matrix Spatial Distribution in Diffusion Tensor Imaging (BSD-DTI) is a method which enables one to discover the real form of the *b*-matrix for every voxel using a precisely defined anisotropic phantom [2]. It allows to maintain the accuracy and precision of the measurement even if the diffusion properties of the phantom are not spatially constant [12].

In this paper the simplified form of the BSD calibration is introduced. It requires measurement of the phantom in only three positions instead of six. Moreover, we present the results of the computer simulations which depict the effectiveness of this method.

II. Theory

A. Diffusion Tensor Calculation

The elements of a diffusion tensor can be calculated from the formula: [13]

$$\ln\left(\frac{S_n(b)}{S(b_0)}\right) = -\sum_{i,j=1}^{3}(b_{ij} - b_{0ij})D_{ij} \qquad (1)$$

which is the form of the Stejskal-Tanner equation, where $S_n(b)$ is a signal intensity for the nth diffusion gradient, $S(b_0)$ is a signal intensity without any diffusion gradient.
D_{ij} and b_{ij} are the components of the diffusion tensor and the *b*-matrix, respectively.

B. B matrix calculation

The *b*-matrix used in eq. (1) is given by [9]

Research financed by the National Centre of Research and Development, contract. No. PBS2/A2/16/2013 and contract No. STRATEGMED2/265761/10/NCBR/2015.

$$b = \int_0^{2\tau} \left[k(t) - 2H(t-\tau)k(\tau)\right]\left[k(t) - 2H(t-\tau)k(\tau)\right]^T dt \,, \quad (2a)$$

where

$$k(t) = \gamma \int_0^t G(t')dt' \,. \quad (2b)$$

G(t) is a column vector which represents the particular gradient in the imaging sequence. γ is the gyromagnetic ratio, 2τ is the echo time, *H(t)* is the Heaviside unit-step function.

If the imaging and other background gradients can be ignored, the *b*-matrix can be calculated from the dyadic product of the adequate gradients [14]; for example

$$b_n = G_n G_n^T = \begin{bmatrix} g_x \\ g_y \\ g_z \end{bmatrix} \begin{bmatrix} g_x & g_y & g_z \end{bmatrix} = \begin{bmatrix} g_x^2 & g_x g_y & g_x g_z \\ g_x g_y & g_y^2 & g_y g_z \\ g_x g_z & g_y g_z & g_z^2 \end{bmatrix} \quad (3)$$

Otherwise, the expressions for components of the *b*-matrix, due to the cross-terms between diffusion and other gradients, has more complex form:

$$b_{ii} = aG_i^2 + bG_i + c \quad (4a)$$

for the diagonal components and

$$b_{ij} = aG_iG_j + bG_i + cG_j + d \quad (4b)$$

for the off-diagonal.

C. Calibration with anisotropic phantom in three positions

To perform the simplified BSD-DTI (sBSD) calibration the phantom is situated inside an MRI scanner in such a way, that its diffusion tensor is diagonal in the laboratory coordinate system and then the diffusion tensor is measured. Subsequently, the phantom is rotated twice about 90° around the two orthogonal axes of the coordinate system respectively and each time the measurement is repeated. In this situation, according to the Stejskal-Tanner equation, the signal attenuation in voxel (*k, l, m*) in laboratory coordinate frame for a particular diffusion gradient is given by

$$\ln\left(\frac{S(b_{klm})}{S(0)}\right) = -b_{klm} : \begin{bmatrix} D1_{klm} & 0 & 0 \\ 0 & D2_{klm} & 0 \\ 0 & 0 & D3_{klm} \end{bmatrix} = b_{xx,klm}D1_{klm} + b_{yy,klm}D2_{klm} + b_{zz,klm}D3_{klm} \,, \quad (5)$$

Providing that diffusion properties of the phantom are well known in a coordinate system associated with the phantom *(r, s, t),* in laboratory coordinate frame (*k, l, m*) they can be found in two steps [15]. First, the (*k, l, m*) coordinates are found by

$$\begin{bmatrix} k \\ l \\ m \end{bmatrix} = R(\alpha,\beta,\gamma) \cdot \begin{bmatrix} r \\ s \\ t \end{bmatrix}, \quad (6)$$

where $R(\alpha,\beta,\gamma)$ is a rotation matrix.

In the next step the diffusion tensor D_{klm} is derived by rotation transformation

$$= R(\alpha,\beta,\gamma) \cdot D_{rst} \cdot R^T(\alpha,\beta,\gamma) \,. \quad (7)$$

In order to calculate the diagonal components of the *b*-matrix, one must solve the system of three equations for every diffusion gradient direction. Assuming that *b*-matrix has a form given by Eq. (3), off-diagonal components are equal to:

$$b_{ij} = \mathrm{sgn}(g_i g_j)\sqrt{b_{ii}b_{jj}} \,. \quad (8)$$

where $sgn(g_ig_j)$ is a sign of the product g_ig_j, where g_i and g_j are the diffusion gradient strengths in *i* and *j* direction, respectively.

If the diffusion properties of a phantom are assumed to be constant in its whole volume, the *k, l* and *m* indexes of the tensor *D* can be removed. BSD calibration under such assumption has been called uniform BSD-DTI (uBSD-DTI) [7].

III. Materials and Methods

In order to examine the effectiveness of simplified BSD-DTI method in comparison to BSD-DTI, uniform BSD-DTI (uBSD-DTI) and standard DTI (S-DTI), the computer simulations were conducted. First of all, the expressions given by Eq. (4a) and (4b) were derived for an exemplary imaging protocol for Bruker BioSpin 9.7 T imaging system. Components of *b*-matrices are given by

$b_{xx} = 598.5\ G_x^2 + 73.5\ G_x + 12.5,$ (9a)
$b_{yy} = 600\ G_y^2 + 72\ G_y + 12.5,$ (9b)
$b_{zz} = 596\ G_z^2 + 76\ G_z + 13,$ (9c)
$b_{xy} = 597\ G_xG_y + 37\ G_x + 37\ G_y + 12.5,$ (9d)
$b_{xz} = 597.5\ G_xG_z + 38\ G_x + 37\ G_z + 13,$ (9e)
$b_{yz} = 598.5\ G_yG_z + 37.5\ G_y + 36\ G_z + 13.$ (9f)

Diffusion gradient directions were chosen as follows

$G_1 = [0.666, 0.333, 0.666],$
$G_2 = [0.666, -0.333, 0.666],$
$G_3 = [0.333, 0.666, 0.666],$

$G_4 = [-0.333, 0.666, 0.666]$,
$G_5 = [0.666, 0.666, 0.333]$,
$G_6 = [0.666, 0.666, -0.333]$.

Then, the *b*-matrix spatial distribution was established by distorting the diffusion gradients and calculating b matrices for every voxel separately using Eqs. (9a)-(9f). The gradients were distorted systematically in following way

$$G'_{n,xyz} = G_n \sigma (x + y + z - 36) / 12.4904,$$

where $G'_{n,xyz}$ is a distorted diffusion gradients for a voxel (x, y, z) in the laboratory coordinate system. G_n is a diffusion gradient in nth direction. σ is the relative standard deviation of the spatial gradient distribution. In described simulations this parameter was set equal to 5%.

Subsequently, simulations of the BSD-DTI calibrations were performed for a virtual anisotropic phantom. The phantom was a cube with a side equal to 41 voxels. This size is required to indicate the *b*-matrix spatial distribution in the 25x25x25 voxels field of view (FOV). Diffusion properties of the phantom were similar to the diffusion properties of the anisotropic plate phantom [16]. The mean eigenvalues of its diffusion tensor were 0.002, 0.002 and 0.0005 mm^2/s. In order to imitate the properties of a real phantoms the tensor was spatially distorted by Gaussian noise with relative standard deviation equal to 1% and mean value equal to 0. The distribution of the *b*-matrix was calculated in four ways. Using the complete BSD-DTI method, uniform BSD, simplified BSD and simplified uniform BSD (uniform BSD with three phantom's positions - usBSD). Finally, diffusion tensor experiment were simulated for other two virtual homogeneous phantoms P1 and P2. The former was characterized by diffusion tensor with eigenvalues equal to 0.001, 0.002 and 0.003 mm^2/s. The latter was isotropic with diffusion coefficient equal to 0.002 mm^2/s. The first phantom was orientated in such a way that in the laboratory coordinate system its edges were inclined at 45^o angle from the axes of the laboratory coordinate system. The tensor was calculated using each of the aforementioned *b*-matrix spatial distributions, spatially constant *b*-matrix (S-DTI) and distribution calculated with approximation given by Eq. (3).

IV. Results

Tables 1 and 2 report the *b*-matrix elements calculated according to expressions (3) and (9a)-(9f), while table 3 reports the relative difference between them.

Mean values and relative standard deviations of diffusion tensor eigenvalues obtained in the simulations are reported in table 4.

In the case of isotropic phantom P2 as well as the phantom P1 orientated in such a way that its diffusion tensor was diagonal and in the case of *b*-matrix calculated with dyadic approximation, obtained results were similar for the simplified BSD and the BSD-DTI approaches.

Computing time of the *b*-matrix spatial distribution calculations performed on a personal computer was 331s, 227s, 72s and 44s for BSD-DTI, uBSD, sBSD and usBSD, respectively.

TABLE I.
B-MATRIX ELEMENTS CALCULATED ACCORDINGLY TO EQS. 9A – 9F.

	bxx	byy	bzz	bxy	bxz	byz
G1	315.00	90.67	315.56	169.67	315.56	169.50
G2	315.00	42.67	315.56	-120.33	315.56	-121.50
G3	91.00	314.67	315.56	169.67	170.11	315.00
G4	42.00	314.67	315.56	-120.33	-120.78	315.00
G5	315.00	314.67	91.56	314.67	170.44	170.00
G6	315.00	314.67	40.89	314.67	-119.78	-120.00

TABLE 1.
B-MATRIX ELEMENTS CALCULATED ACCORDING TO EQ. 3.

	bxx	byy	bzz	bxy	bxz	byz
G1	315.00	90.67	315.56	169,00	315,28	169,15
G2	315.00	42.67	315.56	-115,93	315,28	-116,03
G3	91.00	314.67	315.56	169,22	169,46	315,11
G4	42.00	314.67	315.56	-114,96	-115,12	315,11
G5	315.00	314.67	91.56	314,83	169,82	169,73
G6	315.00	314.67	40.89	314,83	-113,49	-113,43

TABLE 2.
RELATIVE DIFFERENCE BETWEEN BOTH SETS OF B-MATRIX ELEMENTS.

	bxx	byy	bzz	bxy	bxz	byz
G1	0%	0%	0%	0.39%	0.09%	0.21%
G2	0%	0%	0%	3.66%	0.09%	4.50%
G3	0%	0%	0%	0.26%	0.38%	-0.04%
G4	0%	0%	0%	4.46%	4.68%	-0.04%
G5	0%	0%	0%	-0.05%	0.36%	0.16%
G6	0%	0%	0%	-0.05%	5.25%	5.47%

TABLE 3.
AVERAGE EIGENVALUES [MM^2/S], ITS RELATIVE DIFFERENCE Δ BETWEEN REAL AND MEASURED ONES AND RELATIVE STANDARD DEVIATION (RSD) ACROSS THE FOV FOR PHANTOM P1 OBTAINED BY SIMULATING THE S-DTI, BSD-DTI, UBSD, SBSD AND USBSD EXPERIMENTS.

	S-DTI	BSD-DTI	uBSD	sBSD	usBSD
av. E1	1.0021E-03	0.001	9.9859E-04	9.9630E-04	9.9521E-04
av. E2	2.0048E-03	0.002	2.0002E-03	2.0137E-03	2.0118E-03
av. E3	3.0074E-03	0.003	3.0006E-03	3.0520E-03	3.0218E-03
ΔE1	0.206%	0%	-0.141%	-0.372%	-0.481%
ΔE2	0.240%	0%	0.011%	0.682%	0.586%

ΔE3	0.245%	0%	0.021%	1.705%	0.721%
RSD E1	9.10%	0%	3.16%	0.03%	0.49%
RSD E2	9.81%	0%	1.62%	0.08%	4.19%
RSD E3	9.89%	0%	1.18%	0.19%	3.67%

I. Discussion

The calculations of the *b*-matrix for the exemplary imaging sequence show that neglecting the cross-terms leads to a systematic error in the off-diagonal elements up to 5.47%.

As expected the BSD-DTI method ensures the highest accuracy and precision of DTI experiments, but also takes the longest period of computing time. However, in the case of post hoc image analysis it may be irrelevant.

Accuracy of the uBSD-DTI is almost as good as in the case of BSD-DTI. Nevertheless, due to not taking into account the imperfections of the phantom the precision is lower.

In the simplified BSD-DTI, due to the differences in *b*-matrix elements reported in tab. 3, precision is slightly lower. Nevertheless, in comparison with S-DTI the accuracy is significantly improved. It also allows to reduce the computing time.

When one combines the uniform and simplified BSD approaches, in comparison to complete BSD-DTI method, the accuracy and the precision definitely decrease but also the calculations time is drastically reduced.

If the diffusion tensor of imaged object is diagonal in laboratory coordinate system (like in the cases of isotropic media) the off-diagonal components of the *b*-matrix can be neglected, therefore the quality of sBSD-DTI is the same that of BSD-DTI.

II. Conclusion

BSD-DTI is a robust and effective method of improving precision and accuracy of DTI experiment. Nevertheless, sBSD approach can be an expedient trade-off between quality and simplicity, especially when non-diffusion gradient presents during the imaging sequence are negligible.

Acknowledgment

The work was financed by the National Centre of Research and Development, contract. No. PBS2/A2/16/2013 and contract No. STRATEGMED2/265761/10/NCBR/2015.

K.B. thanks Marian Smoluchowski Cracow Scientific Consortium – KNOW for the support

References

[1] A. Lerner, M. Mogensen, P. Kim, M. Shiroishi, D. Hwang and M. Law, "Clinical Applications of Diffusion Tensor Imaging", World Neurosurgery, vol. 82, no. 1-2, pp. 96-109, 2014, doi: 10.4103/0971-3026.38505

[2] A. Krzyżak and Z. Olejniczak, "Improving the accuracy of PGSE DTI experiments using the spatial distribution of b matrix", Magnetic Resonance Imaging, vol. 33, no. 3, pp. 286-295, 2015, doi: 10.1016/j.mri.2014.10.007

[3] D. Güllmar, J. Haueisen, and J. R. Reichenbach, “Analysis of b-value calculations in diffusion weighted and diffusion tensor imaging,” Concepts Magn. Reson., vol. 25A, no. 1, pp. 53–66, Mar. 2005, doi: 10.1002/cmr.a.20031

[4] J. Mattiello, P. Basser and D. Lebihan, "Analytical Expressions for the b Matrix in NMR Diffusion Imaging and Spectroscopy", Journal of Magnetic Resonance, Series A, vol. 108, no. 2, pp. 131-141, 1994, doi:10.1006/jmra.1994.1103

[5] J. Mattiello, P. Basser and D. Le Bihan, "The b matrix in diffusion tensor echo-planar imaging", Magnetic Resonance in Medicine, vol. 37, no. 2, pp. 292-300, 1997, doi: 10.1002/mrm.1910370226

[6] Hsu EW, Mori S. Analytical expressions for the NMR apparent diffusion coefficients in an anisotropic system and a simplified method for determining fiber orientation. Magn Reson Med 1995;34:194–200, doi: 10.1002/mrm.1910370226

[7] Neeman M, Freyer JP, Sillerud LO. A simple method for obtaining cross-term-free images for diffusion anisotropy studies in NMR microimaging. Magn Reson Med 1991;21:138–43, doi: 10.1002/mrm.1910210117

[8] Güllmar D, Haueisen J, Reichenbach JR. Analysis of b-value calculations in diffusion weighted and diffusion tensor imaging. Concepts Magn Reson Part A 2005;25A:53–66. doi: 10.1002/cmr.a.20031

[9] S. Boujraf, R. Luypaert, H. Eisendrath and M. Osteaux, "Effect of accurately calculated b matrix on the evaluation of the diffusion tensor using echoplanar diffusion tensor imaging", ITBM-RBM, vol. 23, no. 6, pp. 340-344, 2002, doi:10.1016/S1297-9562(02)90003-3

[10] Ozcan A. Minimization of Imaging Gradient Effects in Diffusion Tensor Imaging. IEEE Trans Med Imaging 2011;30:642–54. doi:10.1109/TMI.2010.2090539.

[11] [1] G. Nair and X. P. Hu, “Manifestation and Post-hoc Correction of Gradient Cross-term Artifacts in DTI,” Magn Reson Imaging, vol. 30, no. 6, pp. 764–773, Jul. 2012, doi: 10.1016/j.mri.2012.02.021

[12] K. Borkowski and A. Krzyżak, “Simulations of rotation of the anisotropic phantom in BSD-DTI” in Magnetic Resonance Materials in Physics, Biology and Medicine, 2015 vol. 28 suppl. 1, s. 467–468. ESMRMB 2015 : 32nd annual scientific meeting : Edinburgh, UK, doi: 10.1007/s10334-015-0490-7

[13] P. Basher, J. Mattiello and D. Lebihan, "Estimation of the Effective Self-Diffusion Tensor from the NMR Spin Echo", Journal of Magnetic Resonance, Series B, vol. 103, no. 3, pp. 247-254, 1994, doi:10.1006/jmrb.1994.1037

[14] P. Kingsley, "Introduction to diffusion tensor imaging mathematics: Part II. Anisotropy, diffusion-weighting factors, and gradient encoding schemes", Concepts in Magnetic Resonance Part A, vol. 28, no. 2, pp. 123-154, 2006, DOI: 10.1002/cmr.a.20049

[15] A. Krzyżak and K. Borkowski, "Theoretical analysis of phantom rotations in BSD-DTI" in Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE , vol., no., pp.410-413, 25-29 Aug. 2015, doi: 10.1109/EMBC.2015.7318386

[16] K. Kłodowski and A. Krzyżak, "Innovative anisotropic phantoms for calibration of diffusion tensor imaging sequences", Magnetic Resonance Imaging, vol. 34, no. 4, pp. 404-409, 2016, doi: 10.1016/j.mri.2015.12.010

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 939–942
DOI: 10.15439/2016F493
ACSIS, Vol. 8. ISSN 2300-5963

The matrix-based description approach for the multistage differential-algebraic processes

Paweł Drąg
Department of Control Systems and Mechatronics
Wrocław University of Technology
Janiszewskiego 11-17, 50-372 Wrocław, Poland
Email: pawel.drag@pwr.edu.pl

Krystyn Styczeń
Department of Control Systems and Mechatronics,
Wrocław University of Technology
Janiszewskiego 11-17, 50-372 Wrocław, Poland
Email: krystyn.styczen@pwr.edu.pl

***Abstract*—In the article a new insight into an optimal control problem of the multistage processes has been given. The multistage descriptor processes with differential-algebraic constraints are under considerations. The new representation of the descriptor model has been presented. Moreover, the new structures to represent the differential state variables, algebraic state variables, control function, as well the global parameters have been introduced. The generalized description enables the unified representation of a broad group of the multistage processes with differential-algebraic relations and indicates on the physical interpretation of the process variables.**

***Index Terms*—optimal control, descriptor systems, DAE systems, generalized description approach.**

I. Introduction

IN THE article some new aspects of the multistage descriptor control systems and their unified representation have been discussed. The control and optimization of the technological processes with the differential-algebraic constraints have nowadays a great importance [12], [18], [19]. Therefore, one of main aims of the presented considerations is to emphasize the applicability of the obtained theoretical results.

The multistage technological processes are often designed in technology, especially in biotechnology [1], [9], [10], [11], chemical engineering [4], [8], as well as in environmental engineering [13], [14], [15].

The presence of a large number of successive stages is a characteristic feature of the modern technological processes. This reflects a complexity, as well as an arrangement of the designed processes. Therefore, the large number of the state variables, control functions and other process parameters needs a generalized description of the multistage systems. The generalized process description can enable us to unify the new control algorithms design.

The large-scale technological systems should be flexible to change their configurations, as well as to be open for necessary modifications. These requirements indicate a one of the new, unified process description approach, which can enable the process to develop.

To design the methodology, which allows us the integration of the additional stages within the process, is a task that requires the specific theoretical basis preparation. The generalized description methodology is the main result of the carried out considerations.

The rest of the paper has been organized in the following way. In Section II the general optimal control problem with the multistage differential-algebraic constraints has been formulated. In Section III the generalized description approach for the multistage DAE system has been proposed. Then, in Section IV, the generalized description approach has been illustrated with the three-stage chemical process. In Section V the presented considerations have been concluded.

II. Statement of the optimal control problem

The origin of the considerations about the multistage descriptor processes analysis are connected with two articles [16] and [17]. The assumptions and algorithms proposed in these articles are treated as the mile stones in control and optimization of the multistage technological processes. The observed progress in the control algorithms has been presented in details in three monographs [2], [3], [5].

In this section the most important features of the optimal control problem with the multistage differential-algebraic constraints have been given.

Assumption 2.1: The multistage process are consisted on the N successive stages, where N is the known number and $N \in \mathcal{N}$.

Assumption 2.2: [6] Each stage can be described by the system of the differential-algebraic equations. The index of the DAEs system is not greater than 1.

Moreover, to simulate, optimize and control a process, the time range has to be known.

Assumption 2.3: The range of the process time duration is known *a priori* and

$$t \in [t_0 \quad t_F]. \tag{1}$$

According to the Assumptions 2.1 and 2.3, the time domain of the each considered stage can be defined separately.

Definition 2.1: The time domain of the stage number $i = 1, \cdots, N$ can be defined as

$$t^i \in [t_0^i \quad t_F^i]. \tag{2}$$

978-83-60810-90-3/$25.00

In the stage number i and during the time domain t^i the process is governed by the set of the differential-algebraic equations.

Assumption 2.4: [7] At the stage number i, the process is governed by the set of the differential-algebraic relations

$$\begin{array}{rcl} \dot{\mathbf{y}}^i(t) & = & F^i(\mathbf{y}^i(t), \mathbf{z}^i(t), \mathbf{u}^i(t), \mathbf{p}, t^i) \\ 0 & = & G^i(\mathbf{y}^i(t), \mathbf{z}^i(t), \mathbf{u}^i(t), \mathbf{p}, t^i) \end{array} \tag{3}$$

where $i = 1, \cdots, N$ is the number of the stage considered, $\mathbf{y}^i(t) \in \mathcal{R}^{n_{\mathbf{y}^i}}$ is a differential state variable, $\mathbf{z}^i(t) \in \mathcal{R}^{n_{\mathbf{z}^i}}$ is an algebraic state variable and $\mathbf{u}^i(t) \in \mathcal{R}^{n_{\mathbf{u}^i}}$ denotes the unknown control function. The independent variable (e.g. time or length of the chemical reactor) is denoted as $t \in \mathcal{R}$.

Definition 2.2: The relations in (3) are defined as follows

$$F^i : \mathcal{R}^{\mathbf{y}^i} \times \mathcal{R}^{\mathbf{z}^i} \times \mathcal{R}^{\mathbf{u}^i} \times \mathcal{R}^{\mathbf{P}} \times \mathcal{R} \to \mathcal{R}^{\mathbf{y}^i} \tag{4}$$

and

$$G^i : \mathcal{R}^{\mathbf{y}^i} \times \mathcal{R}^{\mathbf{z}^i} \times \mathcal{R}^{\mathbf{u}^i} \times \mathcal{R}^{\mathbf{P}} \times \mathcal{R} \to \mathcal{R}^{\mathbf{z}^i}. \tag{5}$$

The minimized process performance index, which can be treated as the measure of the control quality, is defined as follows

$$\begin{array}{l} \min_{\mathbf{u}^i(t), i=1,\cdots,N} Q \\ = \sum_{i=1}^{N} \int_{t_0^i}^{t_F^i} \mathcal{L}(\mathbf{y}^i(t^i), \mathbf{z}^i(t^i), \mathbf{u}^i(t^i), \mathbf{p}, t^i) dt \\ + \mathcal{E}(\mathbf{y}^N(t_F^N), \mathbf{z}^N(t_F^N), t_F^N). \end{array} \tag{6}$$

III. The generalized description approach

In this section we would like to propose the new generalized description approach for the multistage differential-algebraic processes. The presented methodology extends the proposition from the article [16].

Definition 3.1: **The particular differential state vector** for the i-th stage $\mathbf{y}^i$ is consisted from all the differential state variables in the whole process. The differential state variables, which are constant during the i-th stage, are equal to their initial values.

The particular differential state vector is characteristic for the each stage and for the i-th takes the form

$$\mathbf{y}^i(t) = \begin{bmatrix} y_1^i(t) \\ \vdots \\ y_{n_{\mathbf{y}^i}}^i(t) \end{bmatrix} \in \mathcal{R}^{n_{\mathbf{y}^i}}. \tag{7}$$

Therefore, the particular differential state vector for the each stage has the same size. Moreover, the particular differential state vectors can be combined together to form the differential state matrix.

Definition 3.2: **The differential state matrix** is formed by the particular differential state vectors is the following way

$$\mathbf{Y}(t) = [\mathbf{y}^1(t) \quad \cdots \quad \mathbf{y}^N(t)], \tag{8}$$

where $\mathbf{y}^i(t)$, $i = 1, \cdots, N$ denotes the particular differential state vectors.

The properties of the differential state matrix:

1) The number of the differential state matrix rows is equal to the size of the particular differential state vector. Therefore, the number of the differential state variables can be identified.
2) The number of the differential state matrix columns is equal to the number of the process stages.
3) The differential state matrix indicates the stages, in which the chosen differential state variable has the same physical interpretation.

In the same way like the differential state matrix, **the algebraic state matrix** can be formed

$$\mathbf{Z}(t) = [\mathbf{z}^1(t) \quad \cdots \quad \mathbf{z}^N(t)], \tag{9}$$

the control matrix

$$\mathbf{U}(t) = [\mathbf{u}^1(t) \quad \cdots \quad \mathbf{u}^N(t)], \tag{10}$$

as well as **the matrix of the parameters constant in the time**

$$\mathbf{P} \equiv \mathbf{p}. \tag{11}$$

The differential state matrix, the algebraic state matrix, the control matrix, as well as the matrix of the parameters constant in the time can be used to define the multistage descriptor process

$$\begin{array}{rcl} \dot{\mathbf{Y}}(t) & = & \mathbf{F}(\mathbf{Y}(t), \mathbf{Z}(t), \mathbf{U}(t), \mathbf{P}, t) \\ 0 & = & \mathbf{G}(\mathbf{Y}(t), \mathbf{Z}(t), \mathbf{U}(t), \mathbf{P}, t) \end{array} \tag{12}$$

where

$$\mathbf{F} : \mathcal{R}^{n_{\mathbf{Y}}} \times \mathcal{R}^{n_{\mathbf{Z}}} \times \mathcal{R}^{n_{\mathbf{U}}} \times \mathcal{R}^{n_{\mathbf{P}}} \times \mathcal{R} \to \mathcal{R}^{n_{\mathbf{Y}}} = \mathcal{R}^{n_{\mathbf{Y}}} \times \mathcal{R}^{N}, \tag{13}$$

$$\mathbf{G} : \mathcal{R}^{n_{\mathbf{Y}}} \times \mathcal{R}^{n_{\mathbf{Z}}} \times \mathcal{R}^{n_{\mathbf{U}}} \times \mathcal{R}^{n_{\mathbf{P}}} \times \mathcal{R} \to \mathcal{R}^{n_{\mathbf{Z}}} = \mathcal{R}^{n_{\mathbf{Z}}} \times \mathcal{R}^{N}, \tag{14}$$

and t denotes the duration time of the whole considered process

$$t \in [t_0 \quad t_F]. \tag{15}$$

The generalized system description has been used to obtain the new form of the optimal control problem of three-stage chemical process .

IV. Application in the multistage process

The matrix-based approach for the multistage DAE systems description has been applied to model of the three-reactor process (Fig. 1). The considered system is consisted of two chemical reactors and a mixing stage between them [16].

At the beginning, the first reactor is loaded with the substrate A with the volume 0.1 m^3 and concentration 2000 mol/m^3.

Due to reactions, which take a place in the system, the products B and C are obtained according to the scheme

$$2A \rightarrow B \rightarrow C. \tag{16}$$

Additionally, the first reactor was equipment with a heating exchanger, which can be used to control the process temperature and in this way - to influence the trajectories of the process variables. The concentrations of the substrate and products are changing in the following way

$$\dot{C}_A = -2k_1(T)C_A^2 \tag{17}$$

$$\dot{C}_B = 2k_1(T)C_A^2 - k_2(T)C_B \tag{18}$$

$$\dot{C}_C = k_2(T)C_B \tag{19}$$

with the kinetics constraints

$$k_1(T) = 0.0444\exp(-2500/T) \tag{20}$$

and

$$k_2(T) = 6889.0\exp(-5000/T). \tag{21}$$

Then, in the mixing stage at the time t_2^0, the component B of concentrations C_B^0 = 600mol/m^3 and some volume S is added. Therefore, the volume and concentrations of the substrates are changing, so the following relations are satisfied

$$V_2C_A(t_0^2) = V_1C_A(t_F^1) \tag{22}$$

$$V_2C_B(t_0^2) = V_1C_B(t_F^1) + SC_B^0 \tag{23}$$

$$V_2C_C(t_0^2) = V_1C_C(t_F^1) \tag{24}$$

where V_1 is the volume of substrates loaded at the beginning of the first reactor. Therefore, the volume V_2 in the second reactor is given by

$$V_2 = V_1 + S \tag{25}$$

The volume S is a decision parameter with

$$0 \leq S \leq 0.1(m^3) \tag{26}$$

After the mixing stage, the substrates are loaded into the last reactor, where three reactions are taking a place

$$B \rightarrow D \tag{27}$$

$$B \rightarrow E \tag{28}$$

$$2B \rightarrow F \tag{29}$$

In the 2*nd* reactor, the reactions take place under isothermal conditions. The state variables are changing in the following way

$$\dot{C}_A = 0 \tag{30}$$

$$\dot{C}_B = -0.02C_B - 0.05C_B - 2 \times 4.0 \times 10^{-5}C_B^2 \tag{31}$$

$$\dot{C}_C = 0 \tag{32}$$

$$\dot{C}_D = 0.02C_B \tag{33}$$

$$\dot{C}_E = 0.05C_B \tag{34}$$

$$\dot{C}_F = 4.0 \times 10^{-5}C_B^2 \tag{35}$$

The decision variables are the profile of the temperature $T(t)$, the duration time of the reactions in each stage, and the amount S of component B, which is added at the mixing step.

The process is aimed to maximize the amount of the product D at the output of the 2*nd* reactor

$$\max_{t_1,t_2,S,T(t)} V_2C_D(t_2) \tag{36}$$

subject to the constraints on the temperature profile

$$298 \leq T(t) \leq 398(K), \qquad t^1 \in [t_0^1 \quad t_F^1]. \tag{37}$$

According to the presented methodology, the three-stage technological process can be rewritten in the generalized matrix form. There are six state variables, which indicate the appropriate concentrations

$$\mathbf{y}(t) = \begin{bmatrix} y_A(t) \\ y_B(t) \\ y_C(t) \\ y_D(t) \\ y_E(t) \\ y_F(t) \end{bmatrix}. \tag{38}$$

The particular vector of the state variables takes the form

$$\mathbf{y}^1(t^1) = \begin{bmatrix} y_A^1(t^1) \\ y_B^1(t^1) \\ y_C^1(t^1) \\ y_D^1(t_0^1) \\ y_E^1(t_0^1) \\ y_F^1(t_0^1) \end{bmatrix}, \tag{39}$$

$$\mathbf{y}^2(t^2) = \begin{bmatrix} y_A^2(t_0^2) \\ y_B^2(t_0^2) \\ y_C^2(t_0^2) \\ y_D^2(t_0^2) \\ y_E^2(t_0^2) \\ y_F^2(t_0^2) \end{bmatrix}, \tag{40}$$

$$\mathbf{y}^3(t^3) = \begin{bmatrix} y_A^3(t_0^3) \\ y_B^3(t^3) \\ y_C^3(t_0^3) \\ y_D^3(t^3) \\ y_E^3(t^3) \\ y_F^3(t^3) \end{bmatrix}. \tag{41}$$

Therefore, the matrix of the state variables for the considered process takes the form

$$\mathbf{Y}(t) = [\mathbf{y}^1(t^1) \quad \mathbf{y}^2(t^2) \quad \mathbf{y}^3(t^3)]. \tag{42}$$

The control function is constructed in a similar way

$$\mathbf{u}(t) = \begin{bmatrix} T(t) \\ S \\ C_0 \end{bmatrix}. \tag{43}$$

$$\mathbf{u}^1(t^1) = \begin{bmatrix} T(t) \\ 0 \\ 0 \end{bmatrix}. \tag{44}$$

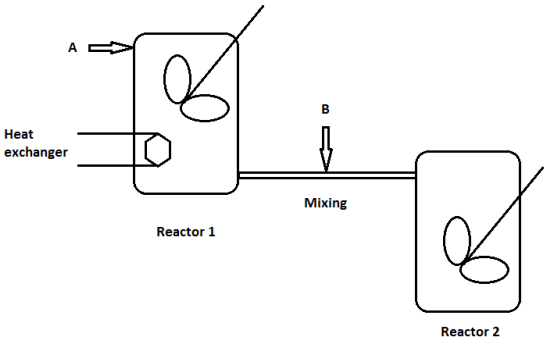


Fig. 1. The three-stage chemical process.

$$\mathbf{u}^2(t^2) = \begin{bmatrix} 0 \\ S \\ C_0 \end{bmatrix}. \tag{45}$$

$$\mathbf{u}^3(t^3) = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}. \tag{46}$$

and finally

$$\mathbf{U}(t) = [\mathbf{u}^1(t^1) \quad \mathbf{u}^2(t^2) \quad \mathbf{u}^3(t^3)]. \tag{47}$$

V. Conclusion

In the presented article the considerations about the new generalized description approach for the multistage technological processes with the differential-algebraic constraints were discussed. Therefore, the proposed approach was characterized by some important features:

a) the general and simple schematic multistage processes description,

b) the physical interpretation of the process variables,

c) easy modification of the process by addition of new elements to the model,

d) integration of processes developed according to different standards,

e) indication the successive stage of the process by the matrix-based structure of the process variables.

From the presented reasons, the carried out theoretical considerations possess a significant practical application. Thus, the generalized description approach has been used to obtain the new form of the three-stage chemical process model.

Acknowledgment

This work has been supported by the National Science Center under grant: UMO-2012/07/B/ST7/01216.

References

[1] J. R. Banga, E. Balsa-Canto, C. G. Moles, A. A. Alonso. 2005. Dynamic optimization of bioprocesses: Efficient and robust numerical strategies. Journal of Biotechnology. 117:407-419, http://dx.doi.org/10.1016/j.jbiotec.2005.02.013

[2] J.T . Betts. 2010. Practical Methods for Optimal Control and Estimation Using Nonlinear Programming, Second Edition. SIAM, Philadelphia, http://dx.doi.org/10.1137/1.9780898718577

[3] L. T. Biegler. 2010. Nonlinear Programming. Concepts, Algorithms and Applications to Chemical Processes. SIAM, Philadelphia, http://dx.doi.org/10.1137/1.9780898719383

[4] L. T. Biegler. 2014. Nonlinear programming strategies for dynamic chemical process optimization. Theoretical Foundations of Chemical Engineering. 48:541-554, http://dx.doi.org/10.1134/S0040579514050157

[5] L. T. Biegler, S. Campbell, V. Mehrmann. 2012. DAEs, Control, and Optimization. Control and Optimization with Differential-Algebraic Constraints. SIAM, Philadelphia, http://dx.doi.org/10.1137/9781611972252.ch1

[6] K. E. Brenan, S.L. Campbell, L. R. Petzold. 1996. Numerical Solution of Initial- Value Problems in Differential-Algebraic Equations. SIAM, Philadelphia, http://dx.doi.org/10.1137/1.9781611971224

[7] M. Diehl, H. G. Bock, J. P. Schlöder, R. Findeisen, Z. Nagy, F. Allgöwer. 2002. Real-time optimization and nonlinear model predictive control of processes governed by differential-algebraic equations. Journal of Process Control. 12:577-585, http://dx.doi.org/10.1016/S0959-1524(01)00023-3

[8] P. Drąg, K. Styczeń. 2012. A Two-Step Approach for Optimal Control of Kinetic Batch Reactor with electroneutrality condition. Przeglad Elektrotechniczny. 6/2012, pp. 176-180.

[9] S. Drozdek, U. Bazylińska. 2016. Biocompatible oil core nanocapsules as potential co-carriers of paclitaxel and fluorescent markers: preparation, characterization, and bioimaging. Colloid and Polymer Science. 294:225-237, http://dx.doi.org/10.1007/s00396-015-3767-5

[10] S. Fidanova, M. Paprzycki, O. Roeva. 2014. Hybrid GA-ACO algorithm for a model parameters identification problem. Proceedings of the 2014 Federated Conference on Computer Science and Information Systems pp. 413–420, http://dx.doi.org/ 10.15439/2014F373

[11] S. Fidanova, O. Roeva. 2013. Metaheuristic techniques for optimization of an E. coli cultivation model. Biotechnology and Biotechnological Equipment. 27:3870-3876, http://dx.doi.org/10.5504/BBEQ.2012.0136

[12] P. S. Harvey Jr, H. P. Gavin, J. T. Scruggs. 2013. Optimal performance of constrained control systems. Smart Materials and Structures. 21:085001, http://dx.doi.org/10.1088/0964-1726/21/8/085001

[13] M. Kwiatkowska. 2015. DAEs method for time-varying indoor air parameters evaluation. In: A. Kotowski, K. Piekarska, B. Kaźmierczak (eds.) Interdyscyplinarne zagadnienia w inżynierii i ochronie środowiska T. 6. Wrocław 2015, pp. 214-220.

[14] M. Kwiatkowska, A. Szczurek, P. Drąg. 2016. Zastosowanie równań różniczkowo-algebraicznych do predykcji zmian parametrów powietrza wewnętrznego. Przeglad Elektrotechniczny. 5/2015, pp. 181-184, http://dx.doi.org/10.15199/48.2016.05.34

[15] K. Matyja, A. Małachowska-Jutsz, A. Mazur, K. Grabas. Assessment of toxicity using dehydrogenases activity and mathematical modeling. Ecotoxicology. 25:924-939, http://dx.doi.org/10.1007/s10646-016-1650-x.

[16] V. S Vassiliadis, R. W. H. Sargent, C. C. Pantelides. 1994. Solution of a Class of Multistage Dynamic Optimization Problems. 1. Problems without Path Constraints. Ind. Eng. Chem. Res. 33:2111-2122, http://dx.doi.org/10.1021/ie00033a014

[17] V. S Vassiliadis, R. W. H. Sargent, C. C. Pantelides. 1994. Solution of a Class of Multistage Dynamic Optimization Problems. 2. Problems with Path Constraints. Ind. Eng. Chem. Res. 33: 2123-2133, http://dx.doi.org/10.1021/ie00033a015

[18] Z.-H. Yang, W.-J. Cui, Y. Tang. 2008. Optimal control with DAE constraints. International Conference on Industrial Engineering and Engineering Management, 2008. pp. 188-192, http://dx.doi.org/10.1109/IEEM.2008.4737857

[19] Z.-H. Yang, F. Guo. 2012. Optimal Control Conditions with Differential-Algebraic Equation Constraints. Advanced Science Letters. 6:654-659, http://dx.doi.org/10.1166/asl.2012.2300

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 943–946
DOI: 10.15439/2016F457
ACSIS, Vol. 8. ISSN 2300-5963

Approximation of the actual spatial distribution of the b-matrix in diffusion tensor imaging with bivariate polynomials

Krzysztof Kłodowski
Faculty of Physics and Applied Computer Science
AGH University of Science and Technology
in Kraków
al. Mickiewicza 30, 30-059 Kraków, Poland
Email: klodowski@fis.agh.edu.pl

Piotr Łukasik, Artur T. Krzyżak
Faculty of Geology, Geophysics and Environmental Protection
AGH University of Science and Technology
in Kraków
al. Mickiewicza 30, 30-059 Kraków, Poland

***Abstract*—The aim of this work was to find an analytical expression describing the b-matrix spatial distribution (BSD) in diffusion tensor imaging, obtained by means of simple calibration to a water isotropic phantom.**

The bivariate second degree polynomial function was fitted for the complete set of spatially distributed b-matrix elements derived through measurements on a 3 Tesla clinical scanner.

Smooth, noise free b-matrices were obtained with clear patterns of systematic errors. Diffusion tensor eigenvalues were derived with much better accuracy than for previous BSD calibration. The proposed approach does not require many averages during the acquisition of the phantom and thus can shorten the BSD calibration.

I. Introduction

DIFFUSION Magnetic Resonance Imaging (dMRI) which appeared in the eighties and quickly developed towards Diffusion Tensor Imaging (DTI) [1] [2], has found number of clinical applications. It became very successful tool for neurological structural and functional imaging [3]. However, DTI is intrinsically prone to the numerous artifacts [4], which makes any quantitative comparison of the images obtained by different scanners questionable.

In order to calculate a diffusion tensor one has to acquire series of images, each with a different diffusion sensitizing gradient applied. The gradients can differ in both magnitude and orientation. Information about diffusion gradient schemes is stored in a b-matrix. The most common way of deriving the b-matrix is an approximated analytical calculation of each element. However, the applied diffusion gradients interact with other magnetic field gradients applied during the imaging sequence and thus the resultant gradients differ from what was expected. Variations of the magnetic susceptibility across the sample volume can also affect the b-matrix [5].

Several methods to partially solve such problems were proposed. Some of them introduced bipolar gradients which cancel out the cross-terms between the applied gradients [6], [7]. The other focused on choosing the best gradient encoding scheme [8], [9], [10]. There were also post processing methods [11] and phantom calibration techniques [12] suggested.

Research financed by the National Centre of Research and Development (contract No. PBS2/A2/16/2013)

A recent report revealed that for particular imaging sequence parameters the errors superimposed on the b-matrix have a systematic character and can be reduced through a calibration procedure [13]. The solution is based on the derivation of the b-matrix spatial distribution (BSD) which substitutes the standard b-matrix, constant for an entire volume. The BSD method improves the accuracy of diffusion measurements, but since the calibration is based on a real data, the derived distribution of the b-matrix is always biased with noise.

In this paper we present a procedure of deriving the noise free approximation of the actual spatial distribution of the b-matrices. The results are compared with both standard DTI and hitherto BSD-DTI.

II. Materials and methods

Standard diffusion tensor imaging of a water isotropic phantom was performed on a 3 Tesla clinical scanner. Six diffusion gradient directions were applied and the b-value was set to $1000\,\mathrm{s/mm^2}$. The imaging was done in axial orientation, 25 interleaved slices with voxel size 1x1x3 mm were taken, with number of averages set to 4. The acquisition was repeated then for the BSD calibration purposes.

The spatial distribution of the b-matrices was obtained through simple calibration to water phantom. The experiments were carried out in stable thermal conditions in temperature of $21\,^{\circ}\mathrm{C}$. The diffusion tensor eigenvalues for an isotropic water phantom in constant temperature are all equal and can be derived theoretically from the Einstein-Smoluchowski equation [14]:

$$\left\langle r^2 \right\rangle = 6Dt, \tag{1}$$

where: $\left\langle r^2 \right\rangle$ – is a mean square displacement, D – is a diffusion coefficient, and t – is a diffusion time. The off-diagonal elements of the tensor can be assumed equal to zero, since they describe the rotations of the diffusion ellipsoid being in this case spherically symmetrical. With the above assumptions

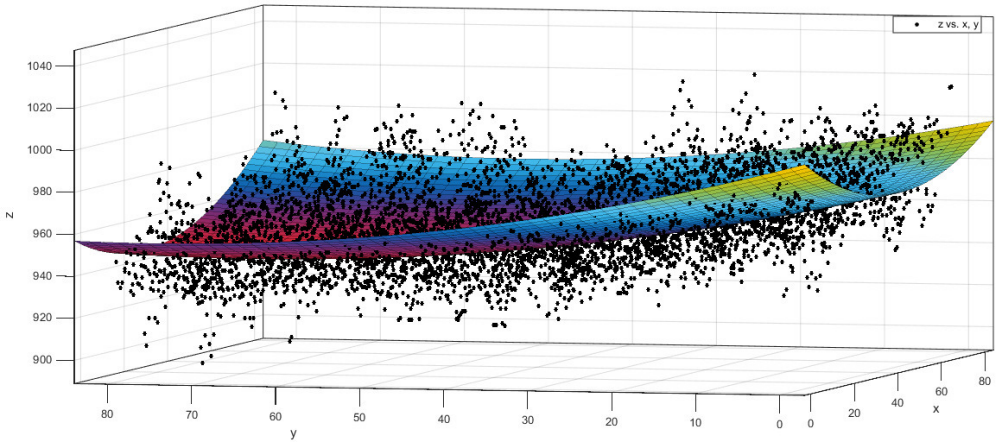


Fig. 1. Plot of bivariate polynomial fitted to data. In this case the diffusion gradient was applied solely along the x axis. The expression of the fitted function is the following: $f(x, y) = 992.5 - 1.021x - 1.043y + 0.013x^2 + 0.004xy + 0.007y^2$.

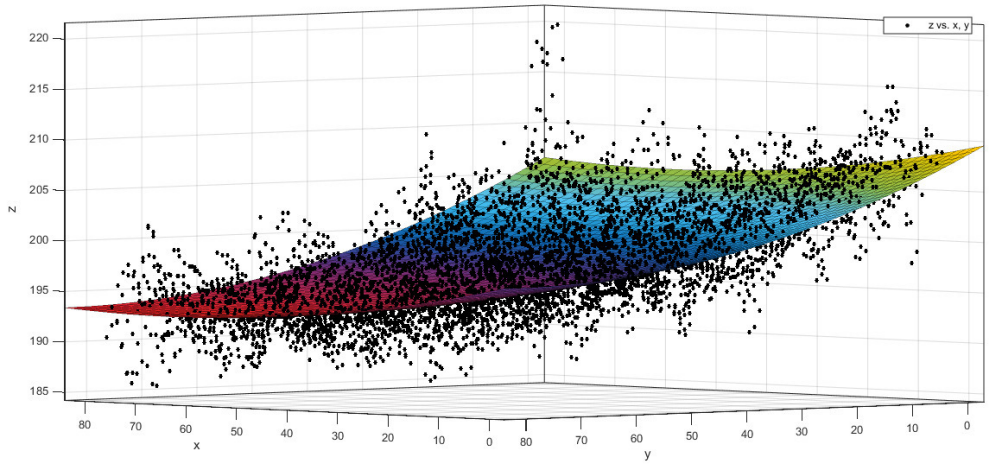


Fig. 2. Plot of bivariate polynomial fitted to data. In this case the dominating diffusion gradient was applied along the y axis with a nonzero x component. The expression of the fitted function is the following: $f(x, y) = 208.4 - 0.1145x - 0.2833y + 0.0009x^2 - 0.0001xy + 0.0017y^2$.

the spatial distribution of the b-matrix elements for each voxel can be derived from Stejskal-Tanner equation [15].

$$\ln\left(\frac{S_x}{S_0}\right) = -\mathbf{b} : \mathbf{D}, \tag{2}$$

where: S_x – is a signal intensity in particular voxel measured with x diffusion gradient applied, S_0 – is a signal intensity in particular voxel measured without diffusion gradient applied, $\mathbf{b}$ – is a b-matrix in particular voxel, $\mathbf{D}$ – is a diffusion tensor in particular voxel. The colon indicates a generalized dot product defined as:

$$\mathbf{b} : \mathbf{D} = \sum_{i,j} b_{ij} D_{ij}, \tag{3}$$

where b_{ij} and D_{ij} are particular b-matrix and diffusion tensor elements, respectively.

The complete set of b-matrices derived from the above formulas consisted of 900 b-maps (25 slices times 6 b-matrix elements times 6 diffusion gradients).

An 80 x 80 pixel square region of interest (ROI) R1 inscribed in the circle of the phantom's axial projection was chosen for the analysis.

Theoretically each element of the b-matrix should be a scalar value derived from the sum of various quadratic gradient terms. The number of terms depends on what is really taken into account in calculating the b-matrix, but the diffusion gradients are always dominating, because of their relatively big strength. The influence of the imaging gradients is of secondary importance. Cross-terms between various types of gradients may also appear. Due to imperfections of the gradients, their nonlinearity and interactions between them the actual b-matrix should be described by a superposition of quadratic functions instead of a single constant value.

The actual character of the b-matrix may be expected to be

TABLE I
COMPARISON OF THE MEAN EIGENVALUES AND THEIR STANDARD DEVIATIONS OBTAINED BY MEANS OF THE THREE APPROACHES. THE IMPROVEMENT FACTOR K_{SD} IS EQUAL TO THE RATIO OF STANDARD DEVIATION OF STANDARD DTI TO THE SD OF A PARTICULAR APPROACH.

Experiment	D [mm^2/s]	sd [mm^2/s]	K_{sd}
DTI	1.9822×10^{-3}	5.73×10^{-5}	1.00
BSD	1.9984×10^{-3}	5.42×10^{-5}	1.05
A-BSD	1.9973×10^{-3}	3.92×10^{-5}	1.46

described by a superposition of quadratic functions instead of a single constant value.

In order to fit a second order bivariate polynomial function:

$$f(x, y) = a_0 + a_1 x + a_2 y + a_3 x^2 + a_4 xy + a_5 y^2, \qquad (4)$$

to all the 900 ROIs, a python script using least-squares fitting procedure was written.

The initial b-matrix values within the ROIs were substituted with values derived from the fitted functions. Eventually, diffusion tensors were calculated in three ways:

1) Standard DTI - using single b-matrix, constant in the entire ROI.
2) BSD - incorporating spatial distribution of the b-matrix, derived directly from the calibration procedure.
3) Actual BSD (A-BSD) - with usage of b-matrix spatial distribution derived from the fitted polynomial. The tensors were calculated for a circle ROI R2 inscribed in the R1 used for the fitting.

III. RESULTS

The calculated diffusion tensors were diagonalized with LU decomposition method in order to obtain the eigenvalues. For each of the 25 slices mean eigenvalue in R2 and its standard deviation were calculated. The results for selected slices together with the improvement factors K_{sd} are presented in table I. Mean eigenvalues are quite similar for all the three approaches, however, standard deviations differ. The lowest value of sd, which indicates the best accuracy, was obtained for the A-BSD approach using the fitting procedure.

Plots of exemplary bivariate polynomial fits obtained for two different gradient directions are depicted in figures 1 and 2, respectively. The first corresponds to the diffusion gradient applied solely in x direction, whereas the latter, to the stronger component in y direction and weaker, nonzero component in x.

Two dimensional maps of the fitted b-matrices for selected slices are depicted in figures 3 and 4.

IV. DISCUSSION AND CONCLUSIONS

The described procedure resulted in a noticeable improvement of the diffusion tensor eigenvalues homogeneity for the isotropic water phantom. The analysis was done in axial direction in which distortions in b-matrices are the smallest, accordingly to the previous studies [13] [16]. Thus also the K_{sd} improvement factor is relatively small. However, fitting the b-matrix data to the bivariate polynomial function decreases the

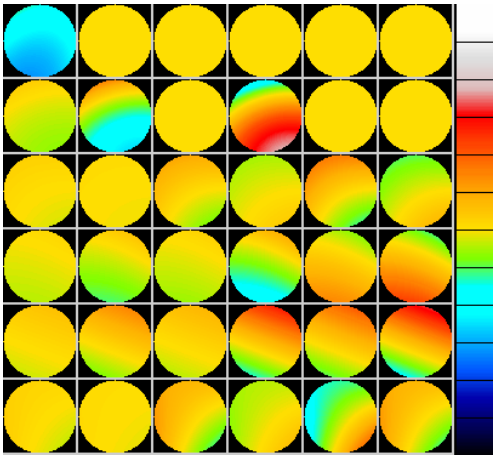


Fig. 3. Two dimensional map of b-matrix elements corrected with fitted bivariate polynomials. Each row represents one of the diffusion gradient directions, and each column is a b-matrix element in order: b_{xx}, b_{yy}, b_{zz}, b_{xy}, b_{xz}, b_{yz}. The maps correspond to the middle slice, positioned in the isocenter of the scanner. The scale ranges from $-50\,s/mm^2$ to $50\,s/mm^2$. The map is differential, i.e. from each element standard, constant b_{ij}-value was subtracted.

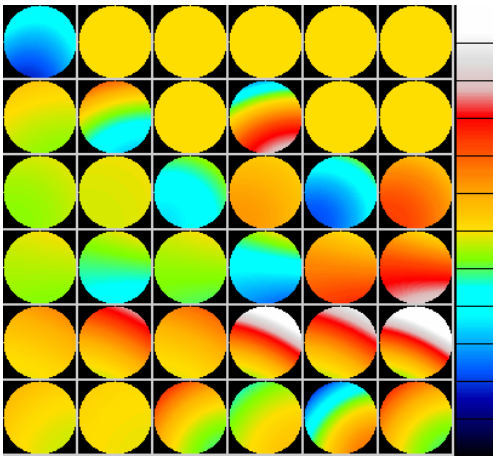


Fig. 4. Two dimensional map of b-matrix elements corrected with fitted bivariate polynomials. Each row represents one of the diffusion gradient directions, and each column is a b-matrix element in order: b_{xx}, b_{yy}, b_{zz}, b_{xy}, b_{xz}, b_{yz}. The maps correspond to the 25 slice, the farthest from the isocenter of the scanner. The scale ranges from $-50\,s/mm^2$ to $50\,s/mm^2$. The map is differential, i.e. from each element standard, constant b_{ij}-value was subtracted.

mean eigenvalue standard deviation to the approximately one third of the sd obtained through hitherto BSD calibration.

The analyzed dataset is characterized by a relatively high noise level, due to only 4 averages. Application of the fitting procedure visibly smooths the b-matrix spatial distribution and emphasizes the shape of the systematic errors. Figures and the equations of the fitted functions show that a_3 coefficient is an order of magnitude bigger than a_5 when the gradient direction was set along x axis and on the contrary the a_5 coefficient is an order of magnitude bigger than a_3 when the gradient is directed mostly along the y axis. The linear coefficients of the fits are also not negligible. It corresponds to the already

discussed in this paper theoretical underpinning of the b-matrix. Complicated, by number of terms which are difficult to take into account in the analytical derivations (including cross-terms), the structure of the b-matrix in the presented approach is revealed in its final shape through an experiment.

The two dimensional maps of the b-matrix elements depict the shape and intensity of the systematic errors. Comparison of figures 3 and 4 confirms the best homogeneity of the b-matrices in the isocenter of the scanner (most of the terms in figure 3 are close to yellow color corresponding to the assumed standard value of b_{ij} element), and intensification of the distortions while moving towards the edges of the scanner (fig 4.).

In conclusion the presented method improves the BSD approach to DTI further. It enables one to derive noise-free spatial maps of b-matrices even from a dataset with relatively high noise level. This gives hope for shortening the BSD calibration time. Second degree bivariate polynomials turn out to define well the b-matrix elements according to the theory.

The presented approach successfully excludes the systematic errors affecting the accuracy of diffusion tensor derivation and opens the path for truly quantitative diffusion tensor measurements independent of a chosen scanner and imaging sequence.

V. Acknowledgments

The work was financed by the National Centre of Research and Development, contract. No. PBS2/A2/16/2013 and contract No. STRATEGMED2/265761/10/NCBR/2015.

K. K. would like to thank the Marian Smoluchowski Cracow Scientific Consortium - KNOW for their support.

References

[1] P. J. Basser, J. Mattiello, and D. LeBihan, "MR diffusion tensor spectroscopy and imaging." *Biophysical Journal*, 1994. [Online]. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1275686/

[2] ——, "Estimation of the effective self-diffusion tensor from the NMR spin echo," *Journal of Magnetic Resonance. Series B*, 1994.

[3] D. Le Bihan and H. Johansen-Berg, "Diffusion MRI at 25: Exploring brain tissue structure and function," *NeuroImage*, 2012. doi: 10.1016/j.neuroimage.2011.11.006. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1053811911012857

[4] J.-D. Tournier, S. Mori, and A. Leemans, "Diffusion tensor imaging and beyond," *Magnetic Resonance in Medicine*, 2011. doi: 10.1002/mrm.22924. [Online]. Available: http://onlinelibrary.wiley.com.atoz.wbg2.bg.agh.edu.pl/doi/10.1002/mrm.22924/abstract

[5] P. J. Basser and D. K. Jones, "Diffusion-tensor MRI: theory, experimental design and data analysis - a technical review," *NMR in Biomedicine*, 2002. doi: 10.1002/nbm.783. [Online]. Available: http://onlinelibrary.wiley.com/doi/10.1002/nbm.783/abstract

[6] M. Neeman, J. P. Freyer, and L. O. Sillerud, "A simple method for obtaining cross-term-free images for diffusion anisotropy studies in NMR microimaging," *Magnetic Resonance in Medicine*, 1991. doi: 10.1002/mrm.1910210117. [Online]. Available: http://onlinelibrary.wiley.com.wiley-online-library.wbg2.bg.agh.edu.pl/doi/10.1002/mrm.1910210117/abstract

[7] A. L. Alexander, J. S. Tsuruda, and D. L. Parker, "Elimination of eddy current artifacts in diffusion-weighted echo-planar images: The use of bipolar gradients," *Magnetic Resonance in Medicine*, 1997. doi: 10.1002/mrm.1910380623. [Online]. Available: http://onlinelibrary.wiley.com.wiley-online-library.wbg2.bg.agh.edu.pl/doi/10.1002/mrm.1910380623/abstract

[8] K. M. Hasan, D. L. Parker, and A. L. Alexander, "Comparison of gradient encoding schemes for diffusion-tensor MRI," *Journal of Magnetic Resonance Imaging*, 2001. doi: 10.1002/jmri.1107. [Online]. Available: http://onlinelibrary.wiley.com.atoz.wbg2.bg.agh.edu.pl/doi/10.1002/jmri.1107/abstract

[9] D. K. Jones, "The effect of gradient sampling schemes on measures derived from diffusion tensor MRI: A Monte Carlo study," *Magnetic Resonance in Medicine*, 2004. doi: 10.1002/mrm.20033. [Online]. Available: http://onlinelibrary.wiley.com.atoz.wbg2.bg.agh.edu.pl/doi/10.1002/mrm.20033/abstract

[10] A. Özcan, "(Mathematical) Necessary conditions for the selection of gradient vectors in DTI," *Journal of Magnetic Resonance*, 2005. doi: 10.1016/j.jmr.2004.10.013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1090780704003544

[11] M. E. Bastin, "Correction of eddy current-induced artefacts in diffusion tensor imaging using iterative cross-correlation," *Magnetic Resonance Imaging*, 1999. doi: 10.1016/S0730-725X(99)00026-0. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0730725X99000260

[12] M. A. Horsfield, "Mapping eddy current induced fields for the correction of diffusion-weighted echo planar images," *Magnetic Resonance Imaging*, 1999. doi: 10.1016/S0730-725X(99)00077-6. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0730725X99000776

[13] A. T. Krzyżak and Z. Olejniczak, "Improving the accuracy of PGSE DTI experiments using the spatial distribution of b matrix," *Magnetic Resonance Imaging*, 2015. doi: 10.1016/j.mri.2014.10.007. [Online]. Available: http://www.mrijournal.com/article/S0730725X1400318X/abstract

[14] A. Einstein and R. Fürth, *Investigations on the theory of Brownian movement*, New York, N.Y., 1956.

[15] E. O. Stejskal and J. E. Tanner, "Spin Diffusion Measurements: Spin Echoes in the Presence of a Time-Dependent Field Gradient," *The Journal of Chemical Physics*, 1965. doi: 10.1063/1.1695690. [Online]. Available: http://scitation.aip.org/content/aip/journal/jcp/42/1/10.1063/1.1695690

[16] K. Kłodowski and A. T. Krzyżak, "Innovative anisotropic phantoms for calibration of diffusion tensor imaging sequences," *Magnetic Resonance Imaging*, 2016. doi: 10.1016/j.mri.2015.12.010. [Online]. Available: http://www.mrijournal.com/article/S0730725X15003057/abstract

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 947–950
DOI: 10.15439/2016F265
ACSIS, Vol. 8. ISSN 2300-5963

Multilayer perceptron for gait type classification based on inertial sensors data

Damian Szczepański
Rzeszów University of Technology al. Powstancow Warszawy 12, 35-959 Rzeszów, Poland
Email: dszczepanski@prz.edu.pl

***Abstract*—In this study, the gait type classification process is considered. Input data are obtained by using the inertial sensors tracking system. Three types of gait are recorded: normal walk, tiptoeing and walk retaining long stance phase. Two data set types, describing the registered motion, are prepared. The most significant input features are selected by means of the sensitivity analysis (SA) procedure. The classification process is conducted using multilayer perceptron (MLP) with various structures. The classification accuracy of the network is computed with the use of a cross validation procedure. The obtained results show that the successful classification of presented gait types can be achieved using relatively simple MLP architecture.**

I. Introduction

NOWADAYS, as the society is getting older, the illnesses typical for the elderly age, are becoming more and more meaningful part of modern medicine. In turn, the doctors need immediate and accurate methods in everyday diagnosing. The analysis of human gait from the very beginning has been a basic method in rehabilitation and early locomotive disease detection. Nowadays, Parkinson's disease (PD) is most common and onerous dysfunction of medicine [1]. At the early stage, PD does not show any significant symptoms, except for small gait and posture changes. Typical gait modification relies on the increasing foot contact in stance phase (shuffling steps), forward-flexed posture and limbs tremor. In order to treat PD effectively, the rehabilitation and exercises should be employed at the very early stage of disease detection. The technology progress in human body motion capture provides new solutions which could be used in human gait and posture analyses.

Since the beginning, the optical systems have been the basic tool in human motion capture. However, due to its complicity, their usage can be very complicated on a daily basis. One of the most promising approaches is the inertial sensors technology, which is simpler in usage and comparable in accuracy with the optical systems . Inertial systems have been used in many fields of medical diagnose assistance[2], i.e.: cerebral palsy [3], [4], determining gait defects, sport medicine [5]. Flexibility advantage of such system is very useful in rehabilitation [6], exercise motivation [7] and automatic exercise advising [8].

The main purpose of this article is to explore the classification problem of gait which is registered on the basis of the motion parameters such as joint angles, segments velocity and accelerations values. The motion acquisition is conducted by using Xsens MVN Biomech suit (MVN) [9] which is based on inertial measure units (IMU) wirelessly connected to the computer. In the classification process, the MLP classifier is used and the best architecture with highest accuracy is chosen. In current study, one also considers the problem of input features' selection. Therefore, the SA procedure is utilized to determine the most significant gait attributes. MLP classification performance is assessed with reduced number of inputs.

All simulations, the classification process and the IMU synthesis are conducted by using Matlab [10], DTReg [11] and MVN Studio [12] software.

The paper is composed of the following sections. Section II presents the input data used in this work. In Section III, the problem statement is formulated and the approaches utilized in the classification process are shortly described. Sections IV and V outline the experiments' settings and the obtained results, respectively. Section VI concludes the paper.

II. Input data

A. Data acquisition

The registration process is conducted by using MVN which utilizes 17 inertial sensors attached to each body segment (see Fig. 1).

Two subjects (M: 90kg, 180 cm, 29 years, M: 75 kg 181 cm, 26 years) take part in the experiment. They are asked to walk in three different gait manners: (a) – normal walk with natural speed, (b) – tiptoeing and (c) – walk retaining long stance phase and very small distance between foot and ground in swing phase. The (c) type of gait is similar to PD walk. Registration is repeated 20 times per gait type by each person, which gives 120 data series for all gait types.

One gait cycle, which is one step, is defined as a time between heel ground contacts. Fig. 2 presents a joint angle registered within a single gait cycle.

B. Data type

Raw data (accelerometer, gyroscope, magnetometer values), which are acquired directly from IMU sensors during registration process, are synthesized by MVN Studio Software. All joint angles, segment accelerations, segment velocity, angular segment accelerations and angular segment velocity are calculated. Each parameter has three values in each dimension what, in turn, gives 612 values per one time frame data. Sample screen-shot from MVN Studio with rendered subject avatar and plotted joint angles are presented in Fig. 3.

978-83-60810-90-3/$25.00

Fig. 1. Inertial sensors attached to body segments.

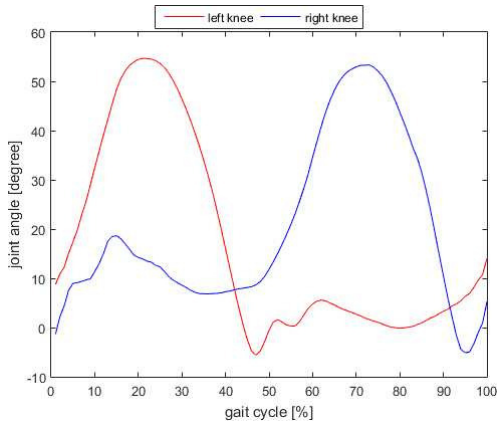


Fig. 2. Flexion - extension knee angle in one gait cycle: left knee (red), right knee (blue).

C. Input vectors

For the purpose of the comparison, two data sets are prepared from the gait recordings. The first data set is represented by statistical values (SV): maximum value, minimum value, mean and standard deviation, which are calculated from acceleration values of 17 body segments. Every IMU accelerometer measure signal in *x, y, z* axis what in result gives 204 values (4 statistical values of 17 segments in 3 axis). These are: pelvis, thoracic spine, head, right shoulder, right upper arm, right forearm, right hand, left shoulder, left upper arm, left forearm, left hand, right upper leg, right lower leg, right foot, left upper leg, left lower leg, and left foot. An exemplary single body segment, i.e. left upper leg, is illustrated in Fig. 3 (red curve).

The second data set is represented by the normalcy index (NI) parameters commonly used in gait analyses [13]. The 15 values proposed in [14] are determined: time of toe off, walking speed, mean pelvic tilt, range of pelvic tilt, mean pelvic rotation, minimum hip flexion, range of hip flexion, peak abduction in swing, mean hip rotation in stance, knee flexion at initial contact, time of peak knee flexion, range of knee flexion, peak dorsiflexion in stance phase, peak dorsiflexion in swing, and mean foot progression angle. All calculations are conducted in Matlab software.

Three types of gait described in subsection II-A impose three class classification problem.

III. Application of MLP network to gait classification problem

A. Problem statement

For the input data set described in Section II, the task is to find the optimal MLP architecture with respect to the highest prediction accuracy (*Acc*) of the model. The accuracy is calculated as a sum of true positives and true negatives divided by number of all input vectors. The experiment is repeated 10 times and after all tests, the average value and the standard deviation (*Std*) of *Acc* are calculated. The accuracy in each test trial is computed using a 10-fold cross validation procedure. Furthermore, an optimal subset of input features is found with the use of the SA procedure. As in the case of all input attributes, the highest *Acc* of MLP is determined for reduced data set.

B. Multilayer perceptron

MLP is the feedforward neural network [15], [16], i.e. the model where the input signal is fed forward through a number of layers [17]. There are three types of layers in MLP: an input layer, one or more hidden layers and an output layer. The input layer consists of the nodes which are the features of an input vector. Each of $i = 1, \ldots, n$ input features is connected to a hidden layer neuron activated as follows

$$y_j^{(1)} = f\left(\sum_{i=1}^{n} w_{ji}^{(1)} x_i + b_j^{(1)}\right), \tag{1}$$

where $f(\cdot)$ is the activation function, $\mathbf{w}_j^{(1)} = [w_{j1}^{(1)}, \ldots, w_{jn}^{(1)}]$ is the weight vector of the jh hidden neuron, $\mathbf{x} = [x_1, \ldots, x_n]$ is an input vector and $b_j^{(1)}$ denotes the bias of the jh hidden neuron. The outputs $y_j^{(1)}$ of the hidden neurons are linearly combined with the second layer weights $w_{kj}^{(2)}$ and the bias $b_k^{(2)}$ and, after activation, create an output of the kth neuron in the second layer

$$y_k^{(2)} = f\left(\sum_{j=1}^{m} w_{kj}^{(2)} y_j^{(1)} + b_k^{(2)}\right). \tag{2}$$

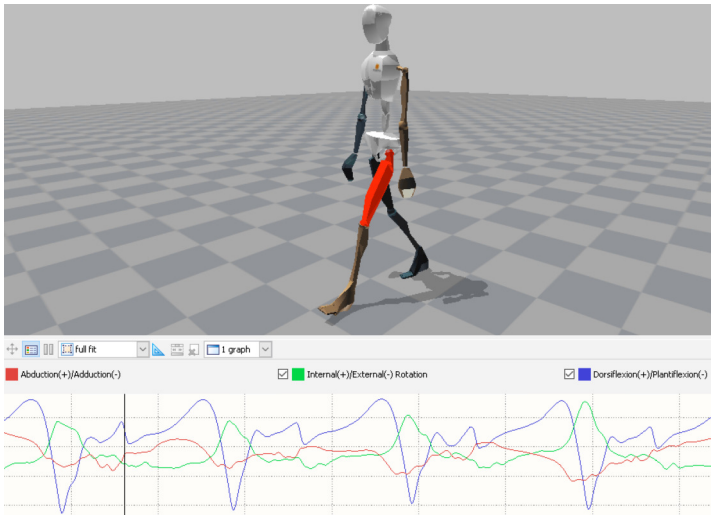


Fig. 3. MVN Studio screen-shot presenting rendered avatar with selected upper leg segment and plotted knee angles.

In (1) and (2), $f(\cdot)$ should be selected in the way it is continuous, differentiable and monotonically increasing. $f(a) = \frac{1}{1+e^{-\lambda a}}$ is the most commonly used activation function. The number of hidden layers and the number of neurons in each hidden layer must be optimized in order to maximize MLP's prediction accuracy. In this study, scaled conjugate gradient algorithm is used for MLP training.

C. Sensitivity analysis procedure

SA procedure steps separately through a single feature in each iteration. It randomly permutes its values across the rows of the data set. The values of this feature remain the same but they are randomly moved to different rows. Then the accuracy the network is evaluated on the modified data set. If the feature is important, randomizing the order of its values greatly degrades the accuracy of the predictions. If the feature is not valuable in prediction, the order of its values has little or no influence on the accuracy of the network. Last step is to determine the rank of the importance of features based on the amount of degradation that occurred by randomizing their values and to scale the scores so the most important feature has a relative importance of 100.

IV. Experiments

MLP used in experiment has 3 outputs (one for each gait type). The number of neurons in the hidden layer is taken from the set $\{2, \ldots, 10\}$. The linear, logistic and tangent activation functions are utilized for MLP training. In the first part of the experiment, MLP is trained by using all input variables both for SV and NI data. As a result, the best network architectures for both data sets are chosen. The second part of the experiments relies on applying the SA procedure to select features for input neurons' representation. Table I and II present the importance of features for NI and SV data set, respectively. Due to the fact that SV data vector is composed of 204 attributes, only 7 significant features are presented. For the second part of the experiment, only 3 most significant features of SV and NI dataset are selected. The classification process is repeated using reduced MLP and best network architectures are chosen.

TABLE I
First 7 most important NI data set features.

Parameter	Importance
peak body dorsiflexion in stance phase	100.00
knee range flexion	61.90
peak body dorsiflexion in swing phase	48.87
range of hip flexion	9.61
mean foot progression angle	7,28
walking speed	5,45
mean hip rotation in stance	3,21

TABLE II
First 7 most important SV data set features.

Parameter	Importance
mean right foot acceleration in Z axis	100.00
mean acceleration of right thigh in Y axis	69,79
mean left foot acceleration in Z axis	69,75
mean right foot acceleration in X axis	48,72
mean left foot acceleration in Y axis	38,11
mean left foot acceleration in X axis	35,08
standard deviation of pelvis tilt acceleration in Z axis	18,75

V. Results

A. Result for SV dataset

For MLP with 204 inputs, the highest accuracy (99.18%) is achieved for the network with 2 hidden neurons. Using only mean right foot acceleration in Z axis (importance 100), mean acceleration of right thigh in Y axis (importance 69.79), mean acceleration of left foot in Z axis (importance 69.74),

TABLE III
RESULTS OF SV DATASET CLASSIFICATION.

Used parameter	Neurons	*Acc*	*Std*
all parameters	2	99.18%	0.11%
mean right foot acceler. in Z axis, mean acceler. of right thigh in Y axis, mean left foot acceler. in Z axis	2	100.00%	0.00%
mean right foot acceler. in Z axis	2	100.00%	0.00%
mean acceler. of right thigh in Y axis	3	91.14%	0.93%
mean left foot acceler. in Z axis	9	94.59%	1.51%

TABLE IV
RESULTS OF NI DATASET CLASSIFICATION.

Used parameter	Neurons	*Acc*	*Std*
all parameters	2	100.00%	0.00%
peak body dorsiflexion in stance phase, knee range flexion, peak body dorsiflexion in swing phase	2	99.92%	0.26%
peak body dorsiflexion in stance phase	9	81.91%	3.25%
peak body dorsiflexion in swing phase	4	88.71%	0.72%
knee range flexion	3	91.40%	1.16%

$Acc = 100\%$ for the model with 2 hidden neurons. Having run the classifications process with a single parameter (mean right foot acceleration in Z axis) for single layer MLP, 100% accuracy is also obtained. Interestingly, $Acc = 91.14\%$ when using only mean acceleration right thigh in Y axis parameter with 3 neurons in hidden layer. Utilizing only the mean acceleration left foot in Z axis parameter, 94.59% accuracy is achieved for MLP with 9 hidden neurons.

B. Result for NI dataset

The classification with all 15 features allows MLP with 2 hidden neurons to achieve 100% accuracy. When using 3 most important features selected by the SA procedure, i.e.: peak body dorsiflexion in stance phase (100 importance), knee range flexion (61.90 importance), peak body dorsiflexion in swing phase (48.87 importance), $Acc = 99.92\%$ for MLP with 2 neurons in hidden layer. By employing only one parameter (peak body dorsiflexion in stance phase), 9 hidden neurons' MLP yields 81.91% of accuracy. Application of the knee flexion parameter allows MLP with 3 hidden neurons to reach $Acc = 91.40\%$. Consequently, the use of 4 neurons in hidden layer of the network with a single input parameter (peak body dorsiflexion in swing phase) provides the accuracy of 88.71%.

VI. CONCLUSION

In this article, the gait classification problem was studied. Three types of gait were registered: normal walk, tiptoeing and walk retaining long stance phase. For the purpose of the comparison, two data sets were created from the gait recordings. The first data set represented the statistical values calculated from 17 body segments. The second data set was composed of the NI parameters, which are commonly used in gait analyses. Furthermore, the SA procedure was applied to select the most important features in the data sets. The classification tasks were performed by MLP for both original, and reduced data. In all cases, the prediction accuracy achieved very high values at relatively simple architecture of MLP. Moreover, in some cases, decreasing the number of neurons in the hidden layer contributed to the increase of the accuracy.

The results presented in this work encourage the author to explore gait classification problem deeper. For this purpose, real gait recordings will registered from the patients in near future.

REFERENCES

[1] Statistics on parkinson's. Accessed: 2016-05-01. [Online]. Available: http://www.pdf.org/en/parkinson_statistics

[2] A. Ferrari, A. G. Cutti, P. Garofalo, M. Raggi, M. Heijboer, A. Cappello, and A. Davalli, "First in vivo assessment of "outwalk": a novel protocol for clinical gait analysis based on inertial and magnetic sensors," *Medical & biological engineering & computing*, vol. 48, no. 1, pp. 1–15, 2010.

[3] A. Ferrari, "Technical innovations for the diagnosis and the rehabilitation of motor and perceptive impairments of the child with cerebral palsy," 2010.

[4] A. Szopa, M. Domagalska-Szopa, Z. Kidoń, and M. Syczewska, "Quadriceps femoris spasticity in children with cerebral palsy: measurement with the pendulum test and relationship with gait abnormalities," *Journal of neuroengineering and rehabilitation*, vol. 11, no. 1, p. 1, 2014.

[5] A. Martinez-Ramirez, P. Lecumberri, M. Gomez, and M. Izquierdo, "Wavelet analysis based on time–frequency information discriminate chronic ankle instability," *Clinical Biomechanics*, vol. 25, no. 3, pp. 256–264, 2010.

[6] H.-P. Brückner, W. Theimer, and H. Blume, "Real-time low latency movement sonification in stroke rehabilitation based on a mobile platform," in *Consumer electronics (ICCE), 2014 IEEE international conference on*, 2014, pp. 242–243.

[7] D. Morris, T. S. Saponas, A. Guillory, and I. Kelner, "Recofit: using a wearable sensor to find, recognize, and count repetitive exercises," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2014, pp. 3225–3234.

[8] A. W. Lam, A. HajYasien, and D. Kulic, "Improving rehabilitation exercise performance through visual guidance," in *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE*. IEEE, 2014, pp. 1735–1738.

[9] Mvn biomech. Accessed: 2016-05-01. [Online]. Available: http://www.xsens.com/products/mvn-biomech

[10] Matlab. Accessed: 2016-05-01. [Online]. Available: http://www.mathworks.com/products/matlab/

[11] Dtreg predictive modeling software. Accessed: 2016-05-01. [Online]. Available: http://www.dtreg.com

[12] Mvn studio software. Accessed: 2016-05-01. [Online]. Available: http://www.xsens.com/mvn-studio-pro

[13] M. Romei, M. Galli, F. Motta, M. Schwartz, and M. Crivellini, "Use of the normalcy index for the evaluation of gait pathology," *Gait & posture*, vol. 19, no. 1, pp. 85–90, 2004.

[14] L. Schutte, U. Narayanan, J. Stout, P. Selber, J. Gage, and M. Schwartz, "An index for quantifying deviations from normal gait," *Gait & posture*, vol. 11, no. 1, pp. 25–31, 2000.

[15] R. Tadeusiewicz, *Sieci neuronowe*. Akademicka Oficyna Wydawnicza Warszawa, 1993.

[16] J. Korbicz, A. Obuchowicz, and D. Uciński, "Artificial neural networks," *AOW PLJ, Warsaw*, 1994.

[17] D. E. Rumelhart, J. L. McClelland, and C. PDP Research Group, Eds., *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 1: Foundations*. Cambridge, MA, USA: MIT Press, 1986.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 951–954
DOI: 10.15439/2016F225
ACSIS, Vol. 8. ISSN 2300-5963

Mathematical model of the VAG gas valve identification algorithms

Adam Trojnar, MSc. Eng.
Institute of Applied Computer Science
Lodz University of Technology
ul. Stefanowskiego 18/22, 90-924 Lodz, Poland
Email: atrojnar@iis.p.lodz.pl

Piotr Ostalczyk, DSc. Eng.
Institute of Applied Computer Science
Lodz University of Technology
ul. Stefanowskiego 18/22, 90-924 Lodz, Poland
Email: piotr.ostalczyk@p.lodz.pl

***Abstract*—This paper present identification algorithms of the VAG gas valve and a comparison between them. One of the mathematical models is based on the differential-integral cal-culus of fractional order. This mathematical tool can be used for modeling devices serving to closed-loop systems synthesis. Admitting fractional orders in difference equations improves the performance of proportional-integral-derivative controller action.**

Index Terms—gas valve, differential-integral cal-culus of fractional order, identification.

I. Introduction

THE AUTOMATION determines a comfort of use and operating costs of the heating system, which is responsible for controlling its operation [1, 2, 3]. In discussed case the VC200 control unit is responsible for it, which along with the VAG gas valve constitutes a set to control the operation of heating boilers. This control unit is designed to maintain the degree of the valve opening based on measured temperature parameters and set temperature by the user. The valve offers linear and constant characteristics of operation of a gas burner to heating boiler with standby position. This solution causes a measurable reduction in gas consumption in older type of systems, elimination of incomplete burning and condensation of the exhaust fumes in emittors.

Many unpredictable factors must be taken into account in design of control systems. Application of coil and steel core in the VAG valve causes non-linear dependence of the degree of opening to current amplitude and hysteresis of valve motion up and down. In case of control with using information about the model, there is a possibility to remove uncontrolled disruptions. In PID type of control units the signal that controls the device is a sum of three functional blocks: multiplication, integration and differentiation. Difference, which appears between the set and currently measured value, is a result of specified operations. It is possible in more accurate way to describe real objects using differential equations of fractional order, and therefore application of fractional order control units in such cases is more beneficial [1, 2, 3].

II. Identification algorithms

In a room was placed the test bench in order to study static and dynamic characteristics of the control system temperature. Setting parameters of the control unit operation is carried out by RS232 communication channel connected to the computer. The computer has installed the Windows operating system that is required to run the DIAG200 program. The test bench

Fig. 1. The test bench for measuring the VAG gas valve.

consists of: 1. the VAG gas valve, 2.the VC200 controller, 3. a computer with the Windows OS, 4. a injector air nozzles, 5. a gas cylinder LPG.

Description of the black box which is a linear system having one input and output is described by a differential equations [3, 4]. In this case, time is a independent variable and binding input values and the observed response. Measured system is described by the following differential equation:

$$\frac{dx(t)}{dt} + a_0 x(t) = b_0 \frac{du(t)}{d(t)} \tag{1}$$

978-83-60810-90-3/$25.00

the solution of the Equation (1) is:

$$x(t) = ae^{-bt} \tag{2}$$

The difference backward v-th row is defined as:

$$_{k_0}\Delta_k^{(v)} x(k) = \sum_{i=k_0}^{k} a_{i-k_0}^{v} f(k + k_0 - i) \tag{3}$$

another form of the difference backward:

$$_{k_0}\Delta_k^{(v)} x(k) = \begin{bmatrix} a_0^v & a_1^v & \cdots & a_{k-k_0}^v \end{bmatrix} \begin{bmatrix} x(k) \\ x(k-1) \\ \vdots \\ x(k) \end{bmatrix} \tag{4}$$

To create the differential equation is applied the backward difference of first order $x(k)$ function:

$$_0\Delta_k^{(1)} x(k) + a_0 x(k) = b_0 u(k) \tag{5}$$

$$x(k) - x[(k-1)h] + a_0 x(k) = b_0 \tag{6}$$

$$x(k) = \frac{1}{1+a_0}[x(k-1)] + \frac{1}{1+a_0} \tag{7}$$

A. Static model

Static model can be described in a form of graph that presents static characteristics of studied object. The graph of this type is not dependent on the time. Static characteristics constitute a material for system identification; it is a relation between the input and output signal.

The controller uploads current temperature in the room from a sensor and according to parameters set by the program is controlled by a needle in the valve. This type of regulation enables to supply of planned quantity of gas to the injector air nozzles, and depending on demand to increase or reduce the temperature in examination room.

Study consisted on leading into the control system, unchanging in time, subsequent values of the input signal, and on measurement of corresponding values of the output signal. As a result of conducted experiment related to determination of the static characteristics, discreet values were obtained and they were marked on a graph (Fig. 2). Obtained graph of static characteristics shows the position of valve in relation to output voltage of the control unit.

The second basic aim in determination of static characteristics is to appoint characteristics equation from the graph (Fig. 2). Using computer synthesis methods and analyses there was a few methods to resolve the task. For the valve an approximation was used with method of the smallest squares. In order to obtain the static characterization of object, an approximation of received extortion values was conducted and corresponding responses on the graph (red line in Fig. 2). Below is presented third degree polynomial that approximates results of conducted measurement:

$$w(x) = -x^3 + 0,0005x^2 + 0,0322x + 15,579 \tag{8}$$

In order to describe appearing phenomena and processes there are used mathematical models. Complexity result of identification depends on its destination. It is possible to describe its correct construction by mapped accuracy of the process flow with reference to applied model. On the one hand, too high complexity of the model structure may require a lot of calculation by which the model may turn out to be useless. However, on the other hand, the simplified model may have large variations in relation to mapped reality [5].

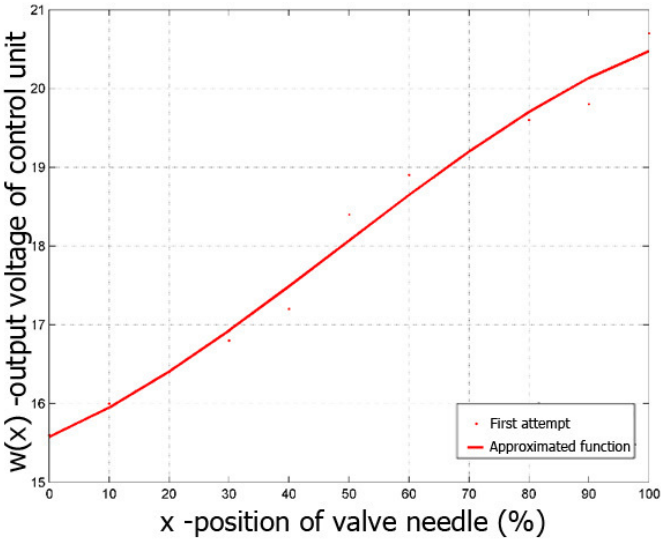


Fig. 2. Graph of static characteristics shows the position of valve in relation to output voltage of the control unit.

B. Dynamic models

An important aspect is identification of dynamic properties of the object. Observation of the process response to stimulation enables to obtain information about its description in the field of variable time. Time line of the output rate triggered by extortion is called dynamic characterization of the object. Dynamic characterization is a transmittance between two established states. Method of dynamic model identification in case of valve consists in evaluation of the object transmittance based on variable characterization. The object of regulation is a movable control needle fixed on the electromagnet core in the VAG valve. As the variable characterization of valve a percentage relation of needle fall was accepted, when set temperature and measured temperature starts to reach standby position.

Temperature in the room, which is the object of regulation, at the time of measurement was 17°C. On the controller was set temperature 23.5°C. After reaching required temperature the valve began to limit supply of gas to the burner by linear needle fall. Figure shows the moment of needle fall to full opening to a minimal position. Duration of a single experiment was 33s, results of sampling every 1s. were presented in points. Figure presents three selected sample measurements, differences between individual measurements result from carrying out examining within the entire day. As the variable characterization of valve a percentage relation of needle fall was accepted.

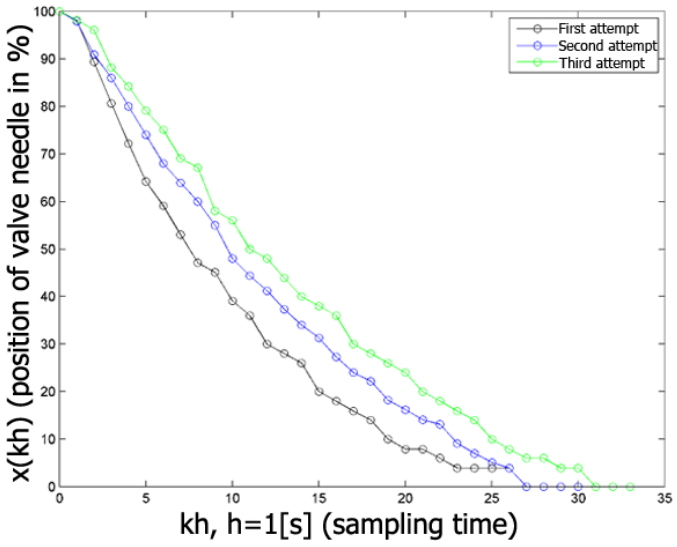


Fig. 3. Measurement of the valve responses to reduce the temperature.

Based on the Fig. 3 it is possible to state that it is the inert object of first order. Inert response type of regulation object to irregular extortion results from the process occurring during the study. Extortion in reductions of opening degree of the valve causes a change of gas supply flow. Reduced power of injector air nozzles resulting for this reason proceeds with certain delay. Other processes are also delayed: heat transfer between the burner and examination room via air and heat transport from surrounding to the temperature sensor. After leveling of new heat loss value in the room with the amount of heat provided by the burner arise a new steady-state and air temperature, and a level of the needle position remains unchanged.

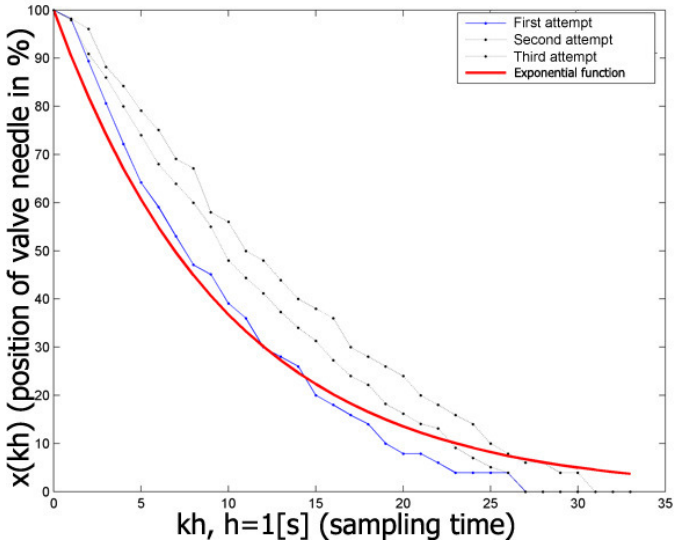


Fig. 4. Exponential function approximate results of the first measurement.

In the measured discreet system described by Equation (2) where continuous time t must be replaced by the value k which is next step, the h value is a sampling time and equal 1 second. On obtained values was approximated exponential function:

$$x(kh) = ae^{-bkh} \tag{9}$$

To the subsequent measurements was matched functions. For the first attempt of response was approximated and illustrated in Fig. 4 exponential function: $x(k) = ae^{\frac{-k}{10}}$, the second attempt of response: $x(k) = ae^{\frac{-k}{12}}$, the third attempt of response: $x(k) = ae^{\frac{-k}{14}}$.

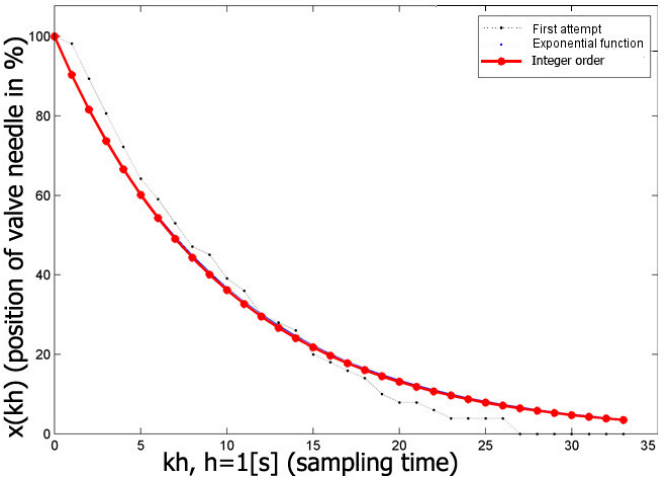


Fig. 5. Function is based on differential equation of the integer order which aproximate the results of the first test measurements.

The second dynamic model is described by a discrete transfer function and the state-space equations [3]. Equation (1) is approximated by the backward difference of first order:

$$\frac{\Delta x(kh)}{h} + a_0 x(kh) = b_0 \frac{\Delta u(kh)}{h} \tag{10}$$

$$\Delta x(kh) + a_0 h x(kh) = b_0 \Delta u(kh) \tag{11}$$

the difference of function $f(kh)$ is:

$$\Delta f(kh) = f(kh) - f[(k-1)h] \tag{12}$$

the difference of function (12) is substituted to the backward difference of first order (11). Based on fact that the measured object got the Dirac delta function extortion it follows $\Delta u(kh)$ is replaced by $\delta(kh)$:

$$x(kh) - x[(k-1)h]a_0 h x(kh) = b_0 \delta(kh) \tag{13}$$

$$(1 + a_0)x(kh) = x[(k-1)h] + b_0 \delta(kh) \tag{14}$$

The solution of the approximation which is based on the backward difference of first order is:

$$x(kh) = \frac{1}{1+a_0} x[(k-1)h] + \frac{b_0}{1+a_0} \delta(kh) \tag{15}$$

The next step is to adjust the parameters a_0, b_0 in order to reproduce the response. The approximate function is based

on the differential equation of integer order (Fig. 6) where: order of differential equation for all attempts $v = 1$ and rate b_0 of differential equation is equal v . Rates of differential equation for the first attempt: $a_0 = 0,107$; for the second attempt $a_0 = 0,085$; for the third attempt $a_0 = 0,073$.

The third dynamic model is also described by a discrete transfer function and the state-space equations. The differ-integral of fractional order is a generalization of the calculus[3]. Equation (1) is approximated by the backward difference of fractional order:

$$\frac{\Delta^v x(kh)}{h} + a_0 x(kh) = b_0 \frac{\Delta u(kh)}{h} \tag{16}$$

The solution of the approximation which is based on the backward difference of fractional order is:

$$\begin{bmatrix} 1 & a_1^v & \cdots & a_k^v \end{bmatrix} \begin{bmatrix} x(kh) \\ x(k-1)h \\ \vdots \\ x(h) \end{bmatrix} + a_0 x(kh) = b_0 \delta(kh) \tag{17}$$

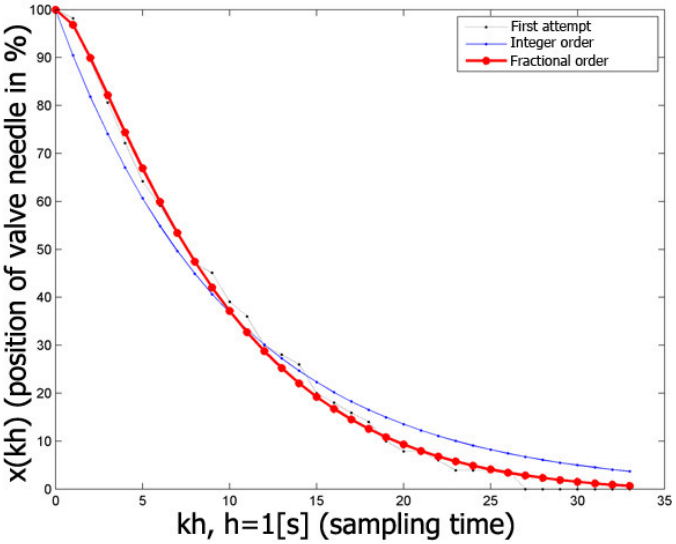


Fig. 6. Function based on the differential equation of fractional order.

The next step is to adjust the parameters a_0, b_0, v in order to reproduce the response. Function based on differential equation of fractional order (Fig. 6), where: rate b_0 of differential equation for all attempts is equal v . Order of differential equation for the first attempt $v = 1,08$, rates of differential equation: $a_0 = 0,115$; Order of differential equation for the second attempt $v = 1,08$, rates of differential equation: $a_0 = 0,092$; Order of differential equation for the third attempt $v = 1,08$, rates of differential equation: $a_0 = 0,073$.

In order to exclude the role of subjective factors in assessing the accuracy of model, as a basis it is necessary to adopt measurable criterion. It is possible to evaluate the level value of control system using integral rates of the course of regulated size. In case of conducted study of valve response, the model accuracy consists on comparison to what extend modeled functions correspond to real measurements. As the correctness criterion of model selection a value of the integral was chosen from the Integral Square Error between measured values and mathematical model [5].

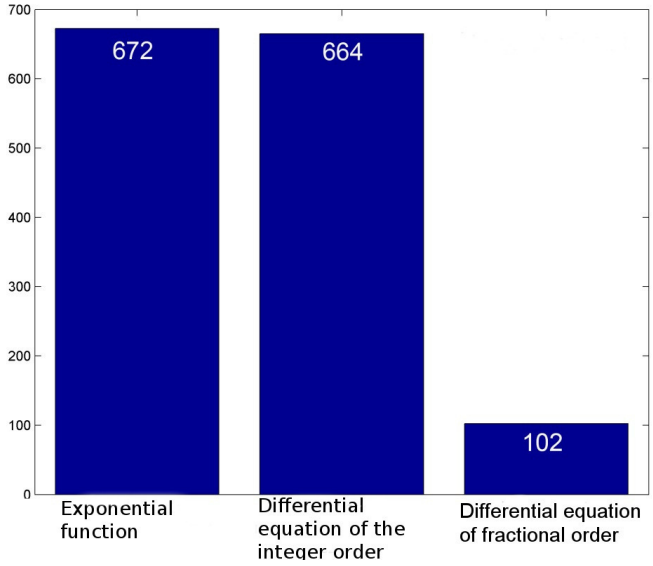


Fig. 7. Comparision a quality of the identification models.

Based on above graphs it is possible to conclude that in any case the best response of device reflects a model based on differential equation of fractional order. The model that reaches the best results in each comparison is several times better than other. Models based on exponential function and differential equation of first order reaches similar degree of approximation.

REFERENCES

[1] YaLi He, RuiKun Gong, *Application of Fractional-order Model Reference Adaptive Control on Industry Boiler Burning System.* Intelligent Computation Technology and Automation (ICICTA), 2010.

[2] Piotr Duch, *Optymalizacja algorytmów numerycznych wykorzystujących równania różnicowe całkowitych i niecałkowitych rzędów,* Praca doktorska, Politechnika Łódzka, 2014, pp 589 - 594 (in Polish).

[3] Piotr Ostalczyk, *Zarys rachunku różniczkowo-całkowego ułamkowych rzędów. Teoria i zastosowania w automatyce,* Wydawnictwo Politechniki Łódzkiej, 2008, pp 1 - 25, 161 - 163, 199 - 209 (in Polish).

[4] Piotr Ostalczyk, *Variable-, fractional-order discrete PID controllers,* Methods and Models in Automation and Robotics (MMAR), 17th International Conference, 2012, pp 534 - 539.

[5] Piotr Ostalczyk, Piotr Duch *Closed — Loop system synthesis with the variable-, fractional — Order PID controller,* Methods and Models in Automation and Robotics (MMAR), 17th International Conference, 2012, pp 589 - 594.

[6] Richard Banach, Pieter Van Schaik, Eric Verhulst *Simulation and Formal Modelling of Yaw Control in a Drive-by-Wire Application,* Computer Science and Information Systems (FedCSIS) 2015, pp 731 - 742.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 955–958
DOI: 10.15439/2016F458
ACSIS, Vol. 8. ISSN 2300-5963

Determination of the quality of results obtained by various numerical methods for BSD

Artur Krzyżak, Piotr Łukasik, Krzysztof Janc
AGH University of Science and Technology, Faculty of Geology,
Geophysics and Environmental Protection,
A. Mickiewicza 30, 30-059 Kraków, Poland
Email: akrzyzak@agh.edu.pl, pioluk@student.agh.edu.pl, krzysztof.janc@gmail.com

***Abstract*—This article describes the determination of the quality of results obtained by various numerical methods for BSD (B-matrix Spatial Distribution). In order to verify the influence of the numerical error on the real data, two datasets acquired using two types of phantoms (isotropic and anisotropic) and the reference random data were analyzed. Additionally examined aspect was the duration of the calculations. The research were carried out on six of numerical methods for solving systems of equations (Gauss, Gauss Jordan, LU, Gauss with partial pivoting, LU (numerical recipes), Gauss-Jordan (numerical recipes)).**

I. Introduction

DIFFUSION Tensor Imaging (DTI) is one the of the most widely used methods of imaging the anisotropic biological structures such as white and grey matter andskeletal muscles. DTI is an imaging technique basing on the phenomenon of nuclear magnetic resonance (NMR), which allows for measurement of diffusion coefficient and its direction, expressed in the form of diffusion tensor. Scientific reports present a broad range of clinical applications of the described method e.g. nerve fibers (tractography), diagnosis of cerebral ischemia, multiple sclerosis, epilepsy, metabolic disorders, tumors of the brain [1-3], and studies on the brain function [4].

Diffusion tensor is a symmetrical 3x3 matrix containing six independent elements. To derive all elements of the diffusion tensor one reference and 6 diffusion weighted images (obtained with various non collinear orientations of a diffusion gradient vector) are required. The relations between the signal and diffusion tensor is described by Stejskal-Tanner equation [6]:

$$\ln\left(\frac{S_n}{S_0}\right) = -\mathbf{b}:\mathbf{D} = -\sum_{i,j=1}^{3} b_{ij} D_{ij} \qquad (1)$$

where $S_n(b)$ and $S(b_0)$ are the signal intensities with and without nth diffusion sensitizing gradient, respectively; b_{ij} is a component of the diffusion gradient b matrix; D_{ij} is a component of the diffusion tensor D. The colon designates the generalized dot product. Each direction of a diffusion sensitizing gradient is described by individual matrix [5].

Computations on floating-point numbers are biased with an error resulting from numerical representation of the numbers. Only a finite length string of binary words can be used for representation of a number, what in the case of irrational values (i.e. with infinite binary expansion) such as π or Euler's number, leads to a necessary rounding and loss of the precision.

Another type of an error is the cut-off error. It occurs during computing as a result of decreasing the number of operations, e.g. during computing an infinite Taylor series while calculating the value of ln, extremely important fact is that a minor numerical error can grow during further calculations (e.g. multiplication of small values by larger ones) and cause a major error in the result. The aim of this work is to test the influence of the numerical error on the result of the DTI method. The errors stemming from the fact that the measurements of physical quantities can be done only with the limited accuracy are neglected here.

Several numerical methods for calculating a system of linear equations such as the method of Gauss elimination and its variations (Gauss-Jordan, Gauss with partial pivoting), LU method and Cramer's rule, were implemented. Additionally the implementations of Gauss elimination method and LU decomposition from the Numerical Recipes were added. The following experimental equipment specification were used: Processor Intel Core Intel Core i7-4700MQ @ 2.4 GHz, 8 GB RAM.

II. Verification of the Methods

The first stage of the work was checking the correctness of the implemented methods on random data. For this purpose, each independent value of bn matrix and D tensor were drawn from the range -1 to 1. By transforming the equation (1) one obtains the formula expressing Sn value:

$$S_n = \mathrm{e}^{-\mathrm{b_n}:\mathbf{D}} S_0 \qquad (2)$$

For final determination of S_n one needs a value of S_0 which is also drawn from the range [-1;1].

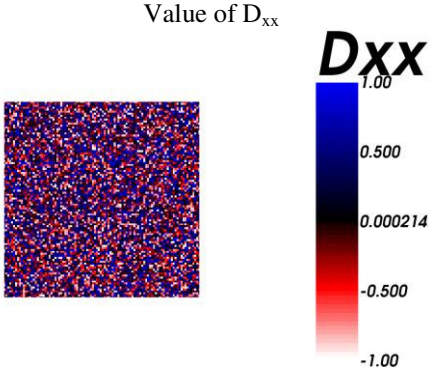
Equation (1) has the following form in the matrix representation:

$$\begin{bmatrix} -b_{xx1} & -b_{yy1} & -b_{zz1} & -2b_{xy1} & -2b_{xz1} & -2b_{yz1} \\ -b_{xx2} & -b_{yy2} & -b_{zz2} & -2b_{xy2} & -2b_{xz2} & -2b_{yz2} \\ -b_{xx3} & -b_{yy3} & -b_{zz3} & -2b_{xy3} & -2b_{xz3} & -2b_{yz3} \\ -b_{xx4} & -b_{yy4} & -b_{zz4} & -2b_{xy4} & -b2_{xz4} & -2b_{yz4} \\ -b_{xx5} & -b_{yy5} & -b_{zz5} & -2b_{xy5} & -2b_{xz5} & -2b_{yz5} \\ -b_{xx6} & -b_{yy6} & -b_{zz6} & -2b_{xy6} & -2b_{xz6} & -2b_{yz6} \end{bmatrix} \begin{bmatrix} D_{xx} \\ D_{yy} \\ D_{zz} \\ D_{xy} \\ D_{xz} \\ D_{yz} \end{bmatrix} = \begin{bmatrix} ln\left(S_1/S_0\right) \\ ln\left(S_2/S_0\right) \\ ln\left(S_3/S_0\right) \\ ln\left(S_4/S_0\right) \\ ln\left(S_5/S_0\right) \\ ln\left(S_6/S_0\right) \end{bmatrix} \quad (3)$$

In the case of the presented test, all values in the above equation are known. In DTI, the problem lies in determination of unknown values of the diffusion tensor D. To do that, a system of equations containing six unknown variables has to be solved. By solving the system of equations for randomly generated b and S_n (3), new values of diffusion tensor D' were determined. Comparing the values of initially determined tensor D with calculated D' one obtains the numerical error ε.

$$\varepsilon = \sum_{i,j=1}^{3} \left| D_{ij} - D'_{ij} \right| \quad (4)$$

Value of D_{xx}



Distribution of the error ε

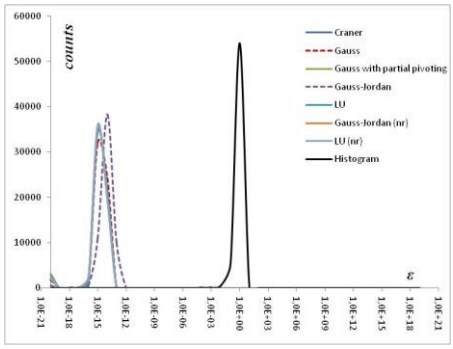


Fig. 1 Distribution of the numerical error ε for all elements of tensor D and value of the xx element of randomly generated diffusion tensor.

On the basis of Fig. 1 one can conclude that the numerical error independently from the method chosen, is 15 orders of magnitude smaller than values of components of the drawn Tensor D, what confirms correctness of the implementation of the methods.

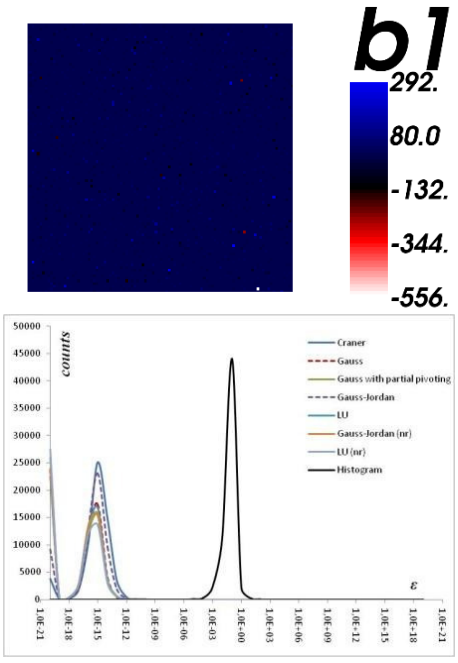
III. Comparison of the Methods

In the case of DTI data, the exact value of tensor D is unknown, what makes the direct determination of the numerical error impossible. However, the error can be estimated alternatively. In DTI one measures values of S_n for the assumed gradient b_n. Solving the system of equations (3) for S_n and b_n data the values of Tensor D (with a certain numerical error) are obtained. Then, using the formula (2) S'_n values were determined. Comparing the input values S_n with re-calculated S'_n, one obtains indirect numerical error ξ:

$$\xi = \sum_{n=1}^{6} \left| S_n - S'_n \right| \quad (5)$$

In order to verify the influence of the numerical error on the real data, two datasets acquired using two types of phantoms (isotropic and anisotropic) and the reference random data were analyzed. It is important that real data belong to distributions which are far from the uniform distribution (-1.0, 1.0).

On the basis of Fig. 2 it can be stated that, for the randomly generated data, the indirect error for all numerical methods is 15 orders of magnitude smaller than the average value of the input data. But on the other hand, in the case of real measurements, the results for the Gauss elimination- and Gauss-Jordan method are different from the others. In theory, the difference between these methods is: If, using elementary row operations, the augmented matrix is reduced to row echelon form (REF), then the process is called Gaussian elimination. If the matrix is reduced to reduced row echelon form (RREF), the process is called Gauss-Jordan elimination. For these methods, a large part of the error is of the same order as the input values.



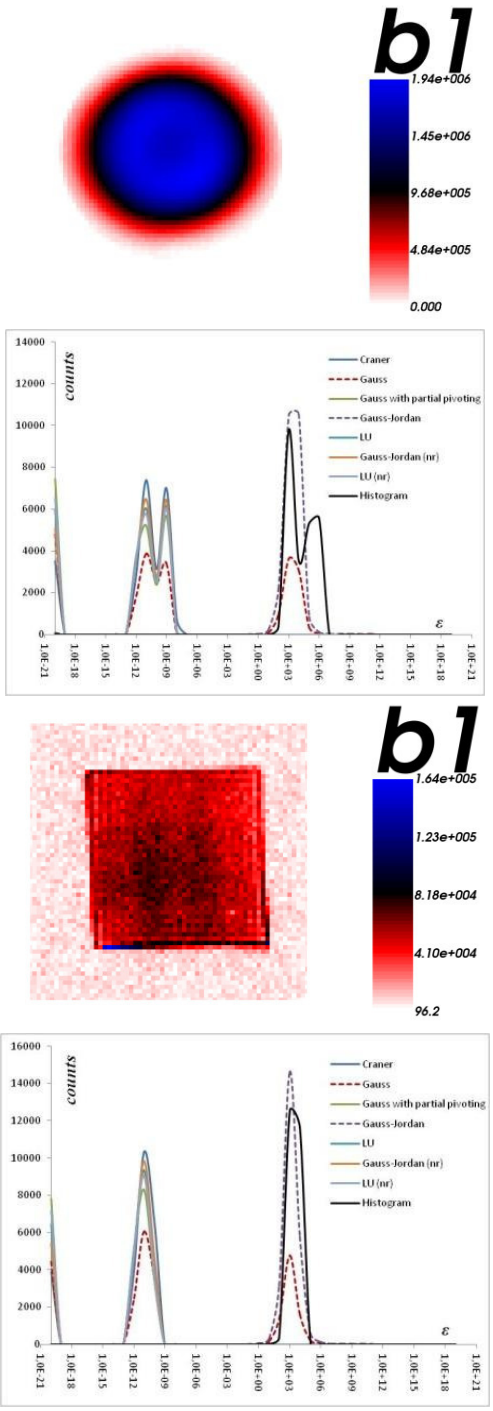


Fig. 2 Distribution of the indirect numerical error for various methods a) random data, b) isotropic phantom, c) anisotropic phantom.

Subsequently the difference between the results obtained with LU method (numerical recipes) and the results obtained in the remaining methods and (Fig.3) were compared. On this basis it can be concluded that the biggest numerical error appears in the noise area. In the case of Gauss and Gauss-Jordan those errors are significantly larger than in the phantom region. In the other methods the errors in both phantom and noise regions are of the same order as the errors in the noise area in Gauss and Gauss-Jordan methods.

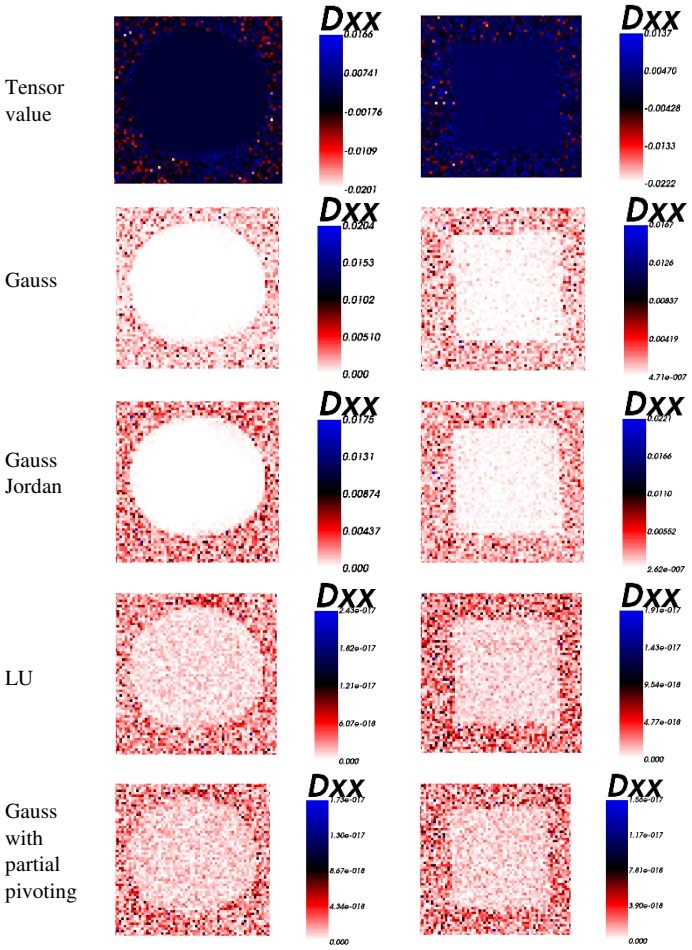


Fig. 3 Distribution of the indirect numerical error for various methods

The last examined aspect was the duration of the calculations. For this purpose, the averaged computing time was determined by repeating the calculations 1000 times for anisotropic phantom (64x64 image) for each of the implemented methods. The fastest method was the Gauss elimination method, the rest were 15-70% slower. From the previous tests it is known that Gauss method is very prone to numerical errors. The fastest method out of ones giving correct results was the LU, which was 15% slower than Gauss. if one takes into account the total duration of computing for DTI, i.e. the memory allocation, calculation of own values etc., then the percent difference between Gauss and LU decreases to 8%.

TABLE I.
DEFINING CHARACTERISTICS OF FIVE EARLY DIGITAL COMPUTERS

	Time of computing tensor D		Time of computing BSD	
Method	**Time [ms]**	**Ratio to Gauss[%]**	**Time [ms]**	**Ratio to Gauss[%]**
Gauss	2.05	100.00	5.49	100.00
Gauss-Jordan	2.88	140.48	6.38	116.21
LU	2.36	115.12	5.95	108.37
Gauss with partial pivoting	3.49	170.24	7.04	128.23
LU (nr)	2.54	123.90	5.92	107.83
Gauss-Jordan(nr)	3.33	162.43	6.72	122.40

IV. Conclusion

Calculating the diffusion tensor D in DTI one has to take into account the impact of the numerical errors, occurring especially in the noise area. Computing time, if all aspects such as the time taken for the allocation of the memory and determining own values and vectors are taken into account, is of minor importance. The methods less affected by the numerical errors, differs in the execution time about 10%. The optimized algorithms will be implement in the B-matrix Spatial Distribution DTI(BSD-DTI) method using a spatial distribution of b-matrix.

Acknowledgment

The National Centre for Research and Development for grant PBS2/A2/16/2013.

References

[1] Basser PJ, Mattiello J, LeBihan D. MR diffusion tensor spectroscopy and imaging. Biophysical Journal. 1994; 66(1):259-267. W.-K. Chen, Linear Networks and Systems (Book style). Belmont, CA: Wadsworth, 1993, pp. 123–135.

[2] Krzyżak A. T., et al., Quantitative assessment of Injury in Rats Spinal Cord in vivo using MRI of Water Diffusion Tensor, App. Magnetic Resonance, 2008; 34; 3-20.

[3] Trivedi R., Rathore R. K., Gupta R. K. Review: Clinical application of diffusion tensor imaging. The Indian Journal of Radiology & Imaging. 2008;18(1):45-52. doi:10.4103/0971-3026.38505.

[4] Mandl R. C. W., Schnack H. G., Zwiers M. P., van der Schaaf A., Kahn R. S., et al. (2008) Functional Diffusion Tensor Imaging: Measuring Task-Related Fractional Anisotropy Changes in the Human Brain along White Matter Tracts. PLoS ONE 3(11): e3631.doi:10.1371/journal.pone.0003631.

[5] Basser P. J., Mattiello J., LeBihan D. Estimation of the effective self-diffusion tensor from the NMR spin echo. J. Magn. Reson. B 1994; 103:247–54.

[6] Krzyżak A. T., Olejniczak Z., Improving the accuracy of PGSE DTI experiment using spatial distribution of b matrix, Magnetic Resonance Imaging, 2015; 33; 286-295.

4th International Conference on Innovative Network Systems and Applications

MODERN network systems encompass a wide range of solutions and technologies, including wireless and wired networks, network systems, services and applications. This results in numerous active research areas oriented towards various technical, scientific and social aspects of network systems and applications. The primary objective of Innovative Network Systems and Applications (iNetSApp) conference is to group network-related events and promote synergy between different fields of network-related research. To stimulate the cooperation between commercial research community and academia, the conference is co-organised by Research and Development Centre Orange Labs Poland and leading universities from Poland, Slovak Republic and United Arab Emirates.

The conference continues the experience of Frontiers in Network Applications and Network Systems (FINANS), International Conference on Wireless Sensor Networks (WSN), and International Symposium on Web Services (WSS). As in the previous years, not only research papers, but also papers summarising the development of innovative network systems and applications are welcome.

- EAIS'16—3rd Workshop on Emerging Aspects in Information Security
- SoFast-WS'16—5th International Symposium on Frontiers in Network Applications, Network Systems and Web Services
- WSN'16 - 5th International Conference on Wireless Sensor Networks
- main iNetSApp'16 track includes remaining topics, related to network systems and addressed not only to one of the tracks listed above

TOPICS

- Architecture, scalability and security of network systems,
- Web services – standards and applications,
- Service delivery platforms—architecture and applications,
- The applications of intelligent techniques in network systems,
- Innovative network applications,
- Network-based computing systems,
- Network-based data storage systems,
- Technical and social aspects of Open API and open data,
- Computer forensic and network security,
- Social, organizational and other aspects of information security,
- Network, cloud and data security,
- Misuse and intrusion detection,
- Traffic classification algorithms and techniques,
- Network protocols and standards,
- Network traffic engineering,
- Wireless communications,
- Control of networks,
- Internet of things,
- Sensor Circuits and Sensor Devices,
- Architectures, Protocols and Algorithms of Sensor Networks,
- Management, Energy and Control of Sensor Networks,
- Data Allocation and Information Processing in Sensor Networks,
- Resource Allocation, Services, QoS and Fault Tolerance in Sensor Networks
- Security and Monitoring of Sensor Networks,
- Software, Applications and Programming of Sensor Network,
- Performance, Simulation and Modeling of Sensor Network,
- Applications of Wireless Sensor Networks,
- Other aspects on network-related research.

STEERING COMMITTEE

- **Sebastian Grabowski, Research and Development Centre Orange Labs Poland,** Poland
- **Zakaria Maamar, Zayed University,** United Arab Emirates
- **Bohdan Macukow, Faculty of Mathematics and Information Science of the Warsaw University of Technology,** Poland
- **Juraj Miček, Department of Technical Cybernetics,** University of Žilina, Slovakia
- **Zbigniew Zielinski, Faculty of Cybernetics of Military University of Technology,** Poland

EVENT CHAIRS

- **Furtak, Janusz,** Military University of Technology, Poland
- **Grzenda, Maciej,** Orange Labs Poland and Warsaw University of Technology, Poland
- **Hodoň, Michal,** University of Žilina, Slovakia

PROGRAM COMMITTEE

- **AbdAllah, Mohamed Mostafa,** Yanbu Industrial College, Saudi Arabia
- **Al-Anbuky, Adnan,** Auckland University of Technology, New Zealand

- **Baranov, Alexander,** Russian State University of Aviation Technology
- **Bataineh, Emad,** Zayed University
- **Ben-Othman, Jalel,** Université Paris 13
- **Chung, Danny Wen-Yaw,** Chung Yuan Christian University
- **Dabrowski, Andrzej,** Warsaw University of Technology, Poland
- **Fouchal, Hacene,** University of Reims Champagne-Ardenne, France
- **Fowler, Scott,** Linköping University
- **Frankowski, Jacek,** Orange Labs, Poland
- **Furnell, Steven,** Plymouth University, United Kingdom
- **Geiger, Gebhard,** Technical University of Munich, Faculty of Economics
- **Ghamri-Doudane, Yacine,** Université La Rochelle
- **Grabowski, Sebastian,** Research and Development Centre Orange Labs Poland, Poland
- **Gu, Yu,** National Institute of Informatics, Japan
- **Guedria, Lotfi,** Centre d'Excellence en Technologies de l'Information et de la Communication
- **Habbas, Zineb,** University of Lorraine
- **Haemmerli, Bernhard,** Hochschule für Technik+Architektur (HTA), Leiter Cisco Regional Academy, Switzerland
- **Howells, Gareth,** University of Kent
- **Husár, Peter,** Technische Universität Ilmenau, Germany
- **Jin, Jiong,** Swinburne University of Technology, Australia
- **Kaczmarski, Krzysztof,** Warsaw University of Technology, Poland
- **Kamoun, Faouzi,** Zayed University
- **Kiedrowicz, Maciej,** Military University of Technology, Poland
- **Kowalski, Andrzej,** Orange Labs, Poland
- **Ksentini, Adlen,** Université de Rennes
- **Laqua, Daniel,** Technische Universität Ilmenau, Germany
- **Legierski, Jarosław,** Orange Labs Poland, Poland
- **Macukow, Bohdan,** Warsaw University of Technology, Poland
- **Marir, Farhi,** Zayed University
- **Micek, Juraj,** University of Žilina, Slovakia
- **Milanová, Jana,** University of Žilina, Slovakia
- **Mokdad, Lynda,** Université Paris-Est, France
- **Monov, Vladimir V.,** Bulgarian Academy of Sciences, Bulgaria
- **Nowicki, Tadeusz,** Military University of Technology, Poland
- **Ouadoudi, Zytoune,** Université IbnTofail, Kenitra, Morocco
- **Scholz, Bernhard,** The University of Sydney, Australia
- **Ševčík, Peter,** University of Žilina, Slovakia
- **Shaaban, Eman,** Ain-Shams university, Egypt
- **Staub, Thomas,** Data Fusion Research Center (DFRC) AG, Switzerland
- **Stokłosa, Janusz,** Poznań University of Technology, Poland
- **Szmit, Maciej,** Orange Labs Poland, Poland
- **Xiao, Yang,** The University of Alabama, United States
- **Yang, Mee Loong,** Auckland University of Technology, New Zealand
- **Zieliński, Zbigniew,** Military University of Technology, Poland

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 961–967
DOI: 10.15439/2016F337
ACSIS, Vol. 8. ISSN 2300-5963

Crowdsourcing based terminal positioning using multidimensional data clustering and interpolation

Noureddine Boujnah
Lodz University of Technology
Institute of Electronics, Poland
Gabes University
Faculty of Sciences, Tunisia
Email: boujnah_noureddine@yahoo.fr

Piotr Korbel
Lodz University of Technology
Institute of Electronics, Poland

Email: piotr.korbel@p.lodz.pl

***Abstract*—Recent years can be characterized by the rapid increase of mobile device usage in people's lives. Contemporary mobile devices are equipped with many sensors and have high computational and processing capabilities. In a crowdsourcing systems, mobile users participate constructively in specific information handling. Data collected by crowd of mobile devices and stored in a database may help offering new services such as operator's radio quality evaluation and tracking of users. In this paper, we focus on mobile positioning by clustering and interpolating data to match fingerprints to positions. Our method is trained during the offline phase and parameters are periodically updated to track possible changes in propagation environment.**

I. Introduction

IN RECENT years a new field of research in radio communication technologies was born taking benefit from wisdom of mobile crowds [1]. In crowdsourcing a large group of users or crowd participate in service gathering, enhancement or evaluation, information sharing and passing in a professional or a social network. Many crowdsourcing applications have been designed so far, and some of them require provision of location information, with different level of accuracy. The traditional way to retrieve mobile location is GPS positioning but its drawback is activating GPS on mobile phone which is power consuming and it does not work well in indoor environment. The actual challenge for crowdsourcing location based techniques, is how to estimate mobile position using only collected statistics and network parameters.

Our paper deals with mobile user location based on crowdsourcing data collected with the use of smartphones. Various techniques exist in literature, the most common is fingerprinting based positioning. Another positioning method is lateration technique, lateration technique depends mainly on propagation model which changes from one area to another, network parameters may also change: transmitted power, antenna parameters, the environment also may changes when there is a new buildings and new trees, some special events may also change model parameters.

Techniques that use mapping between fingerprints and mobile location were also studied and such methods as: $k-$nearest neighbors $k-$NN, neural network and support vector machine (SVM) were applied in order to approximate the location of mobile user [4].

In this work, we investigate the possibility of mobile position determination based on information collected from cellular networks and WLANs. The main idea is based on data clustering and multidimensional interpolation, the use of clustering is justified by the fact that similar fingerprints could exist in different locations, mobile devices sense and retrieve network information periodically, collected data are sent to a database, where received information is organized in tables and columns.

User's position are sparse in outdoor environment and estimation of mobile position is a challenge. The proposed method takes into account the fact that equal fingerprints may occurs in different positions.

The remainder of the paper is organized as follows: in section II we described structure and pre-processing of collected data, in section III we presented the propagation model, related works are discussed in section IV. In section V we propose our method of crowdsourcing based terminal positioning. Performance evaluation and results are presented in section VI we finish by a conclusion and future works.

II. Crowdsourcing data

A. System architecture and data structure

In crowdsourcing context, mobile applications data are collected by user devices and sent to a database. The overall architecture of the application is depicted in Fig. 1. Various data are collected related to the network parameters and also fingerprints, timestamps are also recorded. To summarize, the collected information covers:

1) Received signal strength (RSS) collected from 2G, 3G, 4G serving cells and WLAN access points (AP).
2) Cell and BS identifier (CID or BSID).
3) Primary scrambling code (PSC) or physical cell identity (PCI) for LTE.
4) Timestamp.
5) GPS coordinates.
6) RSS from neighboring cells and access points.

The reference static database contains information about cellular network parameters in the region of interest:

978-83-60810-90-3/$25.00

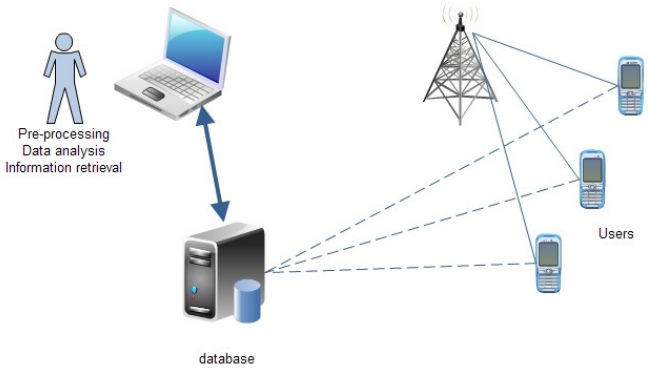


Fig. 1. General overview of a crowdsourcing system

1) Cells in the region of interest.
2) Tilt and Azimuth angle of each cell for cellular technologies.
3) Antenna height and maximum gain.
4) GPS coordinates of a base station.
5) Frequencies.
6) Cell identifier.

B. Data pre-processing

Collected data are periodically sent by mobile phone to the database, user should activate the selected technology for measurement, sometimes there are missing or wrong information. The preprocessing phase is necessary to organize data into tables, each table contain measurements relative to a wireless technology (GSM, UMTS, LTE, WLAN).

Pre-processing phase is compulsory to delete rows with non-significant or missed measurements and to match measurement with the reference database. GPS information is needed in the training phase only. To solve the problem of UTM coordinates, we can refer to conversion method in [3] to use Cartesian coordinate system. For example lateration technique based on fingerprints data uses Euclidean distance.

III. Propagation model

We assume the log-distance shadowing model [5], [9] in most cases of outdoor scenario, spatial received power correlation is also considered [6]. Vector of received power P from K transmitters at a specific position (x, y) is given by:

$$P = \bar{P} + \alpha S(x, y) + \Gamma U \tag{1}$$

Where, P is $K \times 1$ vector of received power at user side, $\bar{P} = P_{TX} + G_{TX}(\theta, \phi) - L_{feeder} - L_c - 20log_{10}(f)$ is the received power at one meter from the transmitter. G_{TX}, P_{TX} and L are: antenna gain, transmitted power and feeder attenuation respectively, and:

$$S(x, y) = \begin{bmatrix} -10\log_{10}(d_1) \\ -10\log_{10}(d_2) \\ \cdots \\ -10\log_{10}(d_K) \end{bmatrix} \tag{2}$$

is the vector of log of distance between location X and each transmitter's location, each element of the vector is given by: $d_i = \sqrt{(x-\alpha_i)^2 + (y-\beta_i)^2}$ for $i = 1 \ldots K$, where (α_i, β_i) is the position of transmitter i. Γ is the result of Cholesky decomposition of the covariance matrix R of the shadowing channel, R is given by:

$$R = \Gamma^T \Gamma \tag{3}$$

U is a vector of normal distribution with mean zero and identity covariance matrix. Elements of U are statistically independent. Many others models based on environment characteristics are reported in the literature. In our work, collected RSS at each user position are assumed to be the mean value of power in equation 1, $RSS = \bar{P} + \alpha S(x, y)$. Statistical properties of the shadowing model are not covered by this paper.

IV. Related works

In outdoor scenario, received power depends on several parameters such as: network configuration, propagation environment, mobile orientation and user's velocity, complexity of exploiting fingerprints information from crowdsourcing data increases when one or more parameters are missing or erroneous. lateration and radio map techniques assume knowledge of network parameters and also radio models. Many propagation models exist in literature, and choice of the right model depends on the environment. In many research works we assume shadowing model with known parameters such as propagation exponent and shadowing variance.

Geolocalization methods of mobile users range from geometrical approaches such as lateration techniques using Cartesian coordinates [8], time and angle of arrival at a given position, to data analysis approach using matching methods between fingerprints and associated positions:

1) Lateration based: In lateration technique, we assume the shadowing propagation model with known parameters, hence, we estimate the distance between current position and transmitter position, the number of received signals should be at least 3, by solving a set of K linear equations, mobile position $X = (x, y)^T$ is estimated as:

$$\hat{X} = (A^T A)^{-1} A^T b \tag{4}$$

Where, A and b are defined as:

$$\begin{cases} A_{i,1} = 2(\alpha_i - \sum_{j=1}^{K} w_{i,j}\alpha_j) \\ A_{i,2} = 2(\beta_i - \sum_{j=1}^{K} w_{i,j}\beta_j) \\ b_i^{(1)} = (\alpha_i^2 + \beta_i^2) - \sum_{j=1}^{K} w_{i,j}(\alpha_j^2 + \beta_j^2) \\ b_i^{(2)} = \hat{d_i}^2 - \sum_{j=1}^{K} w_{i,j}\hat{d_j}^2 \\ b_i = b_i^{(1)} - b_i^{(2)} \\ \sum_{j=1}^{K} w_{i,j} = 1 \\ \hat{d_i}^2 = 10^{-\frac{P_i(d) - P(d_0)}{5\alpha}} \end{cases} \quad (5)$$

This method is known as linear least square (LLS) [7], weighting coefficients $w_{i,j}$ are set to $\frac{1}{K}$. Other techniques based on distance estimation such as nonlinear least square (NLS) and Differential RSS techniques are applied when information on transmitted power and antenna parameters is missing [9].

Geometrical method has some limitations because propagation channel is subject of fast and slow fluctuations that vary from location to another, hence affects positioning accuracy. It is possible to exploit the lateration method in the future in crowdsourcing concept.

2) Neural network based: Neural network is commonly used in pattern recognition in signal processing, in mobile positioning it is used to map received RSS P and associated positions X in a given region [10], during training phase, layer's parameters are tuned in order to minimize the global mean square error MSE. Neural network techniques used in literature are MLP (multilayer perception) and ANN (artificial neural network). Simulation results obtained for indoor environment are good in terms of accuracy. The disadvantage is the computational complexity of the method.
3) Radio coverage map based: During planning phase of radio network, coverage and quality maps should be traced and stored in a database. Strength of the signal received by a mobile phone from nearby base stations is compared to reference points in the radio map. The reference points are usually organized into rectangular or hexagonal grids. This technique is not adapted to the variation of radio propagation channel and parameters changes.

Localization methods are evaluated based on accuracy and algorithmic complexity. When collected data are sparse in space, previous methods seem to be less accurate because there is not enough points to determine model parameters in each area and we may face the problem of appearance of the same fingerprints (SF) in different positions.

V. Proposed Method

In our proposed method we start by subdividing data into two sets, the first one used to perform data clustering using collected information and built up functions that fit well each cluster in order to overcome spatial data sparsity problem, the second set is used to perform tests. Fig. 2 summarizes our approach.

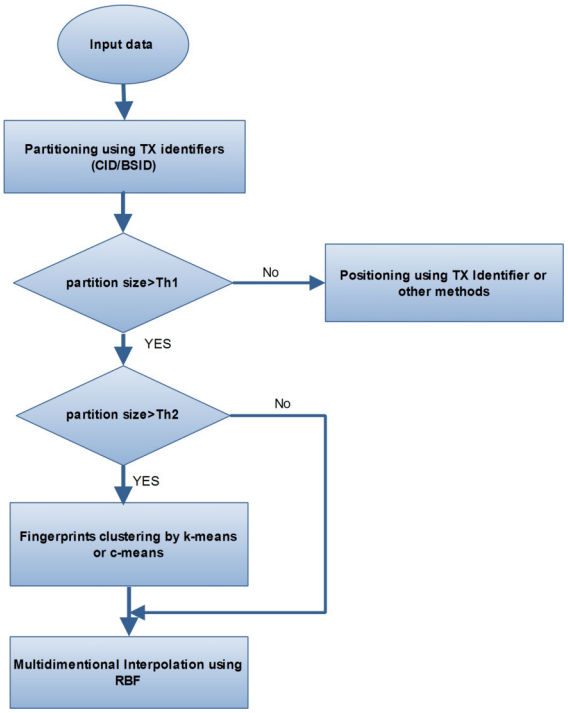


Fig. 2. proposed method

Many clustering techniques are used in literature, such as $k-$means, $c-$means, multimodal classification. First, we cluster collected data based on their CID or BSID, to guarantee that samples belong to one geographical area. Let P_i be the vector of received power in a given position $X = (x, y)^T$, number of element in P_i may change from position to another, depends on sensitivity of the mobile terminal. Clustering using cell identifier or base station identifier is an alternative to determine location area of mobile user, if we consider identifier of more than one cell, we can enhance precision.

After classification of user positions into small regions of interest, we will focus on data interpolation assuming compactness of input data in each cluster, clustering using both transmitter identifier and RSS may reduce the similar fingerprint problem and enhance then positioning accuracy. We look for function f that fits well all points inside a specific class. For example in [7] interpolation using polynomial regression was applied. Our method assumes the use of radial basis function (RBF) with Gaussian kernel.

A. Data clustering

Clustering or data partitioning, is the way of grouping data based on some parameters or criteria, for example, received power, CID or BSID, frequency.

1) Network parameters based:

- Frequency: frequency of the received signal can give information about location of serving cell, if we know frequency plan of region of interest each cell is characterized by its frequencies, but frequency plan may change when needed, hence this method is not good from crowdsourcing point of view.
- Cell identifier or base station identifier: knowledge of CID/BSID increase location accuracy of mobile user, if we consider more than one signal we can approximate better the mobile position, generally, values of CID and BSID are fixed, CID/BSID is considered also as spatial clustering.
- The strongest signals: The number of received signal K in a given position may change from area to another, the RSS vector contains the power of the serving cell/transmitter and the candidate cells or the neighboring cell.

2) Fingerprint based clustering:

$k-$Means: $k-$Means is an unsupervised clustering algorithm taking as input multidimensional data and as input N centroids. Each input fingerprint P_i, power vector of received signal strength, is assigned to the cluster with center C_j as:

$$C_j = \arg\min_{C_l} ||P_i - C_l||^2 \tag{6}$$

Cluster centers C_j are computed by iterative process, the new center of each cluster is the centroid of the old cluster.

Fuzzy $C-$Means: Is a modified version of $K-$means, was proposed by James Bezdek [11], where each sample may belong to a cluster with some membership degree, the mathematical formulation of FCM is given by:

$$\begin{cases} \arg\min_{C_j,U} \sum_{i=1}^{M}\sum_{j=1}^{K} u_{i,j}^m ||P_i - C_j||^2 \\ \sum_{j=1}^{K} u_{i,j} = 1 \\ m > 1 \end{cases} \tag{7}$$

Optimization is performed using the differentiation with respect to U, C and λ_i of:

$$\xi(U,C) = \sum_{i=1}^{M}\sum_{j=1}^{K} u_{i,j}^m ||P_i - C_j||^2 - \sum_{i=1}^{M} \lambda_i (\sum_{j=1}^{K} u_{i,j} - 1) \tag{8}$$

λ_i are the Lagrangian multipliers relative to each condition, cluster centers and weights are computed using iterative algorithm.

B. Class definition

Each data class is represented by a vector V containing: the number of sensed signals K, cell identifier CID/BSID and fingerprint centroid C_j.

$$V = \{CID, K, C_j\}$$

CID may contain two or three identifiers of transmitters. From cellular network point of view, we know that location area with one CID vector, where more than two signals co-exists, are unique, but we can not generalize this assumption in cell's borders. If we increase the number of identifiers that characterize one area, it is possible to reduce the cluster and hence enhance positioning accuracy. Before using interpolation, each measurement is assigned to its corresponding class V, for more accuracy we can use more than one cell for CID clustering.

C. Multidimensional data interpolation

Multidimensional interpolation is a mathematical tool that fit some input data with dimension $N \times K$ to output data with dimension $N \times 2$, Interpolation problem is equivalent to solving set of linear equations in order to determine parameters of interpolating function [12]. In positioning radial basis function was used for indoor localization in [13] where RSS fingerprints are fitted with corresponding positions, mathematical formulation of interpolation is given by:

$$\begin{array}{cccc} X: & \mathbb{R}^K & \longrightarrow & \mathbb{R}^2 \\ & P & \longmapsto & f_K(P) \end{array} \tag{9}$$

Where K is the number of detected signals by the receiver at a given position. P is the input vector and X the associated position. Interpolation by radial basis function (RBF) is given by:

$$X = \sum_{i=1}^{N} W_i \phi_h(||P - P_i||) \tag{10}$$

Where ϕ_h is a radial function, h a shape parameter, and W_i are 2×1 weighting coefficients. To compute weight coefficients, it is necessary to train the network using fingerprints and associated locations.

We assume N input and output data (X_i, P_i), Equation 10 became:

$$X_j = \sum_{i=1}^{N} W_i q_{i,j} \tag{11}$$

Where $q_{i,j} = \phi_h(||P_j - P_i||)$, T_h is $N \times N-$symmetric matrix, and X is $N \times 2$ matrix. Equation 11 is equivalent to:

$$W = Q_h^{-1} X \tag{12}$$

In the case where Q_h is singular matrix, we can modify the solution by introducing small value ϵ as:

$$W_h = (Q_h + \epsilon I)^{-1} X \tag{13}$$

This case occurs when there are the same fingerprints in one cluster. If we want to fix the number of basis functions in each cluster, the weight vector is given as:

$$W = (Q_h^T Q_h)^{-1} Q_h^T X \tag{14}$$

Gaussian kernel function is used for data interpolation.

$$\phi_h(r) = \exp(-(hr)^2) \tag{15}$$

Performance of RBF interpolation depends mainly on choice of h, number of radial basis functions and also the choice of the radial norm.

D. Advantages and limitations of the proposed method

In the test phase, first we assign mobile to its cluster using one clustering technique, then, we estimate its position based on interpolation function.

In order to evaluate the accuracy of our method, we use two approaches: in the first one we compute the distance between from the estimated value of the position and a reference point, with nearest RSS vector. Let start by first order Taylor expansion of $\phi_h(||P - P_i||)$ around P_j:

$$\phi_h(r_i) = \phi_h(r_{i,j}) - 2h^2(P - P_j)^T(P_j - P_i)\phi_h(r_{i,j}) \quad (16)$$

Where P_j is the nearest neighbor to P, $r_i = ||P - P_i||$ and $r_{i,j} = ||P_j - P_i||$. The distance between the nearest point with position X_j and test point with position X, and using function interpolation of user's position, is expressed as:

$$d_{RBF} = 2h^2||\sum_{i=1}^{N} W_i(P - P_j)^T(P_j - P_i)\phi_h(r_{i,j})|| \quad (17)$$

The second level of evaluation is to check if CID cluster of the test point and the reference point are overlapping or not. If the two clusters are not overlapping, the estimation using RBF is wrong and we have to assign a position from the training set that have similar CIDs.

If the value of the distance d_{RBF} is low, and the test cluster overlap with the reference one, the RBF interpolation approximates well the position.

When receiving low numbers of signals from transmitters, the probability of getting SF increases. We conclude that interpolation is more efficient when the number of received signals is high in one hand, in the other hand, we need more sample data to get good approximation by interpolation when the dimension of the received signal is high. The number of collected identifiers may reduce the size of the cluster. Then in the training phase, when the number of samples is high within the cluster region, accuracy increases, otherwise, we cluster the data with lower number of cell identifiers.

Differences between global interpolation and interpolation within cluster are:

- Lower complexity for interpolation within a cluster and matrix inversion do not consume memory.
- Lower number of SF for cluster based interpolation, hence good fitting properties.
- Higher error in positioning for global interpolation with some probability different from zero.

Limitations of the proposed method are:

- Samples are sparse, to overcome this problem we need more measurements and more memory to store data.
- This method should be upgraded in order to track fluctuations of propagation channel, transmitted power and network's parameters changes.

VI. Performance Evaluation

A. Measurement setup

The crowdsourcing system used for evaluation of the proposed approach consisted of application and database servers responsible for data storage and pre-processing, and a group of Android-based mobile devices. The test devices were equipped with dedicated mobile application running in the background and responsible for collecting of the measurement data during normal device operation. The background service is also responsible for communication with the application servers and periodic uploading of the measured data to the system database. The data were collected in the opportunistic manner, i.e. during everyday device usage by a group of test users, and using different devices. The range of data reported to the system covered strengths of the signals received by the device from surrounding radio access network transmitters (the set of parameters depends on the radio access technology as described in Section II) and from nearby Wi-Fi access points, as well as actual GPS-based position of the user.

B. Experimental results

During the initial tests of the system, the measured data were collected in the surroundings of Lodz University of Technology campus area using number of terminals. The locations of reference measurement points are shown in Fig. 3 and 4, measurements could be performed dependently or separately.

Fig. 3. ROI and part of experimental data positions (GSM measurements)

Fig. 4. ROI and part of experimental data positions (LTE measurements)

A selection of 80% of measurements are dedicated to find clusters and interpolate data in each cluster using radial

basis functions, the remaining 20% of RSS measurements are reserved to the test phase.

Fig. 5 shows cell identity pairs from collected data, in y-axis the identifier of the strongest cell or server and in x-axis the best neighbor cell, hence each point describe a set of fingerprints, more the number of fingerprints is more the accuracy we get. In cellular network, theoritically two CIDs are sufficient to characterize a compact location area because antenna are directive, in WLAN technology antennas are generally omni-directional and at least three received signals are required for area identification. It is possible also to cluster data using three or more best cells in order to identify a location area with some serving cells. Fingerprints associated

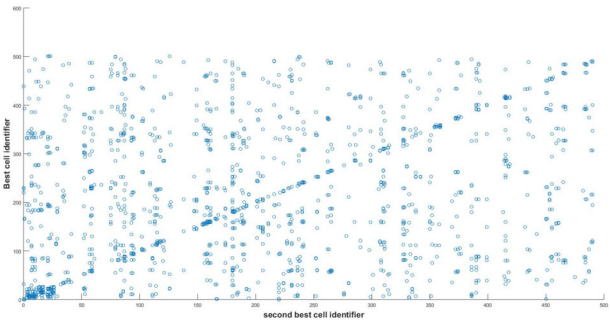


Fig. 5. CID based clustering

to unique CID pair are collected together. Then RSS vectors are clustered based on their dimensionality, dimension K of received RSS vary from 2 to 7.

Clustering using k-means or FCM will be applied only when the number of fingerprints in each fixed (CID, K) set is huge, the splitting of data into smaller clusters may reduce the complexity in the interpolation phase.

Table I shows the number of samples for each signal dimension. The fitting error increase with the number of

TABLE I
DISTRIBUTION BASED ON K

K	Total samples
7	87296
6	55294
5	17813
4	3982
3	225
2	35

existing SF pairs in the training sequence, it is obvious that for global fitting method the error is higher than for clustered method.

In total we have 416 pairs of cell identifiers (CID), retrieved from the collected data, an average number of fingerprints in a CID cluster equal to 348.4. In order to increase accuracy of location area, we can also use triplets of CIDs, in this case the size of the area will be further reduced.

When the number of fingerprints is between $Th2 = 50$ and $Th2 = 10$ for given pairs we use interpolation technique after CID clustering. When the number of fingerprints is higher than $Th2$, we create new sub-clusters using c-means or FCM, then we use interpolation. Table II show the distribution of cluster's size Fig. 6 shows three GSM clusters of size 52, 55

TABLE II
DISTRIBUTION BY CLUSTER SIZE

cluster size ≥ 51	$10 \leq$ cluster size ≤ 50	cluster size ≤ 9
201	143	72

and 171 respectively: Clusters with sample size higher than

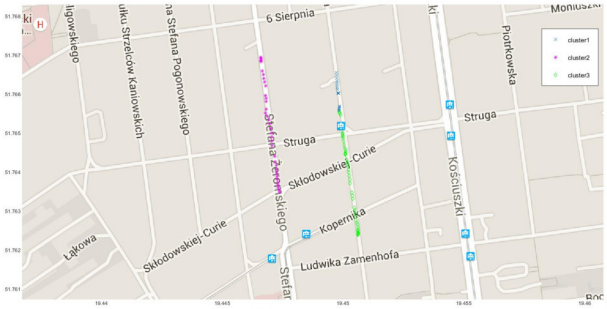


Fig. 6. GSM clusters

$Th2$ are subdivided into subsets using $k-$means or FCM, furthers studies will be focused on how to tune thresholds and also in which case we should use one or the other of the clustering algorithm in order to get more accurate results.

After the clustering phase, we determine the interpolation function assigned to each cluster using RBF. We compare in the table III the fitting error between the proposed method and interpolation only method.

TABLE III
ERROR OF FITTING

cluster based RBF	Global RBF
$8.9357.10^{-10}$	$9.03111.10^{-10}$

Increasing the number of cell identifiers per each cluster may adds more accuracy comparing to the case when using only a pair of CI, but number of samples will be low and interpolation error will increase too.

From measurements we can notice that, a test point could be assigned a position in a different cluster while using global interpolation technique, as we discussed in the limits and advantages paragraph, we can detect this error and assign a new position to the test point based on CID only, in this case we may gain in terms of accuracy. Our method requires many measurements to train and extract associated parameters for each cluster, as we introduce the SF problem, we may also define the same position problem SP and how to solve it. In future we plan to consider such issues as network parameters

tackling as a result of tilts modification, transmitted power and azimuth changes using correlation between measured information with its corresponding cluster. Crowdsourcing adds more consistency to positioning, In our method we add clustering using transmitter identifier with comparison to lateration techniques which based on propagation model, our approach is more realistic and can track possibles changes that occurs in propagation channel. Lateration technique assumes knowledge of APs positions and also transmitter's parameters. Our method can outperform also radio map based technique as many changes may happen in the network after the initial deployment.

Disadvantages of our method is the updating policy of the database. Any change of radio network parameters induces possible changes in received power, and old value of saved RSS will be useless.

VII. Conclusion and future works

This paper deals with crowdsourcing data analysis with special emphasis on localization. Statistical information from data collected by smartphone can be exploited for user localization, approach using techniques like lateration depends on propagation model and the model vary in space and time. We start our work by data partitioning into small or compact area using cell identifier of received signals, if the cardinality of partition is higher than a given threshold, associated RSS fingerprints are clustered into each region using $k-$means or $c-$means methods. Adding clustering may reduce the same fingerprint problem SF that can occur in radio propagation scenario. The second step of the training process was multidimensional interpolation using RBF with Gaussian kernel to estimate the user's position. Clustering and interpolation increase positioning accuracy, in global interpolation method, the probability of similar measurement in two different locations increase with collected data, hence it is possible to get a bad estimate of the position, due to the channel fluctuation it is possible also to get two different fingerprints for the same position. The proposed method, can be updated each time we have new training data, it is possible also to store new sub area and update parameters of interpolation function. As future works, we focus to exploit more correlation techniques between fingerprints and transmitter's identifiers and tracking techniques.

Acknowledgment

This work was partially funded by the European Commission under the Erasmus Mundus E-GOV-TN project (Open Government data in Tunisia for service innovation and transparency) - EMA2; Grant Agreement no. 2013-2434/001-001 and by the Polish National Center for Research and Development under the project PBS2/B3/24/2014.

References

[1] G. Chatzimilioudis, A. Konstantinidis, C. Laoudias, and D. Zeinalipour-Yazti, "Crowdsourcing with smartphones," IEEE Internet Comput., vol. 16, no. 5, pp. 36-44, Sep.-Oct. 2012. DOI: 10.1109/MIC.2012.70

[2] Georgiou, K.; Constambeys, T.; Laoudias, C.; Petrou, L.; Chatzimilioudis, G.; Zeinalipour-Yazti, D. "Anyplace: A Crowdsourced Indoor Information Service," Mobile Data Management (MDM), 2015 16th IEEE International Conference on, On page(s): 291 - 294 Volume: 1, 15-18 June 2015. DOI: 10.1109/MDM.2015.80

[3] "Coordinate Conversions and Transformations including Formulas," IOGP Publication 373-7-2- Geomatics Guidance Note number 7, part 2 - April 2015

[4] H. Liu, H. Darabi, P. Banerjee, and J. Liu, "Survey of wireless indoor positioning techniques and systems," Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on, vol. 37, no. 6, pp. 1067-1080, nov. 2007. DOI: 10.1109/TSMCC.2007.905750

[5] T.K. Sarkar, Z. Ji, K. Kim, A. Medouri, M. Salazar-Palma. "A Survey of Various Propagation Models for Mobile Communication," IEEE Antennas and Propagation Magazine, Vol. 45, No. 3, June 2003. DOI: 10.1002/0471722839

[6] F. Graziosi and F. Santucci, "A general correlation model for shadow fading in mobile radio systems," IEEE Commun. Lett., vol. 6, no. 3, pp. 102-104, Mar. 2002. DOI: 10.1109/4234.991146

[7] J. Yang and Y.Chen, "Indoor Localization Using Improved RSS-Based Lateration Methods," in Proc. IEEE Globecom, Nov. 2009, pp. 1-6. DOI: 10.1109/GLOCOM.2009.5425237

[8] J.Yang, Y. Chen "Indoor localization using improved RSS-based lateration methods," Global Telecommunications Conference, 2009. GLOBECOM 2009, Nov. 30, 2009-Dec. 4, 2009. DOI: 10.1109/GLOCOM.2009.5425237

[9] J. H. Lee and R. M. Buehrer, "Location estimation using differential RSS with spatially correlated shadowing," in Proc. IEEE GLOBECOM, Nov. 2009, pp. 1-6. DOI: 10.1109/GLOCOM.2009.5425272

[10] A. Payal, C. S. Rai, and B. V. R. Reddy, "Analysis of some feedward artificial neural network training algorithms for developing localization framework in wireless sensor networks," Wireless Pers. Commun., vol. 82, no. 4, pp. 2519-2536, 2015.

[11] J.C. Bezdek, R. Ehrlich, W. Full "FCM: The fuzzy c-means clustering algorithm," Computer and Geoscience, Vol.10, No 2-3, pp 191-203,1984.

[12] D.S. Broomhead, D. Lowe "Multivariable Functional Interpolation and Adaptive Networks," Complex system 2: 321-355, 1988

[13] C. Laoudias, P. Kemppi, and C. Panayiotou, "Localization using radial basis function networks and signal strength fingerprints in WLAN," in IEEE GLOBECOM, 2009, pp. 1-6 DOI: 10.1109/GLOCOM.2009.5425278

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 969–977
DOI: 10.15439/2016F398
ACSIS, Vol. 8. ISSN 2300-5963

Highly customizable framework for performance evaluation of *LOOM*-based SDN controllers

Szymon Mentel
Erlang Solutions,
ul. Batorego 25, Kraków, Poland
Email: szymon.mentel@erlang-solutions.com

Marek Konieczny, Sławomir Zieliński
Department of Computer Science
AGH University of Science and Technology,
al. Mickiewicza 30, Kraków, Poland
Email: {marekko, slawek}@agh.edu.pl

***Abstract*—The article presents an innovative method for assessing performance of modular SDN controllers, focusing on test customization. The method was validated by construction of a testing framework that gives its users the opportunity to emulate various network traffic patterns by using arbitrarily chosen applications, rather than simulating workloads. The presented solution is more comprehensive than others available in contemporary SDN environments also because it that takes into account specific features of modular SDN controllers.**

I. Introduction

Software Defined Networking (SDN) is a promising concept for computer networking. The main idea behind it is the separation of control and data planes. The control plane is moved to a logically centralized, software-based controller. However, as the scale of the application grows, the performance of a controller can become a bottleneck and appropriate techniques for controller scaling need to be employed. For example consider using hierarchical controllers [1], increasing the autonomy of data plane components [2], [3], distributing controllers [4] or elastic scaling [5]. It is also possible to use adaptation mechanisms utilized in SOA systems [6]. Another method - which is of focus for the article - is increasing controller's performance by building upon the modularity offered by modern programming and runtime environments, such as Erlang.

Although there are works related to measuring the performance of monolithic SDN controllers, there is little research regarding modular ones. Therefore, this paper is focused on the development of a method and framework for assessing performance of modular SDN controllers. The framework also provides its users with means required to emulate various network traffic patterns. It was implemented upon open technologies, such as the *LOOM Controller Framework* [7], *Mininet Cluster Edition*[1], and *Open vSwitch (OVS)*[2].

The paper organization is as follows. Sections II and III overview the domain of the presented research, including research in the area of measuring performance of SDN controllers. The survey forms a base for choosing a testing approach and designing the framework architecture - both are presented in Section IV. Section V presents a proof-of-concept implementation of a testing environment and in this way demonstrates the options for customizing the framework to a practical use case. Moreover, it presents results gathered from a series of experiments and the conclusions that were drawn from them. The article ends with concluding remarks and acknowledgments, which form Sections VI and VII, respectively.

[1]https://github.com/mininet/mininet/wiki/Cluster-Edition-Prototype/
[2]http://openvswitch.org/

II. Background

In *SDN*, forwarding logic no longer has to run on specific hardware built into the switches, routers or other networking devices. The *control plane* functions, implemented by a *SDN controller*, can be achieved in general purpose programming languages and run on regular servers. Thanks to that, *control plane* development is much more flexible, cheaper and faster.

To make user traffic reach its destinations, the control plane needs to communicate with data plane devices. Currently, the most popular protocol for control to data plane communication is *OpenFlow*, created and maintained by *Open Networking Foundation*[3]. The *OpenFlow* protocol allows to define simple actions (e.g. drop packets or send packets specific ports), but it is also possible to build more complex policies [8]. *SDN* and *OpenFlow* concepts can also be applied to PON networks [9] , vehicular networks [10] or multimedia transmission over *HTTP* [11].

The research in the area of SDN results in many new concepts and technologies being developed. Traffic patterns for new applications can be highly distributed and can require extremely short response times [12], [13]. As the result, various solutions regarding network topologies have been proposed [14], [2]. However, there is no consistent framework that would facilitate early stage performance evaluation of SDN environments designed to host new applications. The framework presented in this article aims to fill the gap.

Because new control plane developments are typically implemented using open controllers and new data plane developments use virtual switches, this section overviews the respective categories.

Controllers. As *SDN* and *OpenFlow* gain popularity, many *controllers* appear on the market. In most cases, they are distributed as open source projects. Table I lists some of the

[3]https://www.opennetworking.org/

controllers along with their core technologies, the number of contributors and popularity indicators. The latter are expressed by the number of cloned repositories (table column *Forks*) and the number of people which found the project interesting (table column *Stars*). Based on this information (and on the last activity times) one can estimate the potential and size of a particular developing community.

Name	Language	Devs	Stars	Forks	Last activity
Ryu	Python	60	457	339	recently
Floodligth	Java	58	313	323	recently
POX	Python	16	282	285	2 years ago
ONOS	Java	74	147	132	recently
Trema	Ruby	18	238	77	recently
OpenDaylight	Java	66	82	114	recently
NOX	C++	6	75	66	one year ago
OpenMul	C	2	18	8	recently
Beacon	Java	3	8	3	4 years ago
LOOM	Erlang	22	26	8	7 months ago

TABLE I: Popularity of OpenFlow controllers (based on *GitHub* data).

In many cases *controllers* are in fact entire *SDN platforms* (they are often referred to as *monolithic controllers*). They consist of many components which deliver rich functionalities to users, such as topology discovery or security analysis tools. *OpenDaylight*[4], the leading production controller, is a representative of this group. On top of the controller developers write single, monolithic network applications. They use supported languages, APIs, and libraries provided by the framework, compile the entire platform and run the created application as a single process. Note however that an error in any part of the application can have adverse effects on the entire system.

Erlang community promises to deliver extensible, robust *OpenFlow* controllers based on the *LOOM* [7] framework. The main goal is to make the controller scalable and distributed (that is the main rationale for using *Erlang* virtual machine). *LOOM* has a modular and layered design. In control plane it consists of *Network Execution* and *Application* layers. LOOM-based controllers share core libraries, such as low-level library for encoding and decoding *OpenFlow* messages, an abstract interface to *OpenFlow* switches and a driver with a common base for *OpenFlow* controllers. LOOM developer creates a controller (specific to his needs) that can be later orchestrated by the network execution layer. Such approach clearly differs from monolithic controllers.

Virtual switches. When it comes to switches supporting *OpenFlow*, there are many hardware and software products on the market. However, from the perspective of early stage testing, software ones are the most important, because it is much easier to build test environment based on software switches (and larger environments can be prepared more easily). Moreover, because software switches often implement the latest version of the *OpenFlow* standard, new features can be used as soon as a reference implementation is available. Software switches are also widely used in production environment (e.g., *VMware NSX* [3]), so their evaluation is also valuable.

Erlang community (together with *Infoblox* and *Erlang Solutions*) is currently working on *LINC-Switch*[5]. It is run in operating system user space to facilitate usage flexibility and quick development. *LINC-Switch* is used as a simulator of an optical network, in one of *Open Network Operating System (ONOS)*[6] controller use cases. *LINCX*[7], which alse comes from Erlang community, is a new, faster version of *LINC-Switch*.

In the presented framework we use *Open vSwitch* because of its popularity and implementation quality. The switch supports multiple protocols and network standards. Moreover, because *OVS* is integrated with *Mininet*[8], it is easier to build test environment based on the switches distributed among multiple physical servers.

III. Related work

There are a few noteworthy research efforts in the area of measuring SDN controllers performance that can be leveraged by new ones, especially to identify the metrics to be measured and key test environment parameters. Shalimov et al. [15] measured latency, and throughput of controllers based on (1) the number of cores the controller uses, (2) the number of switches connected to the controller, and (3) the number of hosts in the network. For their tests, they created their own *Haskell* emulator of *OpenFlow* switches - *hcprobe*[9].

Similar work [16], evaluates performance of two Java-based controllers: *OpenDaylight* and *Floodlight*. The authors used a cluster of hosts running *Cbench*[10]. The tool directly stresses a controller by sending OpenFlow Packet-In messages. Another study [17], describes metrics of *ONOS*, including *Flow installation throughput*. The metric shows a number of flow mods, which can be sent by the SDN control plane in response to requests coming from applications or network events.

While all the mentioned characteristics were studied using network emulators, authors of the [18] proposed a methodology that utilizes network virtualization. To interconnect virtualized hosts, *Mininet* was used. They evaluated performance, as the number of Packet-In messages (requesting installation of new flows), a controller can handle per second.

As an example of more structured and holistic approaches, consider *IETF* draft describing the methodology for reporting SDN controller performance [19]. The document touches three aspects, namely: the number of switch sessions a controller can handle, the network size (number of nodes, links, and hosts) a controller can discover, and forwarding table capacity.

[4]https://www.opendaylight.org/

[5]http://flowforwarding.github.io/LINC-Switch/
[6]http://onosproject.org/
[7]https://github.com/FlowForwarding/lincx
[8]http://mininet.org/
[9]https://github.com/ARCCN/hcprobe
[10]https://goo.gl/CEgQ68

No performance tests are covering modular controllers (especially written in Erlang), although the authors of [15] mentioned that they examined *FlowER*, and it was not ready for evaluation under heavy workloads. Nonetheless, Erlang is designed to be used in the telecommunication industry and seems to be well suited for SDN use cases. Therefore, we decided to focus on controllers developed upon *LOOM* framework. To our best knowledge, this article is the first to present any results on Erlang SDN/OpenFlow controller performance testing.

Additionally, the presented framework aims to facilitate profiling of network topologies in context of specific applications (which generate specific traffic) by enabling the developer to define the virtual network topology to be constructed by the framework, and to deploy the applications in the network.

IV. Architecture of solution

Before starting to test performance of a *SDN* controller, it should be decided what to measure, and how to perform the measurements, i.e., what kind of environment to use to generate load against the controller, how to collect the data, and how to interpret it. This article proposes a method that follows a generic methodology consisting of the following steps:

1) Deciding upon testing objectives.
2) Building the test environment.
3) Configuring the test system.
4) Running tests and collecting data.
5) Interpreting the test results.

The following subsections refer to the steps in more detail.

A. Testing objectives

For measuring controllers performance two characteristics seem to be most important: *latency* and *throughput*. Latency describes an average amount of time that is needed to process a request. Although *OpenFlow* controllers have to be able to process various requests, those related to topology changes and new traffic flows are crucial. Regarding networking devices configuration, when a new traffic flow occurs, a controller has to handle *Packet-In* requests. Handling them, from a controller perspective, usually boils down to installing flow entries in the data plane device that originated the Packet-In, in order to instruct it what to do with packets belonging to the new flow. Note that latency in processing a Packet-In depends on the kind and context of a particular request, so the values measured can be different for particular application, network topology, etc.

Throughput is a metric describing how many requests a controller can handle, on average, in a given period. The requests related to receiving an unrecognized traffic flow may be sent simultaneously from multiple data plane devices to a single controller. Thus, this metric is sensitive to the characteristics of the network, i.e., its topology, number of switches and hosts, etc.

Regarding the controller itself, the following parameters affect its performance (i.e., both latency and throughput): number of cores used by the controller, the amount of memory used by the controller, the number of instances of the controller (if it can operate as a distributed system), and workload distribution strategy.

The framework presented in the article provides means for building test environments that enable the user to manipulate all the mentioned variables and measure metrics including, but not limited to, latency and throughput.

B. Building the Test Environment

There are three generic approaches to building an environment that mimics a real network generating Packet-In messages:

- *Building an SDN network from hardware.* Testing a controller with a real network can provide most accurate results. On the other hand, the approach is expensive, not flexible, and not scaling well. Additionally, it is not easy to gather statistics from all the devices on the network.
- *Emulating an SDN network by using specialized software.* There are switch emulators designed for testing SDN networks, e.g., *Cbench* and *Hcprobe*, which offer easy configuration and automation of test scenarios. It is relatively easy to scale the environments based on emulators – it usually involves adding new servers/virtual machines with emulator instances. This approach has been used in evaluation of *Network Management Systems* [20]. Note however, that emulators are hard to synchronize across instances, making it more difficult to coordinate test scenarios, and to gather and analyze the results.
- *Using network virtualization to simulate an SDN network.* In this approach, a tool such as *Mininet* can be used to virtualize a complex network on hardware hosts by using Linux containers interconnected with software switches. Such a tool allows for easy automation, controlling traffic generation and gathering statistics.

The presented list is not exhaustive. It is easy to imagine an approach that combines all of the above. It is also possible to test scalability using discrete-event network simulator such as *NS-3*[11].

For the implementation of our framework, We chose virtualization-based approach because of the ease of test automation and low cost.we chose virtualization-based approach. Figure 1 presents essential components of a test environment that could be developed with the proposed test framework. It outlines three main modules, as well as the data channels between them.

The core elements of the *Network Module* are those related to network simulation: *OpenFlow switches*, *hosts*, and *topology*. To enable the user to reproduce arbitrarily designed topologies, the framework uses *Mininet Cluster Edition*, and *Open vSwitch*. The emulated network characteristics can be very close to what would be observed in production networks also because the *Mininet* hosts come with a full implementa-

[11] https://www.nsnam.org/

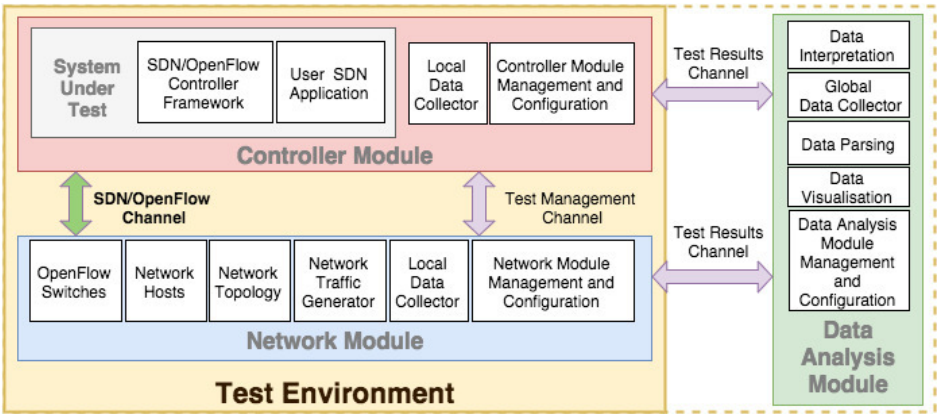


Fig. 1: Test Environment Components

tion of OSI protocol stack, so arbitrarily chosen applications can be used for testing, and act as *traffic generators*.

Local Data Collector is a component responsible for collecting and persisting tests-related data, like traffic metrics or counters of OpenFlow messages sent by the switches (i.e., the load metrics). The *Test Results Channel* indicates the likely transfer of collected data to the *Data Analysis Module*. The *Network Module Management and Configuration* component's task to facilitate automatization of *Network Module* set up and its coordination with the *Controller Module* via the *Test Management Channel*.

The *Controller Module* module encapsulates the *SDN/OpenFlow Framework* and *User SDN Application* components, that together form a fully capable SDN controller, which is the *System Under Test*(SUT). Because the framework is intended to be used mainly with modular, Erlang-based SDN controllers, a natural choice was to leverage the LOOM framework as the basis for SUT implementations, so that custom controllers can be easily built into the framework. The *Local Data Collector* and *Management and Configuration* components have similar responsibilities as their counterparts from the *Network Module*. Similarly to the *Network Module* case, the *Test Results Channel* indicates the likely transfer of collected data to the *Data Analysis Module*.

The *Data Analysis Module* is related to data produced during the tests: gathering, parsing, presenting and interpreting information. Depending on a particular test scenario, this component could be placed either inside the environment (in the case of runtime analysis) or outside (in the case of offline analysis). *Global Data Collector* is a component that gathers all the metrics from other modules so that they can be processed further in a consistent way. *Data Parsing* and *Data Visualization* components are designed for presenting the data in readable formats. *Data Interpretation* denotes additional tools that facilitate interpretation processes. *Data Analysis Module Management and Configuration* component is intended for management and configuration of the data analysis process.

C. Configuring the Test System

The configuration of a test environment can result in changes of traffic patterns. The key network configuration parameters include number of switches, number of hosts per switch, topology, and deployed applications.

Regarding the controller, there are usually less parameters to change. However, changing the hardware specification (e.g., memory, CPU), setting some options at the controller level (e.g., the number of cores it can use) or choosing the controller system structure option (centralized, distributed) have to be considered.

D. Data Collection and Interpretation of Results

Metric probes (i.e., values read from a given metric) can be taken at different intervals, cover different time spans, and have various aggregation rules. For example, when measuring load as the number of Packet-In requests, one needs to decide on how often the value will be measured, what time it will cover (e.g., last 10 minutes), and how it will be aggregated (e.g., maximum, average). Those decisions have to be taken keeping in mind the planned test scenarios.

In the presented framework, the core of the component responsible for collecting test data is implemented by using the *exometer*[12] package - it is run as a user application dependency in the same Erlang Virtual Machine. The package allows for instrumentation of Erlang code, so that data reflecting system performance can be exported to a variety of monitoring systems.

The following *exometer* metrics are implemented:

Application Packet-In Handle Time: histogram metric that captures elapsed time it takes for the SDN application (part of the SUT) to handle a single Packet-In request.

LOOM Packet-In Handle Time: similar as above, but time for the whole SUT (i.e., *LOOM* framework and SDN application) is captured.

Packet-In Count: counts Packet-In messages that are processed by the SDN application.

[12]https://github.com/Feuerlabs/exometer

Packet-Out Count: counts Packet-Out messages that are sent out by the SDN application.

Flow-Mod Count: counts Flow-Mod messages sent out of the SDN application.

In order to make updates to the **LOOM Packet-In Handle Time** metric, the *LOOM* framework code had to be instrumented. In the proof of concept implementation it is assumed that **each Packet-In message has a corresponding Packet-Out message**. Based on that, each Packet-In message that arrives to the controller framework is marked with a timestamp just after being decoded. Then, the mark is copied into the corresponding Packet-Out message. When the message is about to be encoded before sending, the mark is retrieved and the metric value is computed. Note that such instrumentation would not be possible with closed controller code.

The value of **Application Packet-In Handle Time** is based on the same assumption. The first timestamp is made when a Packet-In message enters the application process, and the second when Packet-Out and possibly Flow-Mod messages are sent.

The presented set of metrics is not closed and can be extended by metrics specific for a particular controller. To make use of the metrics collected during a test run, appropriate subscriptions and reporters need to be configured. Each reporter reads values from a metric at given interval and writes them to a file. The values are then processed by the *Data Analysis Module*. Depending on test scenarios' specifics, data can be analyzed (and, e.g., visualized) at run time or saved for later reference.

V. Framework Evaluation

In order to prove the framework usability and test whether it fulfills the requirements, a learning switch application was developed. The switch (called later *LOOM Switch*), was built upon the *LOOM Framework*. Together with the framework it formed a simple, but fully functional, SDN controller. The basic design decisions regarding the *LOOM Switch* were as follows:

1) There is no topology detection service: *LOOM Switch* fills in the forwarding tables based on the standard MAC addresses learning algorithm.
2) There is exactly one forwarding table (FT) for each data plane switch connected to the controller.
3) A packet that does not match any of the flow table entries is buffered in the data plane switch and its portion is sent to the controller using a Packet-In message.
4) At most one flow table entry is installed in the data plane switch that sent a Packet-In message.
5) Flow tables of other data plane switches are not changed.

Figure 2 depicts an example test setup. It demonstrates how the most important parts of the test environment are related to each other. Thick solid lines represent network links. Arrows show OpenFlow channels, connecting each of the switches to the *LOOM Switch* controller. Test management and data collection components were omitted for clarity.

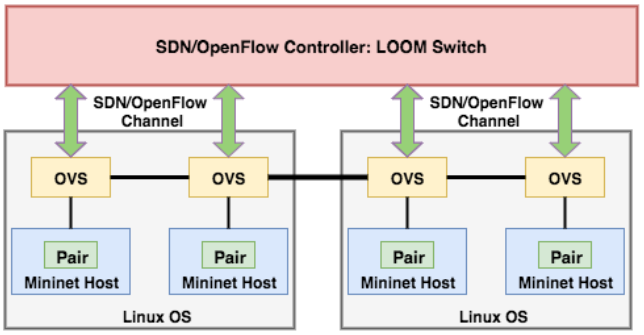


Fig. 2: An example test setup

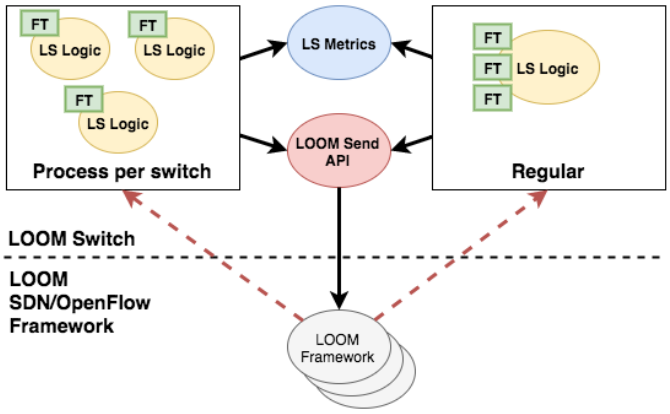


Fig. 3: Options of LOOM Switch Deployment

The network used for proof-of-concept testing was built from a number of switches that formed a linear topology, each of which had an even number of hosts attached. Network traffic was generated by a simple application called *Pair*, that was deployed on *Mininet hosts*. The application was sending out a UDP datagram to a random counterpart, receiving a reply, changing its configuration (including MAC address) and flushing its *ARP* table. Such an application provided an evenly distributed, constant load during the test runs.

A. Testing scenarios

Because Erlang is a highly concurrent functional language, and the presented framework supports the specifics of Erlang LOOM, it was a natural choice to evaluate two options of *LOOM Switch* deployment. In the first, called later *regular*, a single Erlang process handled requests coming from all the switches. In the second, called later *process per switch*, each switch had its dedicated process. The intention was to compare performance offered by the deployment options.

Figure 3 shows the *LOOM Switch* deployment architectures from the perspective of Erlang processes, for *LOOM Switch* application instance that serves 3 switches. The dashed arrows in the figure represent invocations of *LOOM Switch* callbacks from the LOOM framework.

In the case of *process per switch* deployment, each forwarding table (implemented as a hash map) is associated with

a separate Erlang process. In the *regular* deployment case, all the tables are accessed from one process. Regardless of the deployment option, *LOOM Switch* uses a single process for asynchronously handling metrics (**LS Metrics**) and for sending messages to the switches via *LOOM Send API*.

B. Experimental results

We conducted framework proof-of-concept tests for configurations in which the controller (i.e., *LOOM Switch*) used 2, 4 or 8 cores. As a representative test case, we chose the results gathered for 4 cores serving the *LOOM Switch*, run in the *process per switch* deployment option.

Figures 4a, 4b, 4c and 4d contain plots with metrics obtained during respective test runs. Each graph is labeled with *(X, Y)*, where X is a number of switches in the (linear) topology and Y is a number of hosts per switch. The overall number of hosts (240) was constant across all the test runs, and the same number of traffic generator (i.e., *Pair*) instances was used, so the performance was expected to be dependent only on the deployment option and the number of switches used. Note that the same number of hosts and conversations resulted in different loads in terms of Packet-In requests due to the assumptions regarding LOOM Switch and the varying number of data plane switches in the topology.

The following metrics are presented:

- **Application Packet-In Handle Time**, i.e., SDN application (part of the controller) latency, called later **LS Time**,
- **LOOM Packet-In Handle Time**, i.e., SDN controller latency, called later **LOOM Time**.

The collected values correspond to a 20 minute period of a stable load. To get a time frame in which the load is stable, the measurement was started after 15 minutes of test execution. The values were sampled every minute.

The difference in values of **LS Time** and **LOOM Time** is quite impressive. Average **LS Time** values, indicating the time it takes *LOOM Switch* to handle a request are becoming lower and less significant in relation to **LOOM Time**, which represents the overall time needed to serve a request by the controller (which consists of the *LOOM Switch* and the LOOM framework). In percentages, they take $38.7\,\%$, $9.8\,\%$, $1.7\,\%$, and $0.6\,\%$ of **LOOM Time**, respectively.

The phenomenon of decreasing values of **LS Time** metric can be explained as follows. In the *process per switch* deployment option, computations related to switching decisions are spread across the available cores, because each switch is served by a separate process. Along with the increase of Packet-In messages rate the *LOOM* framework gets overloaded and slower in serving requests, in comparison to the *LOOM Switch* application. As a result, as the framework needs more time to process the messages, and they are delivered at a lower pace to the *LOOM Switch*. Consequently, a message spends less time being processed by *LOOM Switch*, because it is able to process it instantly, since a major part of the messages is buffered (and stuck) in the framework.

C. Comparison of deployment options

In the presented test case of *process per switch*, four cores deployment scenario, the LOOM framework was observed to be a bottleneck. As presented in section V-B, the **LS Time** constituted only a few precent of the whole **LOOM Time**, which clearly denotes that the Packet-In messages were stuck in the framework. The highest load measured was about $228\,000$ Packet-Ins per minute with average times of $3\ \mathrm{ms}$ and $572\ \mathrm{ms}$ for **LS Time** and **LOOM Time** metrics, respectively.

Similar values of **LOOM Time** ($580\ \mathrm{ms}$) were observed in case of the *regular* deployment for the load of 73000 Packet-Ins per minute. In that case, the values of **LS Time** and **LOOM Time** metrics were almost equal, indicating that most of the processing time was consumed by the *LOOM Switch* application. Its pace of serving requests was significantly slower than the pace of request delivery performed by the *LOOM* framework.

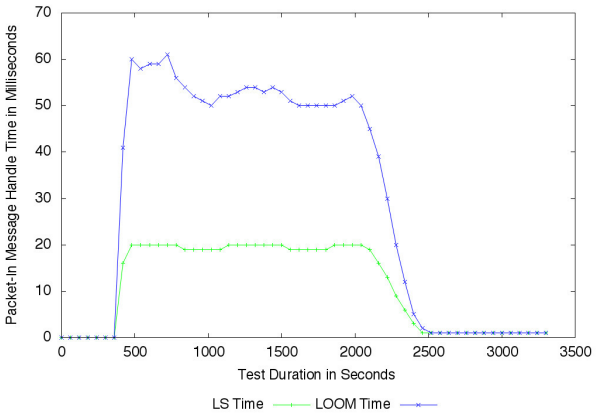
From the results gathered from two test sets for the two *LOOM Switch* deployment options, it is clear that the way of deploying a modular SDN controller has a tremendous impact on its performance. The same hardware configuration, 4 cores per controller, yielded about three times better processing capability in the *process per switch* case than in the *regular* one while offering the same processing time. Note that the processing time values cannot be directly compared with results presented in other articles, because it was measured for the maximum number of Packet-In messages that the controller was able to process. As shown in [5], the processing time grows rapidly after excessing a certain threshold, mainly due to request queuing.

In the following section, we present an experiment that was conducted to check whether adding more cores to the LOOM controller improved the situation, i.e., lowered the processing time for the maximum load identified for 4 cores.

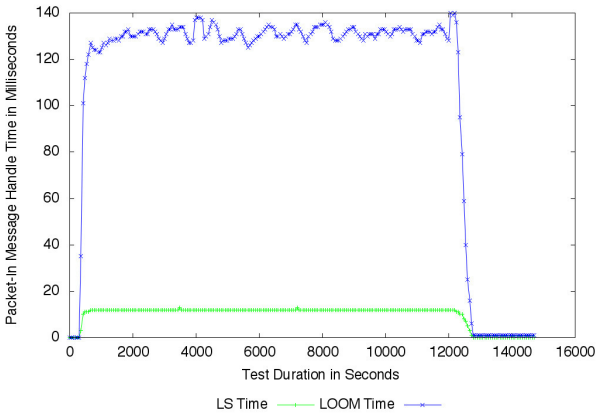
D. Scalability Test for Process Per Switch Deployment

Figure 5 presents a plot for the tested controller running in the *process per switch* mode, that depicts the relation between controller performance and the number of cores available for the schedulers. The plot was created based on experiments conducted on the same topologies that were used for comparing deployment options. Each test scenario (denoted (X,Y) on the X axis) has three values for three corresponding controller configurations that vary only in the number of used cores. Each point represents an average value of the **LOOM Time** metric measured in a given test scenario and controller configuration. Note that the experiment was conducted under heavy load (228000 Packet-In messages per minute processed), which proved to be the limit of processing capability of the tested controller running on 4 cores.

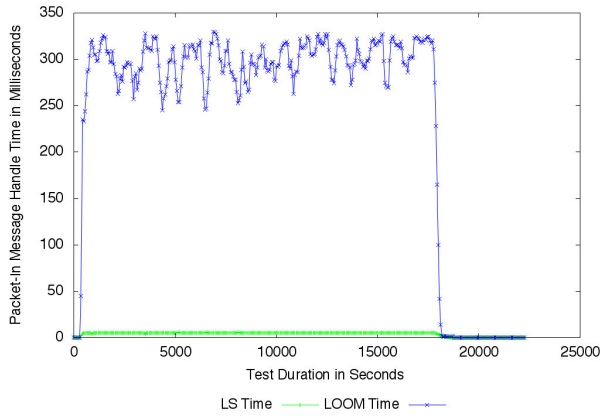
The most important conclusion from comparing the results depicted in figure 5, is that while switching from 2 cores to 4 gives a significant increase in performance in all test scenarios, such situation does not happen when doubling the number of cores again, from 4 to 8. In the case of 5 schedulers the controller performed even a bit worse. As a consequence, it



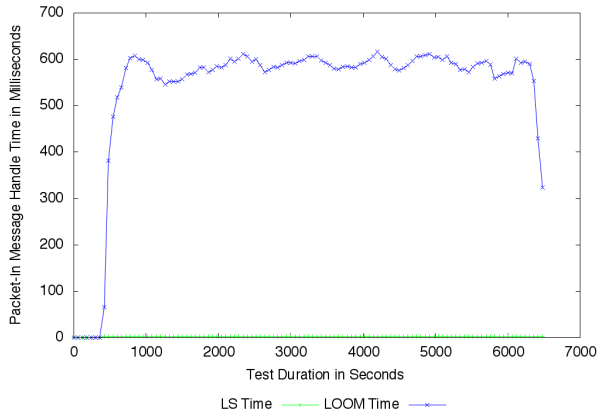
(a) Test Scenario (5, 48);
61000 Packet-Ins/minute



(b) Test Scenario (10, 24);
117000 Packet-Ins/minute



(c) Test Scenario (15, 16);
172000 Packet-Ins/minute



(d) Test Scenario (20, 12);
228000 Packet-Ins/minute

Fig. 4: Handle Request Time in Function of Time for Different Test Scenarios

should be stated that the bottleneck (for 8 cores) does not result from insufficient computing power, but rather from the LOOM framework implementation.

The reason that *LOOM* running *LOOM Switch* in the discussed deployment option scales well only up to the certain point (to 4 cores given the presented results), may be related to *locks*. Some processes in *LOOM* may be acquiring the same *lock*. The more cores available to the system, the more processes can execute simultaneously. As a result, there could be more failed attempts to acquire a particular *lock*. Of course, there could be other potential reasons but checking the locks seems to be a good starting point in searching for an explanation.

VI. Conclusions

In the article, we presented our research on the development of an innovative framework for measuring the performance of SDN controllers. As a part of the presented research, we built a test environment to evaluate the framework practically by testing specific features of modular SDN controllers. The presented results are promising, and form a good base for further development and analysis. We plan to extend our environment with different traffic generators to emulate various application workloads (e.g. multimedia sessions, big data processing). It will allow measuring performance of SDN controllers in real scenarios, with real applications and with different traffic conditions.

The framework presented in this article is designed for assessing various network topologies, and controller deployments, with arbitrarily chosen applications generating network traffic. Such a tool is needed, e.g., for early stage testing of a new network-intensive application that creates network traffic with certain characteristics. Using the framework presented in the article one can define both the network topology and traffic patterns that run on top of it as well and observe the effects of modifying multiple attributes on the performance of the controller.

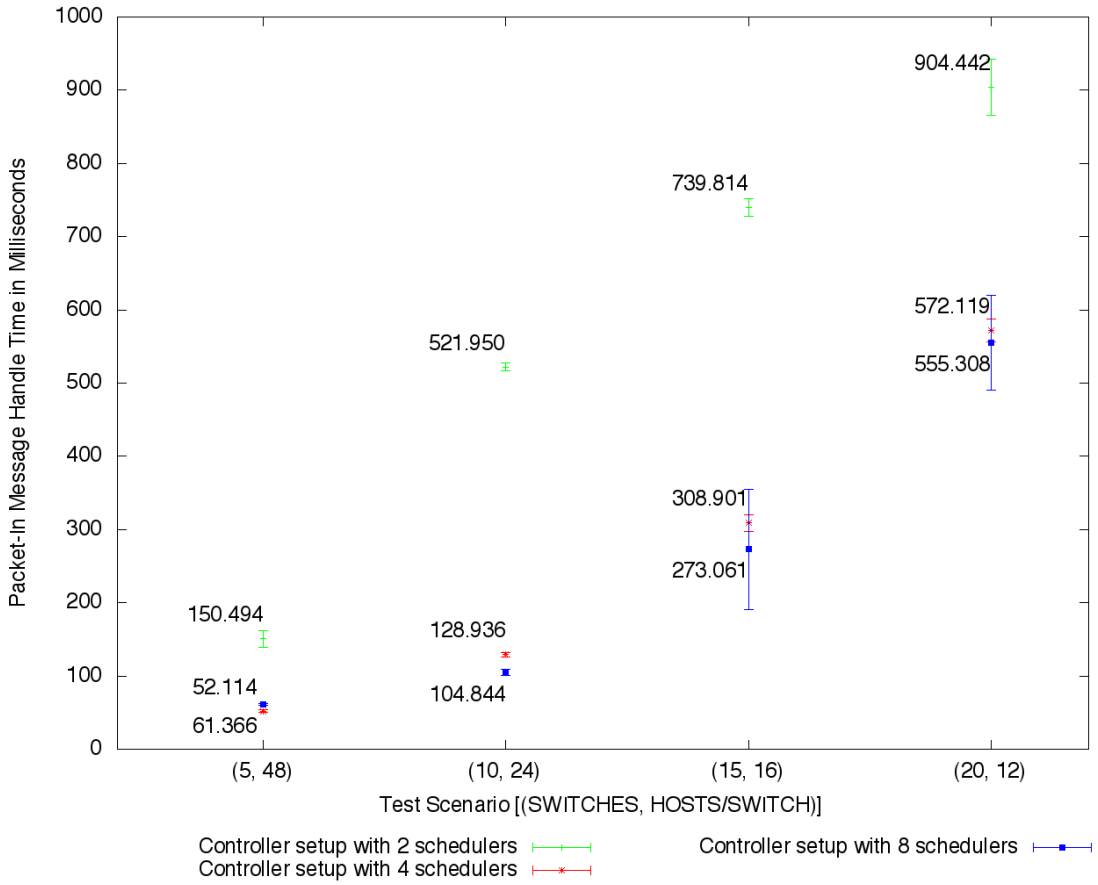


Fig. 5: *LOOM* Handle Request Time in Different Test Scenarios

VII. Acknowledgments

The research presented in this paper was partially supported by the National Centre for Research and Development (NCBiR), Poland, project PBS1/B9/18/2013 and AGH statutory research grant no. 11.11.230.124.

References

[1] S. Hassas Yeganeh and Y. Ganjali, "Kandoo: a framework for efficient and scalable offloading of control applications," in *Proceedings of the first workshop on Hot topics in software defined networks*. ACM, 2012, pp. 19–24.

[2] A. Greenberg, J. R. Hamilton, N. Jain, S. Kandula, C. Kim, P. Lahiri, D. A. Maltz, P. Patel, and S. Sengupta, "Vl2: a scalable and flexible data center network," in *ACM SIGCOMM computer communication review*, vol. 39, no. 4. ACM, 2009, pp. 51–62.

[3] T. Koponen, K. Amidon, P. Balland, M. Casado, A. Chanda, B. Fulton, I. Ganichev, J. Gross, P. Ingram, E. Jackson *et al.*, "Network virtualization in multi-tenant datacenters," in *11th USENIX Symposium on Networked Systems Design and Implementation (NSDI 14)*, 2014, pp. 203–216.

[4] T. Koponen, M. Casado, N. Gude, J. Stribling, L. Poutievski, M. Zhu, R. Ramanathan, Y. Iwata, H. Inoue, T. Hama *et al.*, "Onix: A distributed control platform for large-scale production networks." in *OSDI*, vol. 10, 2010, pp. 1–6.

[5] A. Dixit, F. Hao, S. Mukherjee, T. Lakshman, and R. Kompella, "Towards an elastic distributed sdn controller," *ACM SIGCOMM Computer Communication Review*, vol. 43, no. 4, pp. 7–12, 2013.

[6] T. Szydlo and K. Zielinski, "Adaptive enterprise service bus," *New Generation Computing*, vol. 30, no. 2-3, pp. 189–214, 2012.

[7] LOOM, http://flowforwarding.github.io/loom/.

[8] D. Jullier, M. Konieczny, and S. Zieliński, "Applying software-defined networking paradigm to tenant-perspective optimization of cloud services utilization," in *Computer Networks*. Springer, 2015, pp. 193–202.

[9] P. Parol and M. Pawlowski, "Towards networks of the future: Sdn paradigm introduction to pon networking for business applications," in *Computer Science and Information Systems (FedCSIS), 2013 Federated Conference on*. IEEE, 2013, pp. 829–836.

[10] I. Stojmenovic and S. Wen, "The fog computing paradigm: Scenarios and security issues," in *Computer Science and Information Systems (FedCSIS), 2014 Federated Conference on*. IEEE, 2014, pp. 1–8.

[11] C. Cetinkaya, Y. Ozveren, and M. Sayit, "An sdn-assisted system design for improving performance of svc-dash," in *Computer Science and Information Systems (FedCSIS), 2015 Federated Conference on*. IEEE, 2015, pp. 819–826.

[12] V. Jalaparti, P. Bodik, S. Kandula, I. Menache, M. Rybalkin, and C. Yan, "Speeding up distributed request-response workflows," in *ACM SIGCOMM Computer Communication Review*, vol. 43, no. 4. ACM, 2013, pp. 219–230.

[13] A. Roy, H. Zeng, J. Bagga, G. Porter, and A. C. Snoeren, "Inside the social network's (datacenter) network," in *Proceedings of the 2015 ACM Conference on Special Interest Group on Data Communication*. ACM, 2015, pp. 123–137.

[14] M. Al-Fares, A. Loukissas, and A. Vahdat, "A scalable, commodity data center network architecture," *ACM SIGCOMM Computer Communication Review*, vol. 38, no. 4, pp. 63–74, 2008.

[15] A. Shalimov, D. Zuikov, D. Zimarina, V. Pashkov, and R. Smeliansky, "Advanced study of sdn/openflow controllers," in *Proceedings of the 9th Central and Eastern European Software Engineering Conference in Russia*. ACM, 2013.

[16] Z. K. Khattak, M. Awais, and A. Iqbal, "Performance evaluation of

opendaylight sdn controller," in 20th IEEE International Conference on Parallel and Distributed Systems (ICPADS), 2014.
[17] Open Network Operating System, "Raising the bar on sdn control plane performance and scalability," http://goo.gl/AizqcC, 2015.
[18] M. P. Fernandez, "Evaluating openflow controller paradigms," in IEEE 27th International Conference on Advanced Information Networking and Applications, 2013.
[19] B. Vengainathan, A. Basil, M. Tassinari, V. Manral, and S. Banks, "Benchmarking methodology forsdn controller performance," Working Draft, IETF Secretariat, Internet-Draft draft-bhuvan-bmwg-sdn-controller-benchmark-meth-01, July 2015. [Online]. Available:https://tools.ietf.org/html/draft-bhuvan-bmwg-sdn-controller-benchmark-meth-01
[20] K. Grochla and L. Naruszewicz, "Testing and scalability analysis of network management systems using device emulation," in Computer Networks. Springer, 2012, pp. 91–100.

3rd Workshop on Emerging Aspects in Information Security

ADMITTEDLY, information security works as a backbone for protecting both user data and electronic transactions. Protecting the communication and data infrastructure of an increasingly inter-connected world has become vital nowadays. Security has emerged as an important scientific discipline whose many multifaceted complexities deserve the attention and synergy of the computer science, engineering, and information systems communities. Information security has some well-founded technical research directions which encompass access level (user authentication and authorization), protocol security, software security, and data cryptography. Moreover, some other emerging topics related to organizational security aspects have appeared beyond the long-standing research directions.

The Emerging Aspects in Information Security (EAIS'16) workshop focuses on the diversity of the information security developments and deployments in order to highlight the most recent challenges and report the most recent researches. The workshop is an umbrella for all information security technical aspects. In addition, it goes beyond the technicalities and covers some emerging topics like social and organizational security research directions. EAIS'16 is intended to attract researchers and practitioners from academia and industry, and provides an international discussion forum in order to share their experiences and their ideas concerning emerging aspects in information security met in different application domains. This opens doors for highlighting unknown research directions and tackling modern research challenges. The objectives of the EAIS'16 workshop can be summarized as follows:

- To review and conclude researches in information security and other security domains, focused on the protection of different kinds of assets and processes, and to identify approaches that may be useful in the application domains of information security
- To find synergy between different approaches, allowing to elaborate integrated security solutions, e.g. integrate different risk-based management systems
- To exchange security-related knowledge and experience between experts to improve existing methods and tools and adopt them to new application areas
- To present latest security challenges, especially with respect to EC Horizon 2020

TOPICS

Topics of interest include but are not limited to:

- Biometric technologies
- Human factor in security
- Cryptography and cryptanalysis
- Critical infrastructure protection
- Hardware-oriented information security
- Social theories in information security
- Organization- related information security
- Pedagogical approaches for information security
- Individual identification and privacy protection
- Information security and business continuity management
- Decision support systems for information security
- Digital right management and data protection
- Cyber and physical security infrastructures
- Risk assessment and risk management in different application domains
- Tools supporting security management and development
- Emerging technologies and applications
- Digital forensics and crime science
- Misuse and intrusion detection
- Security knowledge management
- Data hide and watermarking
- Cloud and big data security
- Computer network security
- Security and safety
- Assurance methods
- Security statistics

EVENT CHAIRS

- **Awad, Ali Ismail,** Luleå University of Technology, Sweden
- **Bialas, Andrzej,** Institute of Innovative Technologies EMAG, Poland

PROGRAM COMMITTEE

- **AbdAllah, Mohamed Mostafa,** Yanbu Industrial College, Saudi Arabia
- **Banach, Richard,** University of Manchester, United Kingdom
- **Bun, Rostyslav,** Lviv Polytechnic National University, Ukraine
- **Clarke, Nathan,** Plymouth University, United Kingdom
- **Cyra, Lukasz,** DM/OICT/RMS (UN)
- **Dworzecki, Jacek,** Police Academy in Szczytno
- **Fernandez, Eduardo B.,** Florida Atlantic University, United States
- **Furnell, Steven,** Plymouth University, United Kingdom
- **Furtak, Janusz,** Military University of Technology, Poland

- **Geiger, Gebhard,** Technical University of Munich, Faculty of Economics
- **Grzenda, Maciej,** Orange Labs Poland and Warsaw University of Technology, Poland
- **Hämmerli, Bernhard M.,** Hochschule für Technik+Architektur (HTA), Switzerland
- **Hasssaballah, M.,** South Valley University, Egypt
- **Kalbarczyk, Zbigniew,** University of Illinois at Urbana-Champaign
- **Kapczynski, Adrian,** Silesian University of Technology, Poland
- **Klamka, Jerzy,** Polish Academy of Sciences
- **Kosmowski, Kazimierz,** Gdansk University of Technology
- **Krendelev, Sergey,** Novosibirsk State University
- **Mamojka, Mojmír,** Police Academy in Bratislava
- **Misztal, Michal,** Military University of Technology, Poland
- **Pańkowska, Małgorzata,** University of Economics in Katowice, Poland
- **Rot, Artur,** Wroclaw University of Economics, Poland
- **Soria-Rodriguez, Pedro,** Atos Research & Innovation
- **Stokłosa, Janusz,** Poznań University of Technology, Poland
- **Suski, Zbigniew,** Military University of Technology
- **Szmit, Maciej,** Orange Labs Poland, Poland
- **Thapa, Devinder,** Luleå University of Technology
- **Yen, Neil,** The University of Aizu, Japan
- **Zamojski, Wojciech,** Wrocław University of Technology
- **Zieliński, Zbigniew,** Military University of Technology, Poland

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 981–989
DOI: 10.15439/2016F490
ACSIS, Vol. 8. ISSN 2300-5963

Developing malware evaluation infrastructure

Krzysztof Cabaj, Piotr Gawkowski, Konrad Grochowski, Amadeusz Kosik
Institute of Computer Science
Warsaw University of Technology
ul. Nowowiejska 15/19
00-665 Warsaw, Poland
Email: {K.Cabaj, P.Gawkowski, K.Grochowski, A.Kosik}@ii.pw.edu.pl

***Abstract*—Malware evaluation is a key factor in security. It supposed to be safe and accurate. The contemporary malware is very sophisticated. Usually it uses complex distributed infrastructure an investigation of which is a very challenging task. In the paper, the development of the testbeds toward malware and its infrastructure evaluation is presented. Based on the real-life experience with the subsequent CryptoWall generations analysis, the MESS evaluation system is introduced. A rich set of analytical results is discussed. A new methods of visualization for malware artefacts analysis are given.**

I. Introduction

In the last decade the main motive of attackers' actions was associated with money. In the previous years the most precious treasures were credit cards numbers or data used for accessing to e-banking systems. However, it is worth mentioning that reaction to these threats from financial organizations made such attacks harder. From the last few years, more and more popular are attacks that lock victim's computers and demand some ransom for enabling access to the infected machines. Due to a ransom request, any malware used during these attacks is called ransomware. Reports prepared by antivirus companies show a huge increase in this kind of attacks in the last two years. For example, McAfee shows that only in the first quarter of the 2015 year, the number of observed ransmoware samples rose by 165% [1]. Symantec shows even more horrifying data - accordingly to its report number of ransomware which encrypts files in the hard drive rise almost 45 times, from 8274 samples observed in 2013 to the 373342 in 2014 [2].

At the end of March 2015 our security group had to clean-up an infected machine in the Institute of Computer Science. That malware sample (CryptoWall 3.0) was then examined using dynamic analysis. Performed analysis revealed very interesting behavior concerning the network activity of the ransomware. After the infection the sample contacts attacker's Command and Control server using a list of prior infected Web servers. We called this kind of a Web server a *CryptoWall proxy*. What should be emphasized, these servers are innocent victims, too. Analysis of few more CryptoWall samples showed that these lists of proxies contain many infected servers, and these lists are centrally managed by attackers. Detailed description of this network activity and results from its initial analysis can be found in [4].

During continuation of research many samples of CryptoWall family were soon gathered. Manual analysis of all samples soon became almost impossible. So, the ARTA system (Automatic Ransomware Traffic Analyzer) was develop and deployed. It consists of several modules. The ARTA has a dedicated subnet with a set of HoneyPots and a DNS redirecting the whole traffic to these HoneyPots. The malware sample is executed within the dynamic evaluation system Maltester (see [4]). Moreover, the whole system is remotely controlled with dedicated the Web application. Results and practical experience gathered with ARTA is presented in [9].

The advantages of using ARTA are really great. However, two things are missing (even if the lack of their presence should not be considered as drawbacks). First of all, in ARTA, the whole network traffic was enclosed within the HoneyPots. It is a safe solution and allows to identify the basic (i.e. initial) communication made by the malware. However, there is no further knowledge about the liveness of external parts of malware infrastructure (i.e. nodes it tries to communicate with). Also subsequent communication of the sample is not known (due to limitations of HoneyPots). Secondly, the Maltester is not designed to provide information about detailed actions taken by the sample within the target system. Maltester allows sample execution and comparison of system state only after the sample evaluation is finished. It is an environment for evaluation of malware in a Xen-based virtualized host by state comparison (only network traffic is monitored on-line from the outside).

There are some ready-to-go solutions, like Cuckoo Sandbox [10] - due to its popularity and availability it is common to meet malware samples that detect being executed in environment like this. Another popular system, Anubis (exposed as a service in web application) is no longer available as its developers has created their own company with malware analysis services. That is why it is important to build own solutions. Moreover, as the malware dynamically evolve, the evaluation infrastructure must be open for fast developing. So, we decided to develop and implement our own solutions.

In this paper we present the Malware Evaluation Support System - MESS - an environment based on Hyper-V virtualization that uses on-line, on-site monitoring of the malware activity. Comparing to Maltester, it delivers not only information about changes made by the malware but also how the execution proceeds. One of the main advantage on

978-83-60810-90-3/$25.00

MESS is the ability to remotely control the analysis process including the security settings of the network traffic. This capability was used in long-term experiment with the rich set of Cryptowall ransomware samples. One of the goals was to identify and observe the life-cycle of the so-called *proxy servers* being a vital part of the Cryptowall infrastructure. For the analysis we also propose a graph-based method that, in our opinion, facilitates identification of the most interesting artefacts of malware infrastructure. In the paper the MESS test-bed, methodology and results of different kinds of analysis are presented.

In the next section the basic differences between different kids of malware analysis are discussed. Section III generally presents the most contemporary malware type - ransomware. In more detail the behavior of a CryptoWall family ransomware is described. Then, the insights into the MESS test-bed are given in section IV followed by the description of experiments automation (section V). The paper presents the obtained results in section VI.

II. Malware analysis problems

Analysis of malware can be conducted in several ways. First of all, statically - with the analysis of the de-assembled/decompiled code. Generally, such analysis can be very effective. However, in the case of malicious software it can be challenging or even impossible: to cheat anti-virus scanners and to obstruct such analysis, the malware usually implements several obfuscating techniques [5], [6], [7]. For example, some techniques introduce dynamic code modifications (upon execution - decryption using XOR or ROT13 on some code blocks, garbage instructions overwritten with NOPs, return statements without previous calls).

In dynamic analysis, the black-box model is assumed - the examined application is analyzed through its behavior. Such analysis requires malware execution along with some dedicated monitoring utilities [6]. Gathered logs are then used to investigate actions taken by the malicious code on the host. Here, two main problems arise: how to safely execute malicious software and how to identify malicious behavior. Virtualization may address the problem of maltware sand-boxing. Another advantage is scalability - different malware samples can be examined in parallel and the guest system can be efficiently prepared for the evaluation using snapshot for state recovery. On the other hand, there is a risk of virtualization hypervisor disclosure [7]. One of the simplest way to identify virtualization is to check if the hard disk contains any user activity related files (e.g. changed desktop background, web browser temporary files). More sophisticated one is checking in the system registry for CPU information and verification of the number of threads with the declared by the CPU manufacture.

Creating an environment for malware evaluation a key issue is to assure security of other IT resources in the neighborhood. As the contemporary malware often requires Internet connection to be fully operable, the connectivity limitations may also significantly limit the analysis (e.g. unavailability of command-and-control servers, downloading of other malware). At the same time it is obvious that during our experiments we would like to protect our infrastructure as well as limit possible attacks made by the executed malware. So, in particular, the ports 25 and 587 should be blocked to not allow spamming over the Internet.

Dynamic analysis can be made on-line - the malware execution is monitored while its execution - or off-line - the analysis is based on the comparison of the system state before and after the execution.

In the second case, the analysis is mainly focused on changes in the file-system and system registry (i.e. new, deleted, renamed files and directories, registry entries). The main advantage is low probability of analysis disclosure but the temporal analysis is very limited. In fact, only the network traffic can be analyzed in details as it can be gathered from the outside of the host.

The most detailed information can be collected with the real-time monitoring of malware execution on the host itself. In this case the probability of monitoring disclosure by the malware is very high. Especially using debuggers can be easily identified - it interfere with the execution much more than other monitoring utilities [5], [7], [8].

III. CryptoWall Family Ransomware

The first generation of ransomware only locks access to the computer, preventing logging to the machine. For many skilled users these threats can be easily overcome. In the most severe cases full system reinstallation is needed. However, all user's data stored in the infected machine can be restored. Due to this fact, shortly, a second generation of ransomware become popular which works in more hostile fashion.

In its second generation the malware encrypts various types of files associated with user precious data generated by, for example, word processors, spreadsheets or games - yes, some ransomware encrypts games' saves files. As in most cases the ransomware uses modern encryption algorithms, like AES (Advance Encryption System), the decryption without the key is almost impossible. The first malwares of this generation utilized symmetric-key algorithms, which use the same key for encryption as well as for decryption. In effect, this key could be extracted from its poor implementation (key was not deleted after encryption of the whole data) or during its transfer from the victim's machine to the attacker.

The one of the most sophisticated ransomware family is called CryptoWall which uses an asymmetric-key encryption algorithm. Such algorithm uses two separate keys: public - used for encryption and private - used for decryption. In this situation both keys are generated somewhere in the Internet and only the public key used for encryption of users' data is transferred to the infected machine. Private key used for decryption never appears in the victims' machine. Considering that this malware uses 2048 bit RSA asymmetric-key algorithm, decryption of victim's data without the private key is unfortunately impossible. Detailed analysis of various ransomware families can be found in [3].

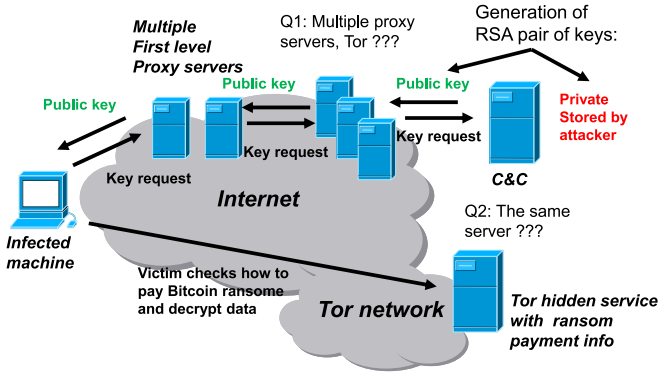


Fig. 1. CryptoWall infrastructure overview

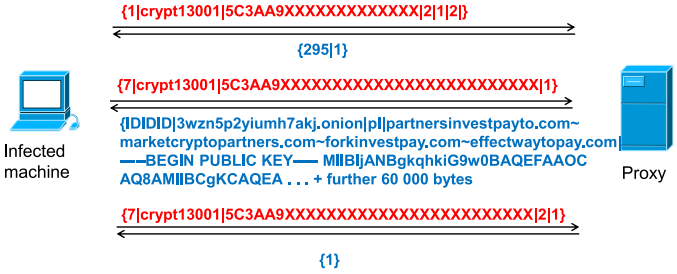


Fig. 2. CryptoWall 4.0 communication overview

The CryptoWall family, due to utilization of asymmetric cryptography, was one of the most sophisticated ransomware in the 2015 year. Usage of this type of cipher has big advantage in contrast to previous ransomware families that utilize symmetric cryptography - the key required for decryption is not present in infected machine at any moment of ransomware activity. However, the one disadvantage of this approach is a need of contact between a victim and an attacker's command and control server, which generates asymmetric key pair and provides a public key used for data encryption.

At the end of January 2015 a new version was observed - CryptoWall 3.0. This version uses infected web servers for hosting proxy script that hinder the location of attackers' command and control server. Detailed description of used communication protocol was presented in [4]. Fig. 1 presents CryptoWall infrastructure.

At the beginning of November 2015, the CryptoWall 4.0 came out. Despite the similar protocol (at first glance), we failed to decrypt its communication for more than a month. Fortunately, due to attackers' mistake, at 6th of December, one of the proxy servers, instead of the execution of the malicious proxy script, was sending its copy. Analysis of the script source code revealed that (in comparison to the previous versions) it has new four lines of code. This part of the code removes some random bytes added by the attacker at the beginning of the encrypted data (within the communication protocol). Probably, this change is introduced for hindering CryptoWall 4.0 activity from detection by Intrusion Detection System. The previous version uses messages that have almost the same length in particular communication phase during communication with proxy. Analysis of decrypted messages revealed a second change in comparison to the CryptoWall 3.0 - the protocol used is simplified. Instead of five transmissions, CryptoWall 4.0 exchanges only three. Decrypted communication of this CryptoWall version is presented in the Fig. 2.

The first transmission (message exchange) informs the attacker that a new machine is infected. This message contains the message of type one (see the first number in sent request - red color), the name of the used campaign (e.g. crypt13001 - see Fig. 2, the machine unique identifier and the encoded description of Windows operating system version. The next transmission (message with the type of 7) is responsible for downloading the public key and the personalized image presented to the victim. The last transmission confirms reception of all the data needed for the encryption process.

IV. Malware Evaluation Support System

Malware Evaluation Support System (called MESS later on) is a system toward dynamic monitoring in real-time of malware sample execution. Contrary to Maltester, the data collection is made on-the-target machine during the runtime.

A. MESS architecture

It consists of several components, as depicted in Fig. 3:

- Executor - responsible for malware sample execution and on-site monitoring tools. It resides on a target system on which the malware sample is executed - a virtual machine VMx
- NAT/Firewall - responsible for recording and filtering the network traffic to and from the Executor systems from/to the Internet
- Controller - responsible for coordinating the whole MESS infrastructure and interaction with the user
- Supervisor - responsible for controlling the Executors through the hypervisor management API.

All these components can be located on a single physical machine or on multiple virtualization servers. In the first, simplest scenario, the Controller and the NAT/Firewall can be located on a single virtual machine along with a set of Executor - separated virtual machines.

The Supervisor (in order to manage the virtual machines) has to be located within the physical host system. In more complex infrastructure (as in Fig. 3) MESS scales-up with the number of physical virtualizators (each requires its own Supervisor) and their virtualization capabilities (on each physical virtualization host a set of virtual machines VM1-VMx with the Executors can be used in parallel). Theoretically, MESS can utilize several different hypervisors (if a proper Supervisor component is available), however, at the moment only Microsoft Hyper-V is supported.

Within the MESS three networks are defined. The NAT/-Firewall machine has to have three network interfaces. The first one is connected to the Internet. The second one serves as a communication channel with the Controller. The third one

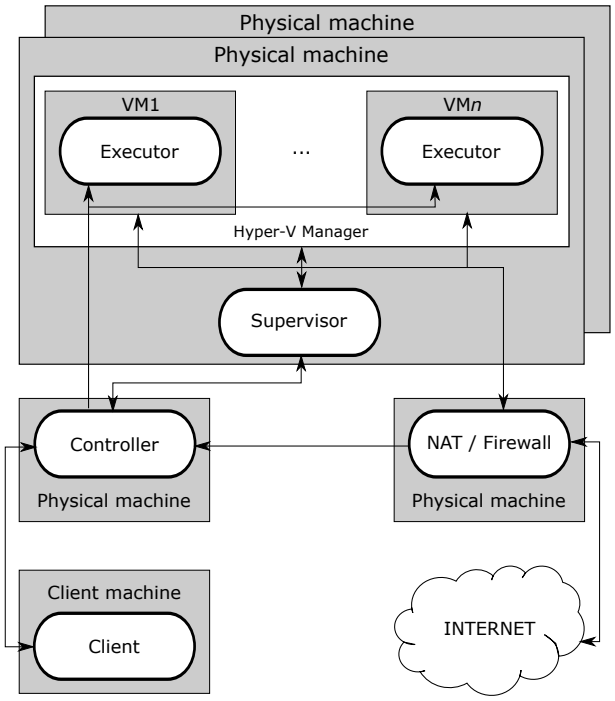


Fig. 3. MESS Architecture

serves the Internet connectivity for the target virtual machines with the Executors. However, due to security restrictions, that network is filtered. Executor's machine is not allowed to communicate with local network as well as with physical LAN (university network in our case). As some ports are commonly used to spread malware (e.g. through automatically sent spam) these ports are also disabled (e.g. 25, 587). During the analysis the settings of the firewall can be changed.

B. Usage scenarios

Typical usage scenario consists of several steps. First of all, a malware sample has to be executed within specially crafted virtual machine. That means, that a set of user-level applications with typical vulnerabilities and a desired set of operating system updates should be installed. A set of system services should be running as well as some user-level emulators (to emulate users' activity). After preparation of such environment, the virtual machine should be frozen in the snapshot. Later on, it will be rolled back to this state before each sample analysis.

A user has to choose at which target system snapshot he wants to conduct the analysis, as well as the malware sample and its filename at which it should be saved in the target machine. He can also pass a set of tools and scripts (in PowerShell) to be executed upon:

- before actual sample execution - some additional preparation tasks
- before restart of the virtual machine - sometimes an analysis might require to restart the machine
- before the end of the analysis - before the gathered results of monitoring utilities are prepared for sending to the user

If the chosen virtual machine is ready, the Supervisor rolls its state to the one saved in the chosen snapshot and runs it. Then, all the informations (scripts, additional monitoring tools are the malware sample) are transferred to the Executor component on the target machine. The monitoring tools are started and finally, the sample is executed - the actual analysis begins. During the analysis any user actions are not required, however, the user may request target machine restart, can interact with the target system (e.g. with remote desktop or console) or preview the network activity. The option to restart the machine is MESS unique feature. Some malware samples, do not start the main activity during the first execution - they just setup itself within the system. In such case, the system has to be restarted to observe the malware. In MESS, the target system will continue the analysis after such restart.

The sample analysis can end in several ways. In normal scenario the MESS is requested to stop the analysis. Then, the Executor on the target system stops the monitoring tools, executed proper user-defined scripts and the results (from embedded monitoring tools as well as the user's) are provided to the MESS as a ZIP file. That file is downloaded by the MESS and sent to the user. After this step, the target virtual machine is stopped and rolled back to its initial state, and ready for the next analysis. Sometimes the user may be not interesting in the results or some unusual circumstances occur (e.g. the sample become a part of DDoS attack or do other unpredicted actions). Then, the analysis is interrupted without result preparation and downloading.

By default, MESS uses Process Monitor tool from the Windows Sysinternals suite to register all the actions made by the sample [11]. It can trace events like system registry accesses (reads, writes of keys and values), filesystem operations and thread/process management. Moreover, the tool provides a rich GUI functionality for further data analysis in off-line. For dumping the network traffic, the Wireshark is used directly on the target system. However, the whole traffic can also be registered on the NAT/Firewall machine.

C. Component integration

Communication between MESS components is implemented with REST approach and XMLRPC protocols (both using HTTP beneath). The REST (Representational State Transfer) is used between the Supervisor(s) and the Controller. It was chosen because of its simplicity, very simple implementation on both, the client and the server side. Moreover it is independent to the implementation technology. It is very important aspect, as the Supervisors (as mentioned earlier) can be located on different kind of operating systems and environments.

Communication between the Controller and the Executors is implemented with XMLRPC (XML Remote Procedure Call). Functionally, it allows to implement remote-like procedure calls. All the parameters and results (even collections or binary files) are passed between the client and server sides very

convenient way. The external interface Controller (from the user side) is also implemented with XMLRPC. Thanks to that, the MESS can be easily integrated with external management systems or scripts.

In order to implement these quite complex tasks, the MESS components consists of subcomponents. For example, the Executor consists of a dedicated system service for keeping the analysis context between system restarts, HTTP service to provide convenient way of analysis result downloading and user-level application for sample startup.

D. Remarks

After several months we gained a lot of experience in using MESS. One of the biggest challenge is to properly prepare the target system environment and trim the monitoring tools. From the one side, the one is interesting in many details of the actions taking place but most of them are not anyhow related to the analysed malware activity. The target system or user-related applications may generate a lot of actions, because, for instance, automatic update services, background tasks etc. It has to be stressed that the Process Monitor can provide very rich set of data. That is great, but the result file might be huge. The proper filter may significantly limit the size of gathered data. We faced these problems as we missed (by oversight) that the Opera browser within the target system was left in the snapshot in the state just before the update. In effect, soon after each analysis, the Opera was starting update downloading and installation. That introduced additional traffic (ca. 40MB) and a lot of system actions (registry and filesystem related).

To properly conduct an analysis the target system should be evaluated in different configurations. Using only updated system might be "too good" for the analysed malware. On the other side, using "too old" version of the system (like WindowsXP, or very outdated version) might be useless (unusual configuration) or suspicious for the malware sample (see section II.

As the MESS utilize Hyper-V technology, there is a risk that the malware sample will detect the presence of the hypervisor. Typically, such sample simply terminates the execution without any malicious actions. Generally, we have met such situation only for a few samples. Only one sample has detected Hyper-V. It was stressfully analyzed in Maltester then. The other sample has detected the Xen hypervisor of Maltester. In this particular case, the sample executed in MESS, after some operations suggesting hypervisor detection procedure, failed to detect Hyper-V and was successfully analysed. These cases proves that both solutions are complementary.

V. Experiment Automation

Manual analysis of each available sample in attempt to retrieve all active servers utilized by it proves to be tiresome and time consuming process. Fortunately preliminary results of manual analysis helped developing tools which used available experiment environment capabilities to automate the process.

Algorithm 1 briefly presents automated experiment procedure. For clarity timeouts calculation and detection in lines 5 and 9 are not shown. Algorithm contains single procedure, repeated for each malware sample. Each sample is analyzed as long as it tries to connect to new servers. During preliminary analysis of malware it was detected that CryptoWall communicates with its servers in semi-random order. It starts to repeat that order after three attempts to connect to each server. That lead to condition used to detect completeness of the sample analysis (line 5). To mitigate potential transient communication problems, configurable timeouts where also applied in that condition, forcing the timed out experiment to be repeated for the given sample. To ensure high quality of gathered data, each active server was detected using separate execution of the sample in the controlled environment (line 6). After executing sample, experiment controller was monitoring network communication for occurrence of incoming malware server response or until some timeout passed (line 9). If response from server was received, it was added to the experiment results, and communication with IP address of that server was blocked for future experiments with the same sample, forcing that sample to try to stimulate another server (lines 10-13). Nevertheless reason for stopping sample run (line 9), all malware requests sent by this sample were gathered (lines 14-18) and used for experiment completeness condition (line 9). For next sample run, all network traffic blocks were lifted (line 4). Procedures used for detection of malware requests and responses were prepared during manual analysis.

Algorithm 1 Acquiring list of active malware servers.

```
1:  for all Sample ∈ MalwareSamples do
2:    Sample.Servers ← ∅
3:    Sample.ActiveServers ← ∅
4:    Unblock all network traffic
5:    while ∃ S ∈ Sample.Servers : S.Retries < 3 do
6:      Launch Sample in controlled environment
7:      repeat
8:        Monitor In-coming and Out-coming traffic
9:      until ∃ P ∈ In : IsMalwareResponse(P)
10:     if ∃ P ∈ In : IsMalwareResponse(P) then
11:       Block network traffic to and from P.SourceIP
12:       Add P.SourceIP to Sample.ActiveServers
13:     end if
14:     for all R ∈ Out do
15:       if IsMalwareRequest(R) then
16:         T ← Sample.Servers[R.TargetIp]
17:         Increment T.Retries
18:       end if
19:     end for
20:   end while
21: end for
```

VI. Cryptowall infrastructure analysis

As was described in the section III, communication to hinder detection infected machine connects to the command and control server via so called proxy servers. These servers are

hacked by the attacker and special proxy script was installed into web server. To raise the chance that an infected machine successfully download public key, each sample of CryptoWall malware contains hard-coded list of multiple proxy servers. The sample tries to connect in a sequence at the beginning of the infection. Our initial analysis reveals that various samples have the same list of proxy servers. Due to this fact we decided to cluster analyzed samples using proxy list as unique group identifier. Our analysis concerns almost 360 samples taken from openly available sources:

- blog `malware-traffic-analysis.net`
- `malwr.com`
- `reverse.it` services

To manage the collected samples during our research, they can be additionally tagged (beside a number) to be easily identified (*cw3-Feb* - for a sample of CryptoWall 3.0 obtained in February, *cw3-Mar*, etc.).

The samples use 59 unique proxy lists; average proxy list contains almost 40 unique URL, however maximal observed number of URL was 70. During our research we detect more than 2000 unique URLs, hosted in 1945 domains. Detailed analysis concerning domains and addresses is presented later in this chapter. Initial analysis of gathered data was performed with the help of graph theory. Data is used for generation of graphs, which represent connections between proxy lists and used domains. In the constructed graph two types of vertexes are introduced - blue and green. The blue vertexes represent name of proxy list. The green vertexes represent domains. Connection between vertexes indicates that this particular proxy list contains URL which is provided by server in this domain. Analysis of constructed graphs can reveal interesting patterns rapidly and help the person, who is performing analysis of gathered data. Analysis of samples from the beginning of the 2015 shows, that each new proxy list uses completely independent set of domains. Plotted graphs from this period are rather simple, especially in comparison to more complex plots from the end of the year. Fig. 4 presents sample graph of this type, in our security team called "flower".

However, from the middle of the 2015 year we started to observe more connections between various proxy lists. Sample graph of this type is presented in the Fig. 5.

As can be seen in the presented image there are many domains which are associated with two or even with three distinct proxy lists. What is interesting, some domains are used both for samples of CryptoWall version 3.0 and 4.0. This is an evidence that this malware is operated by one group of attackers. Moreover, in our opinion this reuse of domains can be sign that attackers have some problems with hacking or buying new machines, which hosts proxy script for various complains. Additionally, due to vast amount of data, rapid finding of interesting domains which should be investigated in the first row is very important. For this purpose graphs constructed in this fashion, can be beneficial, too. The most interesting domains are those, which connects two or more proxy lists. Shouting down such proxies we can eliminate the broadest spectrum of malware. These domains can be easily

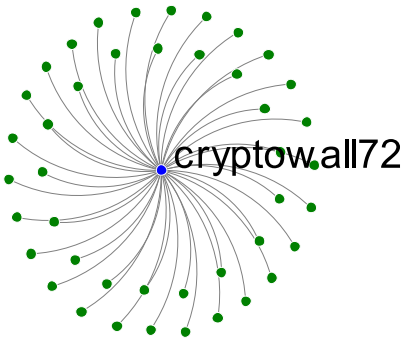


Fig. 4. Simple graph of member domains (green vertexes) associated with given proxy list (blue vertex)

automatically found, they are a green vertexes which degree value is greater than one. Additionally to the static analysis of all domain contained in each malware sample, we performed analysis which reveals information how many proxy servers in given instant provides access to the command and control servers. Accordingly to generally published information infected servers are easily detected and rapidly shut down by administrators. Our research confirms the first part of this statement. Unfortunately, we cannot agree with its second part. The most long lived proxy server observed during our research, allows access for victims to the C&C server for 11 weeks and 1 day. What is alarming, such servers are common. Fig. 6 presents how many servers in given proxy servers list associated with for CryptoWall 3.0 still allows access to the attackers C&C.

We executed samples in controlled environment, provided by the Maltester and MESS dynamic analysis systems. After successful reception of the public key form the C&C server we marked this server as alive, block its IP address in firewall and execute another analysis. Due to manual method we could perform no more than one check of given proxy list in a week. Because achieved in such way data was very valuable we decided to develop and deploy automatic system which could perform analysis more frequent. Details of the system were presented in the section IV. In the plot two instants are very interesting: the one in the middle of the September and the second just before the end of the December (both are marked in the figure with red arrows). In these two instants almost at the same time all proxy servers stopped forwarding the traffic to the command and control server. Due to the fact that these servers are placed in various countries such simultaneous actions with high probability was performed by the attackers. The first situation confirms our assumptions, because almost immediately many new samples of CryptoWall 3.0 come out with completely new proxy servers list. The second event marks the end of the CryptoWall 3.0 activity. After this we

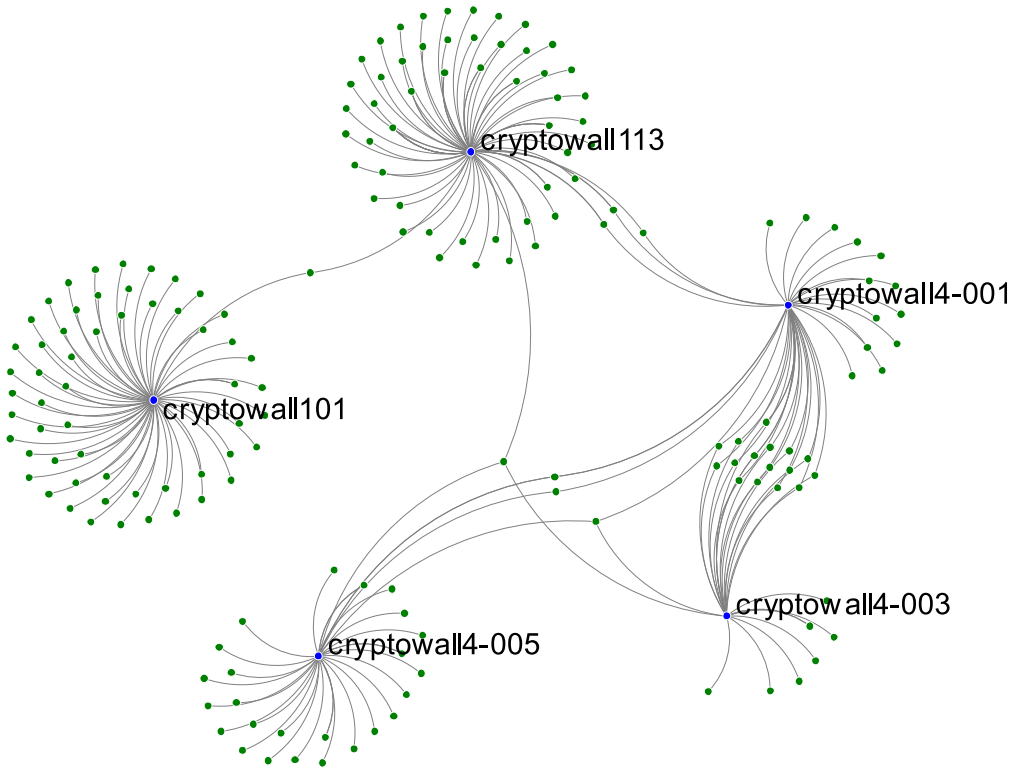


Fig. 5. Complex graph of member domains of proxy lists and its associations, observed at the end of 2015

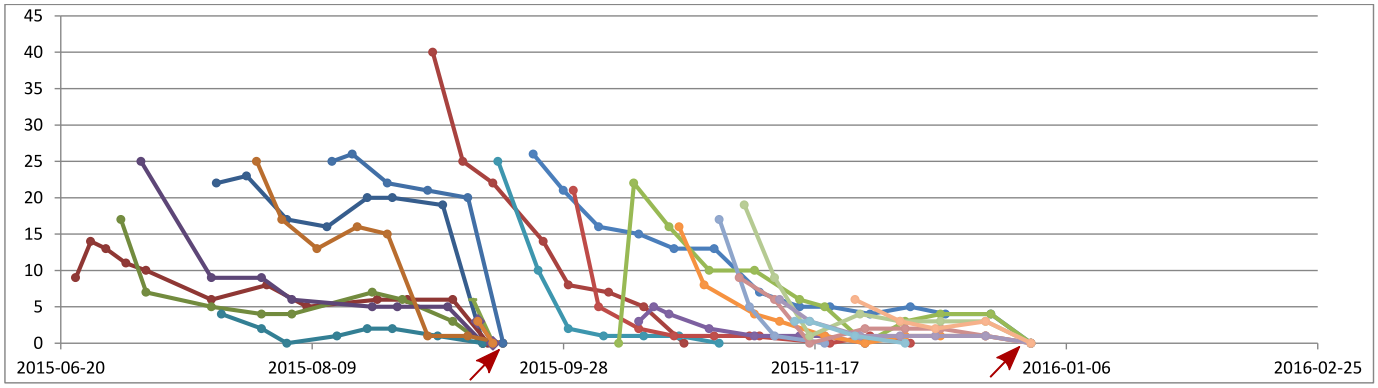


Fig. 6. Number of active CryptoWall proxy servers during 2015

have never observed responding proxy servers for CryptoWall 3.0 samples. However, the new threat came out - CryptoWall 4.0.

In addition we investigated the domain IP addresses and countries of origin of the infected proxy servers. The second aspect is very important, because it specifies to which national CERT or law enforcement agency (LEA) report concerning detected hostile activity should be provided. At the beginning, this aspect of analysis seems to be very simple. We have at least two source of such information: top level domain or geolocalization information associated with IP address. However, our research shows that this information can be inconsistent. We observed numerous examples, when country associated with DNS top level domain was other than country provided by the geolocalization database. To eliminate errors in gelocalization database, we investigated real localization of a server using `traceroute` utility program. In all such situations, information provided by the geolocalization database was accurate - the last few routers observed in the output have country top level domain associated returned country.

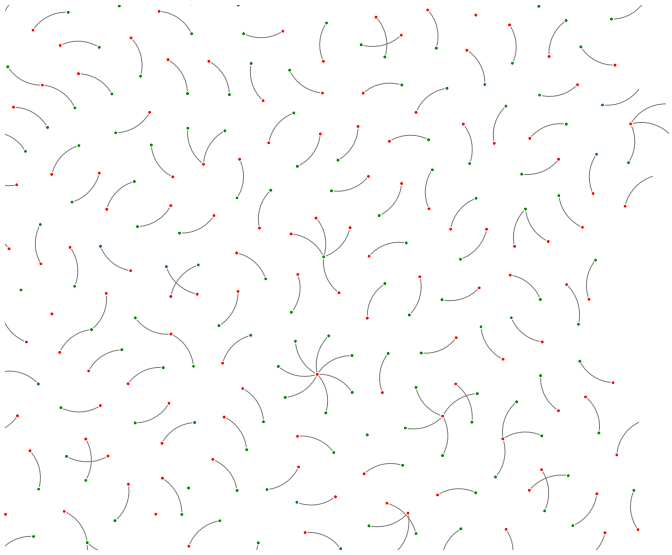


Fig. 7. Examples of graphs presenting associations between domains (green vertexes) and used IP address (red vertexes)

The good thing from this situation is that we could provide information concerning infection, both: to country of top level domain as well as to country where physically server is located. The vast amount of detected domains and IP addresses cannot be investigated all in short time. Due to this fact we introduces method which shows the most interesting data, which should be investigated in the first place. For this purpose we construct custom graph which have two types of vertexes - red and green. The green vertex represents detected domain. The red vertex represents associated with detected domain IP address. Connection between green and red vertexes determines that this domain is resolved to this particular IP address. Fig. 7 presents sample visualization of graphs, constructed in described manner.

Presented plot at first look is completely illegible. The most of presented graphs have only two vertexes, and these simple irrelevant ones hinder interesting knowledge. The first step in analysis of such data is removal of all irrelevant graphs. In data recorded during analysis of CryptoWall there are 1286 such graphs. Remaining ones consist of more than two vertexes. What is interesting, in remaining graphs only three have all vertexes which degree is greater than one. They are presented in the Fig. 8 (left). In remaining graphs there is always one vertex with degree greater than one and all other vertexes have degree equal to one. These graphs can be divided into two categories, depending on the type of vertex, which have degree greater than one. Two types of such graphs are presented in the Fig. 8 (right).

The first type of graph, with green vertex with degree greater then one, represent domains that have multiple IP addresses. The second one in contrast have multiple domains which are hosted in one IP address. The latter one is very promising from security perspective. In such situation, disabling this one address, can stop all domains hosted on it. Our research

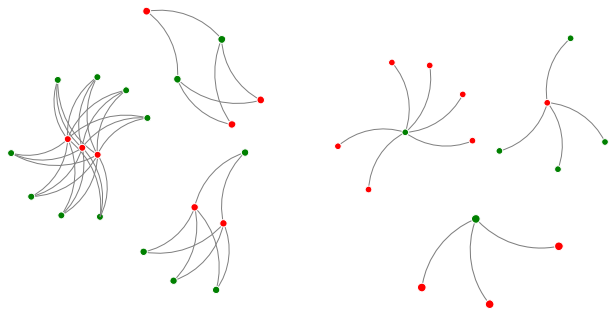


Fig. 8. Some of the most interesting graphs: with all vertexes of degree greater than one (left). Graphs with only one vertex of degree grater than one (right)

shows that attackers from the middle of the 2015 start reusing this same domains. In effect shutting down such domain can protect not only this one analyzed complain, but some before unknown, too. The first type of graphs can be useful, too. Because this same domain is hosted on various IP addresses, they are managed by on organization. In effect one contact with responsible person, can deactivate all of them. Results of our research lead to methodology which can be used for automatic prioritization of received date. In the first step graphs concerning domains and used IP addresses are constructed. In the second step degrees of all vertexes are calculated. These, which all vertexes have degree one, are removed from data presented to the person performing analysis. In effect, the most interesting from the security perspective domains or IP addresses can be easily detected and presented to the security officer in the first row, which can speed up whole process of finding and disabling hostile machines in the Internet.

VII. Conclusion

Starting on March 2015 the CryptoWall ransomware become a point of the authors interest. In order to evaluate its behaviour a rich set of experimental runs were conducted. To make these analysis safe and effective a special toolset and methods are needed.

Based on previous experience with malware analysis, the new, dynamic on-line analysis system, called MESS, was proposed. It proved to be an effective solution toward automated analysis, in particular, in data collection upon the malware behaviour within the operating system. The Hyper-V virtualization is quite effective approach in our case. There is a need to operate on different platforms in order to avoid hypervisor detection. In this sense, the MESS is complimentary to other similar systems.

Analysing the nowadays malware it has to done in two domains: actions within the target system and actions (i.e. communication) over the Internet. In the second domain, it is important to discover and investigate the infrastructure associated with the malware. Sometimes, to restrain the spread and side effects of malware, the easiest way is to identify and limit connectivity to its proxies or Command&Control servers.

In the paper we present the methodology of identification of CryptoWall proxies as well as propose new graph-based method for more effective analysis. The sad true is that the proxy servers, even when identified, are very hard to be turned off or cleaned. In several cases the authors successfully contacted proxy administrators (7 in Poland). However, the obtained life-time of the set of CryptoWall proxies is not optimistic. Further work will concentrate on the analysis of new malware families, like Locky or TeslaCrypt.

ACKNOWLEDGMENT

We would like to thank Wojciech Mazurczyk PhD., Dsc. for his assistance with finding and labeling samples of CryptoWall ransomware provided in the malwr.com service.

REFERENCES

[1] McAffe Labs, *Threats Report,* May 2015, www.mcafee.com/us/resources/reports/rp-quarterly-threat-q1-2015.pdf

[2] Symantec, *Internet Threat Report,* April 2015, www4.symantec.com/mktginfo/whitepaper/ISTR/21347932_GA-internet-security-threat-report-volume-20-2015-social_v2.pdf

[3] A. Kharraz, W. Robertson, D. Balzarotti, L. Bilge and E. Kirda, "Cutting the gordian knot: A look under the hood of ransomware attacks," *DIMVA 2015, 12th Conference on Detection of Intrusions and Malware & Vulnerability Assessment,* July 9-10, 2015, Milan, Italy, http://dx.doi.org/10.1007/978-3-319-20550-2_1

[4] K. Cabaj, P. Gawkowski, K. Grochowski, and D. Osojca, "Network activity analysis of CryptoWall ransomware", *Przegląd Elektrotechniczny,* Vol 91, No 11, 2015, http://dx.doi.org/10.15199/48.2015.11.48

[5] E. Skoundis and L. Zeltser, *Malware. Fighting Malicious Code,* Pearson Education Inc. ; 2004.

[6] U. Bayer, A. Moser, Ch. Kruegel and E. Kirda, "Dynamic analysis of malicious code," *J. in Comp. Virology*, vol. 2, 2006, pp 67-77., http://dx.doi.org/10.1007/s11416-006-0012-2

[7] X. Chen, J. Andersen, Z.M. Mao, M. Bailey and J. Nazario, "Towards an understanding of anti-virtualization and anti-debugging behavior in modern malware," in IEEE Int'l Conf. on Dependable Systems and Networks, 2008, pp. 177-186., http://dx.doi.org/10.1109/DSN.2008.4630086

[8] P. Ferrie, *The "Ultimate" Anti-Debugging Reference*, 2011 http://anti-reversing.com/Downloads/Anti-Reversing/The_Ultimate_Anti-Reversing_Reference.pdf

[9] K. Cabaj, Management System for Dynamic Analysis of Malicious Software, Information Systems In Management, 2015

[10] Cuckoo Sandbox website, https://www.cuckoosandbox.org, May, 2016

[11] Process Monitor website, https://technet.microsoft.com/pl-pl/sysinternals/bb896645.aspx, May, 2016

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 991–998
DOI: 10.15439/2016F94
ACSIS, Vol. 8. ISSN 2300-5963

Pseudo-random Sequence Generation from Elliptic Curves over a Finite Field of Characteristic 2

Omar Reyad
Warsaw University of Technology
Warsaw, Poland
Email: ormak4@yahoo.com

Zbigniew Kotulski
Warsaw University of Technology
Warsaw, Poland
Email: zkotulsk@tele.pw.edu.pl

***Abstract*—In this paper, the randomness of binary sequences generated from elliptic curves over a finite field of characteristic 2 is studied. A scheme of construction based on the Chaos-Driven Elliptic Curve Pseudo-random Number Generator (C-D ECPRNG) is proposed. The generators based of this scheme are verified by using tests from the NIST Statistical Test Suite to analyze their statistical properties. An elliptic curve used in the numerical example is defined over $\mathbb{F}_{2^8}$. The investigations which made for the generated series of two output sequences of the lengths of 2^{10} and 2^{20} bits shown that 14 generators working according to our general scheme exhibit good randomness properties. Next, the binary sequences generated by these 14 schemes were used for encrypting a 256×256 grayscale Lena image as an application example and the security analysis of the ciphered images was carried out.**

I. Introduction

IN 1985, Neil Koblitz [1] and Victor Miller [2] independently proposed one of the most important public-key cryptosystems named the elliptic curve cryptosystem, whose security rests on the discrete logarithm problem over points on an elliptic curve (EC) [3]. Elliptic curve public-key cryptosystems over finite fields ($\mathbb{F}_{2^m}$ or $\mathbb{F}_p$) have become widely used in applications such as smart cards which provide limited space for implementation of modular computations. Recently, the operations (Add, Double, Multiply) of points on elliptic curves over $\mathbb{F}_{2^m}$ or $\mathbb{F}_p$ have a well-developed technology in both hardware and software implementations.

Elliptic curves applications in, both, cryptography and communications are currently the subject of extensive investigation, as means for increasing security in transmission and reception of data over an insecure communication channel. The advantage is that elliptic curves over finite fields ($\mathbb{F}_{2^m}$ and $\mathbb{F}_p$) provide an inexhaustible supply of finite abelian groups. It is found that different elliptic curves defined over the same field have a different structure as finite fields of the same order are isomorphic to each other. With the increase in available computation power, it is found for a given key size that an EC public-key cryptosystem has higher security compared to RSA cryptosystem [4]. EC operations which used in the generation of pseudo-random sequences with strong cryptographic properties have been studied in the literature, such as [5], [6], [7].

In this paper, new constructions for the generation of pseudo-random sequences based on the properties of random numbers and elliptic curves over a finite field of characteristic 2 ($\mathbb{F}_{2^m}$) are proposed. These constructions are based on the C-D ECPRNG which takes benefits from a chaotic generator to reinforce the quality of an Elliptic Curve Pseudo-random Number Generator (ECPRNG). The addition of chaos will define a family of ECPRNGs that are chaotic while being fast, statistically perfect and cryptographically secure as discussed in [8], [9]. The randomness properties of the new constructions are also tested and found to pass tests in the NIST randomness test suite [26]. Such sequences can be used for generating random numbers in the EC digital signature algorithm and a session key in their encryption phases.

The paper is organized as follows. In Section II, the preliminaries of EC are discussed. An overview of various EC based pseudo-random sequence generators are given in Section III. In Section IV, we present several construction methods of binary sequences obtained from the C-D ECPRNG. An illustrative example is presented in Section V. In Section VI, randomness properties of the proposed sequences are discussed. A simple application of the proposed sequences for image encryption is executed in Section VII while conclusions are given in Section VIII.

II. Preliminaries

The definition of elliptic curves over a finite field of characteristic 2 and their arithmetic are given here to provide the general background for our exposition.

A. Elliptic Curve over a Binary Finite Field

The field $\mathbb{F}_{2^m}$ called a *characteristic-two* finite field or a *binary* finite field, can be viewed as a vector space of dimension m over the field $\mathbb{F}_2$ which consists of the two elements $\{0,1\}$. A non-supersingular elliptic curve E over the binary field $\mathbb{F}_{2^m}$ is defined by an equation of the form

$$y^2 + xy = x^3 + ax^2 + b \tag{1}$$

where the parameters $a, b \in \mathbb{F}_{2^m}$ with $b \neq 0$. The set $E(\mathbb{F}_{2^m})$ consists of all points $(x,y), x \in \mathbb{F}_{2^m}, y \in \mathbb{F}_{2^m}$, which satisfy the defining equation (1), together with a special point O called the point at infinity. These set of points form an abelian group with respect to the addition rules given in the following section.

978-83-60810-90-3/$25.00

B. Arithmetic of Elliptic Curve Group over $E(\mathbb{F}_{2^m})$

As mentioned in the previous section, when $b \neq 0$, the set of all points on the elliptic curve E along with a point at infinity constitute an abelian group under addition operation with O serving as its identity element [10]. It is to be noted here that this addition operation $(+)$ is not the "conventional addition" operation as it is based on the arithmetic of elliptic curves [11].

The algebraic formula for the sum of two points and the double of a point are the following:

1) $P+O=O+P \quad$ for all $P \in E(\mathbb{F}_{2^m})$.
2) If $P=(x,y) \in E(\mathbb{F}_{2^m})$, then $(x,y)+(x,x+y)=O$, note that the point $(x,x+y)$ is denoted by $-P$, and it is called the negative of P; observe that $-P$ is indeed a point on the curve E.
3) Point addition: Let $P=(x_1,y_1) \in E(\mathbb{F}_{2^m})$ and $Q=(x_2,y_2) \in E(\mathbb{F}_{2^m})$, where $P \neq \pm Q$. Then $P+Q=(x_3,y_3)$ where
$$x_3 = \left(\frac{y_1+y_2}{x_1+x_2}\right)^2 + \left(\frac{y_1+y_2}{x_1+x_2}\right) + x_1 + x_2 + a \tag{2}$$
and
$$y_3 \quad = \left(\frac{y_1+y_2}{x_1+x_2}\right)(x_1+x_3) + x_3 + y_1. \tag{3}$$
4) Point doubling: Let $P=(x_1,y_1) \in E(\mathbb{F}_{2^m})$, where $P \neq -P$. Then $2P=(x_3,y_3)$ where
$$x_3 = x_1^2 + \left(\frac{b}{x_1^2}\right) \ . \tag{4}$$
and
$$y_3 = x_1^2 + \left(x_1 + \frac{y_1}{x_1}\right)x_3 + x_3 \ . \tag{5}$$

III. Literature Review

A pseudo-random number generator (PRNG) is a deterministic algorithm which takes a random binary sequence of length k and outputs a binary sequence of length $n \gg k$ which "appears" to be random [12]. The input to the PRNG is called the seed, while the output is called a pseudo-random sequence. Different EC-based PRNG schemes suggested in literature use different ways to proceed from seed value for ith iteration to that for $(i+1)$th iteration and different predicates the output sequences, while the used one-way function is the EC point addition operation. Various suggestions for PRNG which based on ECs and their brief analysis are presented below.

A. The EC Power Generator

The Power Generator on EC (EC-PG) is introduced in [13], [14]. The definition of EC-PG for a given point $G(x,y) \in E(\mathbb{F}_p)$ of high order ℓ and an initial secret key $e \geq 2$ provided that the greatest common divisor (gcd) of $(gcd(e,\ell)=1)$ is generated by:

$$U_i = [e]U_{i-1} = [e^i]G \quad ,i=1,2,\ldots, \tag{6}$$

where $U_0(x,y) \in E(\mathbb{F}_p)$ is the "initial value". The output point sequence is the truncated x-coordinate of the resulted points $U_i(x,y)$.

B. The EC Linear Congruential Generator

The Linear Congruential Generator on EC (EC-LCG) has been suggested in [15] and then studied in a number of papers such as [16], [17], [18]. For a given point $G(x,y) \in E(\mathbb{F}_p)$ of high order ℓ, the EC-LCG is defined as the following sequence:

$$U_i = G + U_{i-1} = [i]G + U_0 \quad ,i=1,2,\ldots, \tag{7}$$

where $U_0(x,y) \in E(\mathbb{F}_p)$ is the "initial value". The output point sequence is generated as the resulted points $U_i(x,y)$ and passes through the complete cyclic subgroup of the point $G(x,y)$.

C. The Pseudo-random Bit Sequence Generator-B

The Pseudo-random Bit Sequence Generator (PBSG-B) which presented in [19] is a modification of the EC-LCG such that the periodicity is independent of the order of point $G(x,y)$ and the output sequence does not have any symmetric properties which makes the cryptanalysis easier. For security, the authors of [19] assume that both point $G(x,y)$ and the seed value of the Linear Feedback Shift Register (LFSR) are kept secret.

D. The Chaos-Driven Elliptic Curve Pseudo-random Number Generator

The Chaos-Driven Elliptic Curve Pseudo-random Number Generator (C-D ECPRNG) which presented in [20] for the finite field $\mathbb{F}_p$ is considered to be the EC-LCG driven by a chaotic map. Such a modification improves randomness of the sequence generated and increases it's periodicity. The C-D ECPRNG for a given seed point $G(x,y) \in E(\mathbb{F}_{2^m})$ as the secret key, is defined as the following sequences generated by additive EC-points operation:

$$\begin{aligned} U_i &= [i(1+b_i)]G + U_0 \\ &= \begin{cases} [i]G+U_0 & if \quad b_i=0 \\ [2i]G+U_0 & if \quad b_i=1 \end{cases} \quad ,i=1,2,\ldots \end{aligned} \tag{8}$$

where $U_0(x,y) \in E(\mathbb{F}_{2^m})$ is the "initial value" and b_i is the random bits generated by a chaotic map Φ

$$b_i = \begin{cases} 0 & if \quad \Phi^i(s) \in S_0 \\ 1 & if \quad \Phi^i(s) \in S_1 \end{cases} \quad ,i=1,2,\ldots \tag{9}$$

where the state space $S=[0,1]$ is the interval and $S_0=[0,0.5]$, $S_1=(0.5,1]$ are two subsets of the interval equal to 0.5. (For more details see [21]).

E. The EC Based Random Number Generator

The random number generator proposed in [22] has reduced latency and increased periodicity with a single point multiplication operation in each iteration. The output point sequence is $U_i=[k_i]G$ and $k_i=(i-1)+x_{i-1}$ where x_{i-1} is the x-coordinate of the point $U_{i-1}(x,y)$. The random number generator has good statistical properties and high periodicity.

F. The Dual-EC Generator

The Dual-EC generator has appeared in NIST recommendations [23]. It makes use of two points $G(x,y)$ and $Q(x,y)$ on a non-super singular elliptic curve $E(\mathbb{F}_p)$ for generation of random numbers. One point for generating the iterating key k as $k_i = x([k_{i-1}]G)$ and the other point for generating the output bit sequence as $t_i = x([k_i]Q)$ where t is the truncation function. The Dual-EC generator mechanism represents an EC scalar multiplication operation, followed by the extraction of the $x-$coordinate for the resulting points followed by truncation to produce the output sequence. We mention here that this recommendation is now withdrawn.

G. The Pseudo-random Bit Sequence Generator-A

A modification of the Dual-EC generator with increased periodicity named Pseudo-random Bit Sequence Generator-A (PBSG-A) is published in [19]. In PBSG-A, the iteration key k is modified as $k_{i+1} = [k_i]G + [i]C$ where $C = x([e]G)$ and "e" is the seed value. In addition to two point multiplication operations the modified algorithm requires a finite field multiplication of iteration number "i" and the value "C" to be carried out in each iteration. This increases both the hardware complexity and the time complexity of the system.

IV. Proposed constructions for EC Binary Sequences

In this section, we propose 22 different schemes resulted from different construction methods based on the C-D ECPRNG discussed in Section III-D.

The points resulted $U_i(x,y)$ with $x-$ and $y-$coordinates of each point are used to obtain the binary sequences. We will shortcut the word sequence to (Seq) throughout this paper and mention that an Initialization Vector (IV) is a fixed initialization vector that should be specified with the scheme. The exclusive or (XOR) logical operation with the symbol $\oplus$ is used here and one can show that it can be replaced by any operation that is an *easy-to-invert* permutation of one of its inputs when the second input is fixed.

After applying the C-D ECPRNG, we get the resulted points $U_i(x,y)$ and the $x-$ and $y-$coordinates of these points are used according to the construction methods listed in table I. These construction methods result in 22 different schemes by applying the ith iteration function R_i. For example, R_i for the first scheme is given by:

$$R_i = [R_{i-1} \oplus X_i], \quad , i \geq 1 \tag{10}$$

where $R_0 = IV$ and X is the $x-$coordinate of the first point U_1. The output R_i is the pseudo-random bit sequence. We will use the notion of a permutation operation (appear in table I as *Perm*) of the results from XOR operation for mapping the bit elements then considering them into the output sequence R for the next iteration process in some schemes. In other schemes, the substitution-box operation ($S-box$) which considers the heart of some ciphers because they are highly nonlinear is also used. $S-box$ takes the results from XOR operation and transforms them into the corresponding output then considering them into the output sequence R. $S-box$ is a basic component of symmetric key algorithms which performs substitution. In our calculations we used the Advanced Encryption Standard (AES) block cipher $S-box$ which discussed in [24].

V. Implementation Example

For experimental results we consider the EC defined over $\mathbb{F}_{2^8}$ given by:

$$E : y^2 + xy = x^3 + \alpha x^2 + 1 \tag{11}$$

where the parameters $a = \alpha, b = 1 \in \mathbb{F}_{2^8}$ with $b \neq 0$ and the EC is based on the irreducible polynomial $x^8 + x^4 + x^3 + x^2 + 1$ over $\mathbb{F}_2$. The total number of EC points is found to be 288 including O (point at infinity) and the element α is a generator of $\mathbb{F}_{2^8}$. The EC point $G = (\alpha^{186}, \alpha^{225})$ is chosen as the base point, which has the order $\ell = 288$ and the initial point is $U_0 = (\alpha^{34}, \alpha^{99})$. Also, $\{G, [2]G, \ldots, [288]G\}$ generates all the elements of EC over $\mathbb{F}_{2^8}$, hence the given elliptic curve group is cyclic. In the case of C-D ECPRNG, we use the Logistic map [25] as our chaotic map to generate the random bits b_i defined in (9).

VI. Randomness Properties

The purpose of this section is to check experimentally the randomness properties of the sequences generated in Section IV. The whole sequences generated by Section IV should have good statistical properties, we also decided to check the statistical properties and test the randomness using six basic statistical tests from [26], [27]. These tests are:

1) **Frequency (Monobit) Test**, it verifies if the number of "1" bits in the sequence lies within specified limits.
2) **8-bit Poker test**, it verifies whether bytes of each possible value appear approximate the same number of times.
3) **Runs Test**, it checks whether the number of runs (the test is carried out for runs of zeros and runs of ones) of length 1, 2, 3, 4 and 5 as well as the number of runs which are longer than 5, each lies within specified limits.
4) **Discrete Fourier Transform (Spectral) Test**, it detects the periodic features in the tested sequence that would indicate a deviation from the assumption of randomness.
5) **Linear Complexity Test**, it determines whether or not the sequence is complex enough to be considered random. Random sequences are characterized by longer LFSRs. An LFSR that is too short implies non-randomness.
6) **Cumulative Sums (Cusums) Test**, it determines whether the cumulative sum of the partial sequences occurring in the tested sequence is too large or too small relative to the expected behavior of that cumulative sum for random sequences. The test has two modes, which are either forward through the sequence or backward through the sequence, named in the Tables *Cusums (forward)* and *Cusums (reverse)*, respectively.

All the generated sequences from $Seq-1$ to $Seq-22$ is tested using the six basic tests discussed above. The test

TABLE I
THE 22 PROPOSED SEQUENCE SCHEMES

No.	scheme expression	No.	scheme expression
1	$[R_{i-1} \oplus X_i]$	12	$[R_{i-1} \oplus Y_i]$
2	$Perm[R_{i-1} \oplus X_i]$	13	$Perm[R_{i-1} \oplus Y_i]$
3	$S-box[R_{i-1} \oplus X_i]$	14	$S-box[R_{i-1} \oplus Y_i]$
4	$X_i \oplus Perm[R_{i-1} \oplus X_i]$	15	$Y_i \oplus Perm[R_{i-1} \oplus Y_i]$
5	$Y_i \oplus Perm[R_{i-1} \oplus X_i]$	16	$X_i \oplus Perm[R_{i-1} \oplus Y_i]$
6	$R_i \oplus Perm[R_{i-1} \oplus X_i]$	17	$R_i \oplus Perm[R_{i-1} \oplus Y_i]$
7	$[R_i \oplus Y_i] \oplus Perm[R_{i-1} \oplus X_i]$	18	$[R_i \oplus X_i] \oplus Perm[R_{i-1} \oplus Y_i]$
8	$X_i \oplus S-box[R_{i-1} \oplus X_i]$	19	$Y_i \oplus S-box[R_{i-1} \oplus Y_i]$
9	$Y_i \oplus S-box[R_{i-1} \oplus X_i]$	20	$X_i \oplus S-box[R_{i-1} \oplus Y_i]$
10	$R_i \oplus S-box[R_{i-1} \oplus X_i]$	21	$R_i \oplus S-box[R_{i-1} \oplus Y_i]$
11	$[R_i \oplus Y_i] \oplus S-box[R_{i-1} \oplus X_i]$	22	$[R_i \oplus X_i] \oplus S-box[R_{i-1} \oplus Y_i]$

TABLE II
TEST RESULTS FOR SEQ-1 AND SEQ-5

	Seq-1		Seq-5	
Test name	2^{10}	2^{20}	2^{10}	2^{20}
Monobit	0.5737	0.6157	0.4917	0.2597
Poker	0.2122	0.2220	0.2872	0.2869
Runs	0.1204	0.8510	0.1735	0.8507
DFT	0.2561	0.6139	0.6264	0.8313
L. Comp.	0.9196	0.2846	0.9196	0.4670
Cusums (F)	0.3999	0.7256	0.8035	0.3911
Cusums (R)	0.8831	0.9280	0.3999	0.1933

TABLE III
TEST RESULTS FOR SEQ-12 AND SEQ-16

	Seq-12		Seq-16	
Test name	2^{10}	2^{20}	2^{10}	2^{20}
Monobit	0.1691	0.8177	0.4917	0.1351
Poker	0.2771	0.0548	0.0849	0.7276
Runs	0.1028	0.9765	0.1071	0.9068
DFT	0.4905	0.3124	0.5980	0.4599
L. Comp.	0.9196	0.6583	0.1246	0.2629
Cusums (F)	0.0488	0.5020	0.8579	0.2025
Cusums (R)	0.2219	0.3369	0.3011	0.1273

results are shown that 14 schemes of the proposed 22 schemes exhibits good randomness properties. The other 8 schemes are found to have non-random properties especially with long binary sequences (2^{20} bits) and fail to pass most of the six tests. We presented in Tables II and III the test results for four schemes as examples to discuss. In Table II are presented results for the sequence $Seq-1$ and $Seq-5$. As it is noted, the generator works correctly for short and long binary sequences (2^{10} and 2^{20}bits). In Table III, results for the sequence $Seq-12$ and $Seq-16$ are presented. Also it is clear that the C-D ECPRNG enables generating correctly short and long sequences and the generator passes all the presented tests. For the rest of the paper, we will consider only the 14 schemes (namely $Seq-1$, $Seq-2$, $Seq-3$, $Seq-5$, $Seq-8$, $Seq-9$, $Seq-11$,$Seq-12$, $Seq-13$, $Seq-14$, $Seq-16$, $Seq-19$, $Seq-20$, $Seq-22$) that had good randomness properties.

VII. IMAGE ENCRYPTION APPLICATION EXAMPLE

Image encryption is a potential application where stream cipher is highly preferred over block cipher due to the bulky nature of the data and high correlation between the adjacent pixels. The pseudo-random sequence used for image encryption must have good randomness properties and high periodicity so that the encrypted image is secure. Recently, several attempts for using ECs in image encryption has been proposed in literature such as [28],[29],[30]. In this section, the pseudo-random sequences generated by the considered 14 schemes are used for encrypting a 256×256 grayscale Lena image in which each pixel has a 8-bit value of between 0 and 255 and the security analysis of the ciphered images are carried out.

A. Entropy Analysis

Entropy is defined to express the degree of uncertainties in the system. It is well known that the entropy $H(m)$ of a message source m can be calculated as:

$$H(m) = -\sum_{i=0}^{255} P(m_i) log_2 P(m_i) \tag{12}$$

where $P(m_i)$ represents the probability of symbol m_i. For all the considered cipherimages shown in Figs. 3(a – n), the number of occurrence of each gray level is recorded and the probability of occurrence is computed. Table IV indicates the various values of the entropies for the plain and encrypted images by the considered 14 schemes. It can be noted that the entropy of the encrypted images are very near to the theoretical value of 8 indicating that all the pixels in the encrypted images occur with almost equal probability. Therefore, the information leakage in the considered cipher schemes is negligible, and it is secure against the entropy-based attack. Also it is comparable

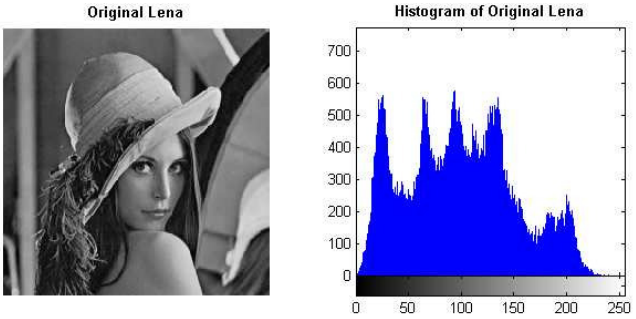


Fig. 1. Lena image and it's Histogram

to the entropy values presented by references [30], [31] and [32].

TABLE IV
ENTROPY AND CORRELATION COEFFICIENTS FOR LENA IMAGE

Scheme	Entropy	Horizontal	Vertical	Diagonal
Lena	7.5807	0.93915	0.96890	0.91686
Seq-1	7.9973	-0.00201	0.04720	0.00132
Seq-2	7.9971	-0.00431	0.00513	-0.00443
Seq-3	7.9975	0.00061	-0.00319	-0.00572
Seq-5	7.9970	0.00390	0.00879	-0.00030
Seq-8	7.9972	0.00591	-0.03651	0.00481
Seq-9	7.9977	-0.00007	-0.00345	0.00378
Seq-11	7.9972	0.00638	-0.01068	-0.00391
Seq-12	7.9973	-0.00071	0.03101	0.00501
Seq-13	7.9972	0.00220	-0.01799	-0.00583
Seq-14	7.9973	-0.00331	-0.00323	0.00588
Seq-16	7.9968	0.00690	0.01358	0.00325
Seq-19	7.9972	-0.00287	-0.04253	-0.00174
Seq-20	7.9973	-0.00076	-0.00670	0.00003
Seq-22	7.9967	0.00026	0.00259	-0.00645
Ref.[30]	7.9964	-0.00079	-0.0013	-0.0046
Ref.[31]	7.9885	0.0132	0.0017	0.0034
Ref.[32]	7.9968	0.0025	0.0037	0.0011

B. Correlation Analysis

It is known that two adjacent pixels in a plainimage are strongly correlated vertically, horizontally and diagonally. This is the property of any ordinary image. The maximum value of correlation coefficient is 1 and the minimum is 0. A robust encrypted image to statistical attack should have a correlation coefficient value of ˜0 as discussed in [33]. Results of horizontal, vertical and diagonal directions are obtained as shown in Table IV for Lena plainimage and the ciphered images by the considered 14 schemes respectively. These results demonstrate that there is negligible correlation between the two adjacent pixels in the encrypted images, even when the two adjacent pixels in the plainimage are highly correlated.

C. Sensitivity Analysis

In order to avoid the known-plaintext attack, the changes in the cipherimage should be significant even with a small change in the plainimage. If one small change in the plainimage can cause a significant change in the cipherimage, with respect to diffusion and confusion, then the differential attack actually loses its efficiency and becomes practically useless. To quantify this requirement, two common measures are used: Number of Pixels Change Rate (NPCR) and Unified Average Changing Intensity (UACI) [34]. We have tested the NPCR and UACI with the considered 14 sequence schemes to assess the influence of changing a single pixel in the plainimages on the encrypted images. From the results, we have found that the average values of the percentage of pixels changed in encrypted image is greater than 99.60% for NPCR and 30.50% for UACI for all the 14 generated sequences. This implies that the considered 14 schemes are very sensitive with respect to small changes in the plainimage.

D. Histogram Analysis

To prevent the leakage of information to an adversary, it is important to ensure that cipherimage does not have any statistical resemblance to the plainimage. A good image encryption scheme should always generate a cipherimage of the uniform histogram for any plainimage. In this work, the histograms are plotted for Lena plain and encrypted images. The histogram of Lena plainimage contains large spikes as shown in Fig. 1 while the histograms of it's cipherimages are almost flat and uniform which indicates equal probability of occurrence of each pixel as shown in Figs. 2(a – n). They are significantly different from the respective histogram of the Lena plainimage and hence does not provide any clue to employ any statistical attack on the considered 14 image encryption schemes.

VIII. CONCLUSION

In this paper, we have presented several construction methods based on a common general scheme for generating binary sequences from EC over a binary finite field ($\mathbb{F}_{2^m}$). The proposed scheme is based on the C-D ECPRNG with simple arithmetic transformations (XOR and permutation or $S-box$) to produce long size binary sequences with good randomness properties. The generated sequences are tested using tests from the NIST randomness test suite to analyze their statistical properties. It is found that 14 schemes of the 22 proposed specific schemes have passed the selected six complementary tests and the sequences generated by these 14 schemes work correctly

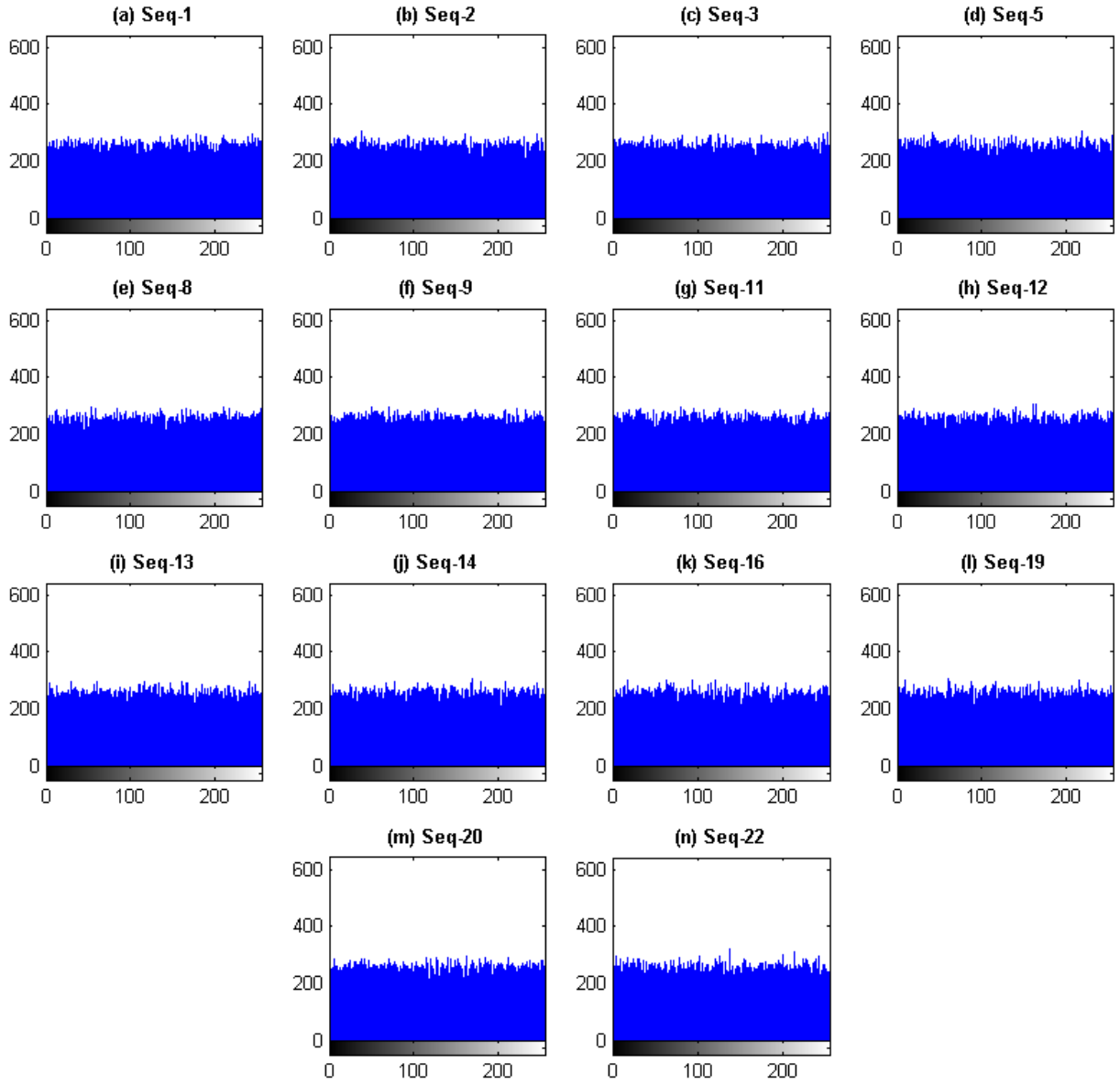


Fig. 2. Histogram of encrypted Lena image with the considered 14 sequences

with short (2^{10} bits) and long (2^{20} bits) size sequences. The pseudo-random sequences generated by these 14 schemes are applied to image encryption as an application example and the security analysis of the ciphered images are carried out. It is found also that the sequences generated using the C-D ECPRNG had high periodicity so that the encrypted images are secure. In addition, it has large key space, which is by far very safe for image encryption applications, and outperforms the competitive image encryption algorithms in terms of efficiency comparing to other encryption schemes.

ACKNOWLEDGMENT

This work has been supported financially by the Ministry of Higher Education of Egypt. Calculations in this paper were performed at the Interdisciplinary Center for Mathematical and Computational Modeling (ICM) of the University of Warsaw part of the grant calculation No. G63-2.

REFERENCES

[1] N. Koblitz, "Elliptic curve cryptosystems," *Mathematics of Computation,* vol. 48, 1987, pp. 203–209.

[2] V. Miller, "Uses of elliptic curves in cryptography," *Advances in Cryptology-CRYPTO'85,* vol. 218, Springer, Heidelberg, 1986, pp. 417–426, doi:10.1007/3-540-39799-X_31

[3] A. Menezes, *Elliptic Curve Public Key Cryptosystems,* Kluwer Academic, Dordrecht 1993, doi:10.1007/978-1-4615-3198-2

[4] N. Gura, A. Patel, A. Wander, H. Eberle and S.C. Shantz, "Comparing elliptic curve cryptography and RSA on 8-bit CPUs," *Cryptographic Hardware and Embedded Systems (CHES),* vol. 3156, Springer, Heidelberg, 2004, doi:10.1007/978-3-540-28632-5_9

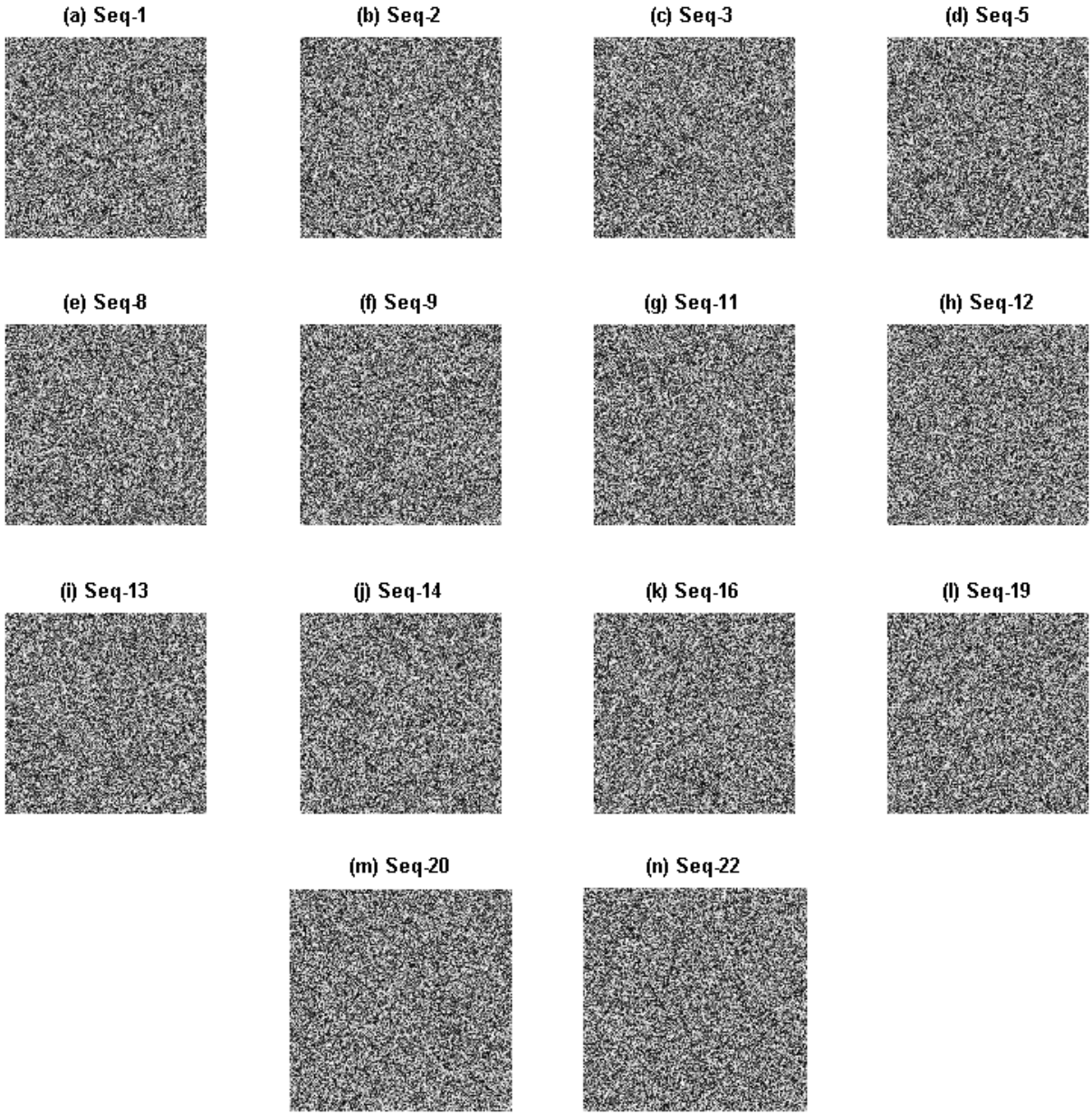


Fig. 3. Encrypted Lena image with the considered 14 sequences

[5] B. S. Kaliski, "One-way permutations on elliptic curves," *Journal of Cryptology 3,* 1991, pp. 187–199, doi:10.1007/BF00196911

[6] Z. Chen, S. Li and G. Xiao, "Construction of pseudo-random binary sequences from elliptic curves by using discrete logarithm," *In: G. Gong, et al. (eds.): SETA 2006. LNCS,* vol. 4086, Springer, Heidelberg, 2006, pp. 285–294, doi:10.1007/11863854_24

[7] S. V. Sathyanarayana, M. A. Kumar and K. N. H. Bhat, "Random binary and non-binary sequences derived from random sequence of points on cyclic elliptic curve over finite field $GF(2^m)$ and their properties," *Information Security J.: A Global Perspective,* vol. 19, 2010, pp. 84–94, doi:10.1080/19393550903482759

[8] J. M. Bahi and C. Guyeux, *Discrete Dynamical Systems and Chaotic Machines: Theory and Applications,* CRC Press, Numerical Analysis and Scientific Computing, London 2013.

[9] R. L. Tataru, "Image hashing secured with chaotic sequences," *Proceedings of the 2014 Federated Conference on Computer Science and Information Systems (FedCSIS),* IEEE, 2014, pp. 735–740, doi:10.15439/2014F250

[10] D. Johnson, A. Menezes and S. Vanstone, "The Elliptic Curve Digital Signature Algorithm," *International Journal of Information Security,* vol. 1, 2001, pp. 36–63, doi:10.1007/s102070100002

[11] J. H. Silverman, *The arithmetic of elliptic curves,* Springer-Verlag, New York 2009, doi:10.1007/978-0-387-09494-6

[12] D. Szalkowski and P. Stpiczynski, "Template Library for Multi-GPU Pseudorandom Number Recursion-based Generators," *Proceedings of the 2013 Federated Conference on Computer Science and Information Systems (FedCSIS),* IEEE, 2013, pp. 515–519.

[13] T. Lange and I. E. Shparlinski, "Certain exponential sums and random walks on elliptic curves," *Canad. J. Math.,* vol. 57, 2005, pp. 338–350, doi:10.4153/CJM-2005-015-8

[14] E. El Mahassni and I. E. Shparlinski, "On the distribution of the elliptic curve power generator," *Proc. 8th Conf. on Finite Fields and Appl.,*

Contemp. Math., vol. 461, Amer. Math. Soc., Providence, RI, 2008, pp. 111–119.
[15] S. Hallgren, "Linear congruential generators over elliptic curves," *Preprint CS94-143*, Dept. of Comp. Sci., Cornegie Mellon Univ., 1994.
[16] G. Gong, T.A. Berson and D.R. Stinson, "Elliptic curve pseudorandom sequence generators," *Selected areas in cryptography,* vol. 1758, Springer, Berlin, 2000, doi:10.1007/3-540-46513-8_3
[17] O. Reyad and Z. Kotulski, "On Pseudo-random Number Generators Using Elliptic Curves and Chaotic Systems," *J. Appl. Math. Inf. Sci.,* vol. 9, 2015, pp. 31-38, doi:10.12785/amis/090105
[18] P. Beelen and J. Doumen, "Pseudorandom sequences from elliptic curves," *Finite Fields with Applications to Coding Theory,* Cryptography and Related Areas, Springer, Berlin, 2002, doi:10.1007/978-3-642-59435-9_3
[19] P. P. Deepthi and P. S. Sathidevi, "New stream ciphers based on elliptic curve point multiplication," *Computer Communications,* vol. 32, 2009, pp. 25–33, doi:10.1016/j.comcom.2008.09.002
[20] O. Reyad and Z. Kotulski, "Statistical Analysis of the Chaos-Driven Elliptic Curve Pseudo-random Number Generators," *In: Z. Kotulski, et al. (eds.) CSS 2014. CCIS,* vol. 448, Springer, Heidelberg, 2014, pp. 38–48, doi:10.1007/978-3-662-44893-9_4
[21] J. Szczepanski and Z. Kotulski, "Pseudorandom number generators based on chaotic dynamical systems," Open Systems & Information Dynamics, vol. 8, 2001, pp. 137–146, doi:10.1023/A:1011950531970
[22] L. P. Lee and K. W. Wong, "A random number generator based elliptic curve operations," *Computers & Mathematics with Appl.,* vol. 47, 2004, pp. 217–226, doi:10.1016/S0898-1221(04)90018-1
[23] E. B. Barker and J. M. Kelsey, "Recommendation for Random Number Generation Using Deterministic Random Bit Generators (Revised)," *US Department of Commerce, Technology Administration, National Institute of Standards and Technology,* Computer Security Division, Information Technology Laboratory, 2007.
[24] J. A. Buchmann, *Introduction to Cryptography,* Undergraduate Texts in Mathematics, Springer-Verlag New York, 2004, doi:10.1007/978-1-4419-9003-7
[25] S. C. Phatak and S. S. Rao, "Logistic map: A possible random-number generator," *Physical Review E* vol. 51, 1995,pp. 3670–3678, doi:10.1103/PhysRevE.51.3670
[26] A. Rukhin, J. Soto, J. Nechvatal, et al., "A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications," *NIST Special Publication 800-22 with revisions,* May 2001.
[27] T. Rachwalik, J. Szmidt, R. Wicik and J. Zablocki, "Generation of Nonlinear Feedback Shift Registers with special-purpose hardware," IEEE Transl. Military Communications and Information Systems Conference (MCC), pp. 1–4, October 2012.
[28] S. Maria and K. Muneeswaran, "Key generation based on elliptic curve over finite prime field, " *Int. J. Elect. Sec. and Digital Forensics,* vol. 4, 2012, pp. 65–81, doi:10.1504/IJESDF.2012.045391
[29] O. Reyad and Z. Kotulski, "Image Encryption Using Koblitz's Encoding and New Mapping Method Based on Elliptic Curve Random Number Generator," *In: A. Dzich, et al. (eds.): MCSS 2015. CCIS,* vol. 566, Springer, Heidelberg, 2015, pp. 34–45, doi:10.1007/978-3-319-26404-2_3
[30] S. V. Sathyanarayana, M. Aswatha Kumar and K. N. Hari Bhat, "Symmetric key image encryption scheme with key sequences derived from random sequence of cyclic elliptic curve points," *Int. J. Netw. Secur.,* vol. 12, 2011, pp. 137–150.
[31] A. Soleymani, M. J. Nordin, Z. M. Ali and L. Golafshan, "A Binary Grouping Approach for Image Encryption Based on Elliptic Curves over Prime Group Field," IEEE Transl. 11th Malaysia International Conference on Communications (MICC), pp. 373–378, November 2013, doi:10.1109/MICC.2013.6805857
[32] J. Payingat and P. P. Deepthi, "Pseudorandom Bit Sequence Generator for Stream Cipher Based on Elliptic Curves," *Mathematical Problems in Engineering,* Hindawi Pub. Cor., vol. 2015, 2015, pp. 1–16, doi:10.1155/2015/257904
[33] G. Zhang and Q. Liu, "A novel image encryption method based on total shuffling scheme," *J. Optics Communications,* vol. 284, 2011, pp. 2775–2780, doi:10.1016/j.optcom.2011.02.039
[34] Y. Wu, J. P. Noonan and S. Agaian, "NPCR and UACI Randomness Tests for Image Encryption," IEEE Transl. J. of Selected Areas in Telecommunications (JSAT), pp. 31–38, April 2011.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 999–1008
DOI: 10.15439/2016F158
ACSIS, Vol. 8. ISSN 2300-5963

An initial insight into Information Security Risk Assessment practices

Gaute Wangen
NISLab and CCIS, NTNU Gjovik
Teknologiveien 22, 2802 Gjovik, Norway
Email: gaute.wangen2@ntnu.no

***Abstract*—Much of the debate surrounding risk management in information security (InfoSec) has been at the academic level, where the question of how practitioners view predominant issues is an essential element often left unexplored. Thus, this article represents an initial insight into how the InfoSec risk professionals see the InfoSec risk assessment (ISRA) field. We present the results of a 46-participant study where have gathered data regarding known issues in ISRA. The survey design was such that we collected both qualitative and quantitative data for analysis. One of the key contributions from the study is knowledge regarding how to handle risks at different organizational tiers, together with an insight into key roles and knowledge needed to conduct risk assessments. Also, we document several issues concerning the application of qualitative and quantitative methods, together with drawbacks and advantages. The findings of the analysis provides incentives to strengthen the research and scientific work for future research in InfoSec management.**

I. Introduction

THE PRIMARY goal of InfoSec is to secure the business against threats and ensure success in daily operations by ensuring confidentiality, integrity, availability, and non-repudiation [1]. Best practice InfoSec is highly dependent on well-functioning InfoSec risk management (ISRM) processes[2]. While ISRM is the practice of continuously identifying, reviewing, treating and monitoring risks to achieve acceptance[3].

This paper investigates the practitioners view of research problems within information security (InfoSec) risk assessment (ISRA). While there is plenty of available material regarding what ISRA frameworks contain and how they compare with each other [4], the literature is scarce regarding the current ISRA industry practices. There are several known theoretical problems in ISRA[4], [5], however, we do not know if the risk practitioners agree that these problems are either relevant or representative. Thus, there is the possibility that existing literature is incomplete and that academia is missing the important issues. This paper contains the results and analysis from a combined quantitative and qualitative study of the practitioners view, and represents a step towards a more holistic picture of industry ISRA practices.

Part one of this study [6] researched practices in InfoSec (ISRM) with emphasis on the risk management part and issues, while this study emphasizes the risk assessment and analysis parts. We provide new knowledge regarding where the research in ISRA should be focusing the efforts, making the ISRA community and researchers the primary beneficiaries of this study. Improving ISRA is essential in making progress in the InfoSec research field as it is this process that helps organizations determine what and how to protect. Thus, the intended audience of this paper is InfoSec professionals and academics, together with other ISRA practitioners and stakeholders.

The main research problem investigated in this article is "How do the ISRA problems outlined in previous work ([4]) reflect problems experienced in the industry?". The scope of this article covers the ISRA process, including risk identification, estimation, evaluation, and risk treatment practices [3], and is limited to the practitioner point-of-view. We separate between risk assessment (ISRA) and analysis (ISRAn), where the assessment is defined as the overall process of risk identification, estimation, and evaluation. While risk analysis is the practical hands-on parts of risk identification and estimation, for example, a practitioner may choose ISO/IEC 27005:2011 as the overall approach to ISRM/ISRA, while prioritizing *Fault tree analysis* for ISRAn.

The remainder of this article has the following structure: First, we briefly describe the related work, before presenting the research method in the form of data collection approach, demographics, and analysis. Following this is a combined analysis and discussion of the results, where we start with findings on the high-level risk assessment practices, before diving into the deeper aspects of InfoSec risk analysis (ISRAn) and risk treatment. Lastly, we summarize our findings, including limitations of this study, and conclude the paper.

A. Related work

This work primarily builds on previous work conducted on the topic of research problems in ISRM/ISRA. Both Wangen and Snekkenes [4] and Fenz et al. [5] have published articles on current challenges in ISRM; The former is a literature review that categorizes research problems into a taxonomy. The latter discusses current challenges in ISRM, pre-defines a set of research challenges, and compares how the existing ISRM methods support them. The primary purpose of the Fenz et al. study was to categorize and present known research problems at different stages in the ISRM/RA areas and activities. These two articles provide the primary literature foundation for this study. The data for this study was gathered in one comprehensive questionnaire, where the first part concerning ISRM was published in [6].

II. Research Method

This study was conducted to investigate ISRM industry practices and the respondents' views of several known challenges within the research field. 46 respondents participated in

Fig. 1. How respondents ranked themselves (x-axis) and how they were rated in the survey (Y-axis)

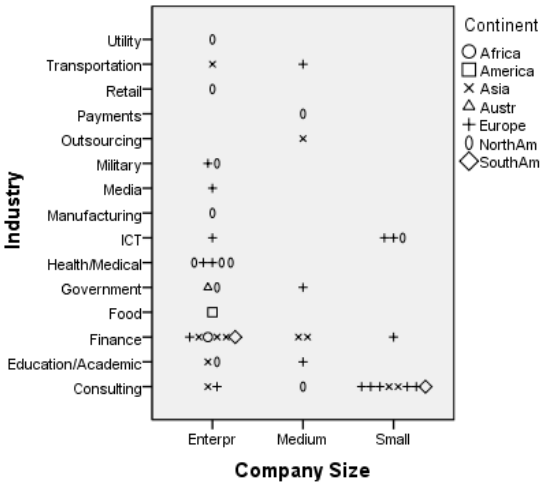


Fig. 2. Respondent demographics, based on company size (x-axis), industry (Y-axis) and Continent.

our online survey. The first sub-section addresses the choice of data collection method and measurement, followed by the demographics, and a brief overview of the statistical methods used for data analysis.

A. Data Collection, Sample, and Measurement

In their study, Kotulic and Clark [7] highlights that one of the most prominent problems in InfoSec studies is getting in touch with the target group and acquiring respondents. They propose several potential explanation for this: Where one is that InfoSec research is one of the most intrusive types of organizational studies. Also, that there is a general mistrust of any ”outsider” attempting to gain data about the actions of the security practitioner community [7]. Thus, we consider non-intrusiveness an essential requirement when designing the data collection tool. The narrow target group, industry professionals, made obtaining respondents a challenge as the study was subject to geographical limitations. To overcome said limitations we attempted to recruit participants from InfoSec risk specialized online forums. We considered this approach as non-intrusive, and it exposed the survey to many within the target group. However, it presents several problems; with this strategy the researcher has little control of participants except that they are members of particular forums, Table I. We, therefore, included self-rating questions in the questionnaire for the respondents to rate their knowledge, expertise and experience, together with our knowledge-based control questions. We designed a classification scheme based on this information, see Fig. 1.

We designed the questionnaire in Google Forms according to the procedure for developing better measures [8]. As for the level of measurement, the questionnaire had category, ordinal, and continuous type questions. Category type questions mainly for demographics and categorical analysis, while the main bulk of questions were designed using several mandatory scale- and ranking questions. The main categories applied for analysis is seen in Fig. 1, together with company size, and work type. The questionnaire also included several non-mandatory fields for commenting on previous questions or just for sharing knowledge about a subject. It had four pages of questions in total; the first page was demographics and self-rating questions. The questionnaire consisted of 37 questions in total, with an estimated completion time of 15-40 minutes depending on how much information the respondent shared. This paper consists of the results from questions regarding risk assessment and analysis.

TABLE I. GROUPS AND FORUMS WHERE THE QUESTIONNAIRE WAS POSTED

LinkedIN Forum name	**Members (at release time)**
IT Risk Management	3 443
CRISC (Official) *(Certified in Risk and Information Systems Control)*	1 400
Information Security Risk Assessment	441
ISO27000 for Information Security Management	22 620
Information Security Expert Center	8 906
Risk Management & Information Security *(Google+)*	521

B. Demographics

We received 46 accepted answers, See Table II for the classification of respondent expertise and work type (technical or administrative). While Fig. 2 displays respondent demographics categorized on company size, industry, and geographical affiliation. For the analysis, we applied the following definitions of company size: Small equals 1-249 employees, Medium 250 -1000, and Enterprise more than 1000.

TABLE II. CLASSIFICATION OF RESPONDENTS, TOTAL 46.

	Expert	Proficient	Competent
Administrative Work	13	10	6
Technical Work	7	7	3

C. Analysis

We applied a variety of statistical data analysis methods specified in the results, and the IBM SPSS software for the statistical analysis. A summary of the statistical tests used in this research is as follows:

For *Descriptive analysis* we have considered distributions including range and standard deviation. On continuous type

questions, we applied measures of central tendency mean, median and mode. We also conducted *Univariate* analysis of individual issues, and *Bivariate* analysis for pairs of questions, such as a category and a continuous question, to see how they compare and interact. However, we have restricted the use of mean and standard deviation for Likert-type questions and ordinal data where there was not defined a clear scale of measurement between the alternatives, as the collected data will seldom satisfy the requirements of normality. We have, therefore, analyzed the median together with an analysis of range, minimum and maximum values, and variance. This study also analyses the distributions of the answers, for example, if they are normal, uniform, binomial, or similar. *Crosstabulation* was applied to analyze the association between two category type questions, such as "Company Size" and "Expertise." We have used Pearson two-tailed *Correlation test* to reveal relationships between pairs of variables as this test does not assume normality in the sample.

The questionnaire also had several open-ended questions. We have treated these by listing and categorizing the responses. Further, we counted the occurrence of each theme and summarized the responses. Also, each continuous question had the possibility for the respondent to write a comment and offer further qualitative insight on an issue, where the most valuable comments are a part of this paper.

III. InfoSec Risk Assessment practices

This section contains the results and discussion of the statistical analysis regarding the ISRA practices. We start at a high-level; with the ISRA practices in organizational tiers, who should attend the ISRA, and what knowledge is important to have included in the process.

A. ISRA and Organizational Tiers

It is common to differentiate between risks at different tiers of abstraction when assessing an organization, such as Operational/Information Systems (low level), Tactical (mid-level), and Strategic (high level) information risks (for example [9]). The strategic and tactical type-risks can provide the risk analyst more time to estimate, risks in the operational environment often has to be handled ad-hoc or within a limited period. As these tiers are quite different and come with different types of risk, we asked if the practitioners distinguish between ISRA methods for them. 28% answered that they do, while the remainder answered no or other. There was no significant difference between groups in this question, Fig. 3. There were three detailed technical insights offered by the participants to shed light on practices, one technical (tech) expert responded: *"We apply the same methodology but are far less formal with tactical solutions. While a strategic solution would require formal sign off, tactical solutions need only require an email approval."*

While an administrative (admin) expert answered:*"High or Very High risks require detailed documented analysis (eg Bowtie diagrams) At each organisational level the risks are assessed against consequences at that level and mitigation applied at that level - if mitigation are insufficient at that level, the risk is escalated to the next higher level and re-assessed."*

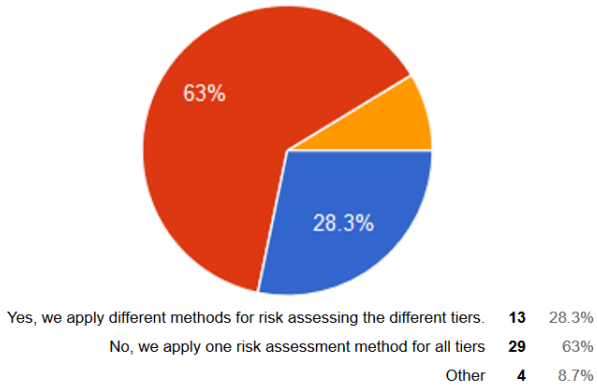


Fig. 3. ISRA practices on different organizational tiers

A tech proficient respondent answered: *"We use different methods for financial risk, IT (security) risk and business strategic risk. method for financial risk is "FOCUS" (successor of "FIRM"), as prescribed in regulations; method for IT security risk based on ISO 27005/31000, method for strategic risk is not formalised."*

The three answers show that there are several nuances to this problem that has not yet been highlighted in academia. The lower organizational tiers may be handled less informally, as it is likely these need faster decision-making. Our results show that some organizations have implemented different approaches to dealing with this problem, while others stick to one approach for all risk types. Awareness around this issue is also something that can be further researched in academia.

B. Who attends and conducts InfoSec risk assessments?

Having people with the right expertise and knowledge about the target system attending the risk assessment is one crucial success factor. Our results should provide a pointer on how to organize thee risk assessment and who should attend.

To get a generic overview of who attends and conducts ISRA in the practitioners organizations, we asked the participants who attends risk assessments in their organization. As two respondents pointed out, this picture depends on the type of risk assessment being conducted, yet, frequencies of attendance can still be estimated. Table III holds an overview of who attends ISRAs in the respondents organizations. The alternatives was "Never attends" (1), Sometimes attends (2), Always attends (3), Leads assessments (4), and we removed the respondents opting *Not present* for the statistical analysis, Table IV.

The results show that the CSO/CISO (Chief InfoSec Officer) most frequently leads risk assessments, while ICT security personnel most frequently attends. With the Head of ICT department and Operations personnel also attending with a high frequency. IT architects and software developers also attend the ISRA process frequently.

We found that in smaller companies, the CEO and CTO is much more likely to attend/lead risk assessments than in medium and enterprise sized companies, Table IV. Although, in some organizations, especially small ones, employees will have overlapping roles. One admin expert provided a caveat

about having high management involved: *"Having C[EO] or high management inside Information Security assessment will not allow the participants to be open when providing input for risk identification."*[1]

Comments on the results in table III, were from six admin experts and one admin proficient. Out of the seven written comments, five of them specified that the composition of the risk assessment team is dependent on the scope of the assessment; *"If business processes or systems are included in the scope, system owners or users with good knowledge of the processes attend."*

TABLE III. ROLES ATTENDING IN RISK ASSESSMENTS.

Attends/ Roles	Never present	Sometimes	Always	Leads	Not present in Organiza.
CEO	34.8%	28.3%	15.2%	13 %	8.7%
CSO/CISO	4.3%	15.2%	34.8%	32.6%	13 %
CTO	15.2%	17.4%	30.4%	8.7%	28.3%
CIO	19.6%	19.6%	28.3%	13 %	19.6%
Head of IT Dep	10.9%	26.1%	32.6%	21.7%	8.7%
ICT sec. personnel	4.3%	8.7%	50 %	30.4%	6.5%
IT architects	8.7%	34.8%	30.4%	13%	13%
Softw. dev	8.7%	39.1%	30.4%	10.9%	10.9%
Operations Personnel	8.7%	32.6%	37 %	15.2%	6.5%
External Consultants	21.7%	43.5%	15.2%	6.5%	13 %

TABLE IV. NOTICEABLE DIFFERENCES BETWEEN ATTENDS, SCALE FROM 1 (NEVER ATTENDS) - 4 (ALWAYS ATTENDS). (NOTE: THE RESPONDENTS CHOOSING "NOT PRESENT IN ORG." HAS BEEN REMOVED FROM THE SAMPLE)

		N	Minimum	Maximum	Range	Median	Grouped Median
@CEO	Small	12	0	4	4	3,00	2,67
	Medium	8	1	4	3	2,00	1,71
	Enterpr	26	0	4	4	1,00	1,53
CTO	Small	12	0	4	4	3,00	2,10
	Medium	8	0	4	4	2,00	2,00
	Enterpr	26	0	4	4	2,00	1,69

C. Critical knowledge areas in ISRA

Conducting an ISRA is a complex task with several different variables to consider, having discussed who attends risk assessments we look into critical knowledge areas to succeed with a risk assessment. So, we asked the participants to rank the importance of having knowledge about a set of items for the results of the ISRA (scale: 1 equals "not important" - 6 "very important"), Table V. For the comparison of knowledge areas the median is 5 for all but the *Organizational Structure* option, meaning that all were ranked highly by the respondents. Knowledge of *information assets* as the most important according to the mean score. Second, knowledge about *Laws & regulations* and *Information systems* were ranked equally, knowledge about *ISRA methods* was ranked the lowest. The diversity of the alternatives and the density of the results, supports that InfoSec is a very diverse field which demands a broad range of knowledge form its practitioners.

There was three noticeable differences between the expertise categories, the difference in view between experts and the two other groups on the importance of software, threat intelligence, and ISRA methods, Table VI. Whereas the experts valued threat intelligence less (grouped median = 4.75) than the proficient and the competent (grouped median = 5.13 and 5.47). There was also a slight difference in views between administrative (median=5, grouped median = 4.71) and technical workers (median=4, grouped median=4.71) on having knowledge of the organizational structure.

Two experts commented on the criticality of experience, *"The assessors experience is critical to a effective and accurate risk assessment"*, and *"Any method in use is only as good as the person(s) executing it and overall understanding of the business (or the part of business to evaluate) is critical to get results that are business beneficiary and useful to work with"*. Both comments highlights the need for experience, while the latter also highlights business understanding as key knowledge items. Our results also support this, as the top three ranked knowledge items relate to business understanding.

TABLE V. VIEWS ON IMPORTANCE OF KNOWLEDGE AREAS FOR ISRA. (*1 - Not Important* TO *6 - Very Important*

	1. Laws & Regulations	1. Info Assets	3. Info Systems	4. IT Infrastr & Hardware	5. Business Processes	6. Software
N	46	46	46	46	46	46
Min	2	3	3	3	1	3
Max	6	6	6	6	6	6
Median	5	5	5	5	5	5
Range	4	3	3	3	5	3
Mean	5,09	5,28	5,09	5,02	4,96	4,72
Std. Dev.	1,05	0,861	0,839	0,856	1,173	1,004
	7. Stakeholders & Employees	8. Organizat. Structure	9. ICT Architecture	10. Threat Intelligence	11. ISRA Methods	12. Pers. Expert & Experience
N	46	46	46	46	46	46
Min	1	2	3	2	1	3
Max	6	6	6	6	6	6
Median	5	4	5	5	5	5
Range	5	4	3	4	5	3
Mean	4,83	4,57	4,85	4,98	4,52	4,93
Std. Dev.	1,122	0,981	0,788	1	1,11	0,879

TABLE VI. NOTABLE DIFFERENCES ON KNOWLEDGE AREAS BETWEEN EXPERTISE GROUPS

		N	Min	Max	Range	Median	Grouped Median
Software	Competent	9	4	6	2	5,00	5,14
	Proficient	17	3	6	3	4,00	4,50
	Expert	20	3	6	3	4,50	4,67
Threat intel	Competent	9	4	6	2	5,00	5,13
	Proficient	17	2	6	4	6,00	5,47
	Expert	20	3	6	3	5,00	4,75
ISRA Methods	Competent	9	3	6	3	5,00	4,83
	Proficient	17	1	6	5	4,00	4,56
	Expert	20	3	6	3	5,00	4,50

IV. RISK ANALYSIS PRACTICES

Risk analysis (ISRAn) is the hands-on tasks performed during the assessment, primarily risk identification and estimation related tasks. This section starts with addressing some common issues regarding information assets, before investigating common risk analysis issues. We then survey the views of ISRAn methods and concepts.

We started the inquiry by asking an optional question on what the respondents thought to be working well in ISRAn. We got sixteen valid answers (eighteen total) with few common denominators, notably six respondents rated the risk assessment process to be working well, where two specified the risk identification phases to be well-developed. Two tech experts and one admin expert mentioned quantitative (numerical) ISRAn methods to be working well. While one tech and one admin expert answered that risk assessment on an overall works well, while *"implementation of risk mitigation and measurement follow up lags in many organizations."*

[1]Edited by author for readability, original answer *"having C or high management inside Information Security assessment not allow the participants to be open when providing input for risk identification."*

A. Views on Information Assets

Asset evaluation is one of the key challenges in ISRA [4], [10]. Due to being intangible, information assets can be particularly elusive to monetize and quantify. Which makes it hard to estimate, evaluate, and predict consequences of asset breaches in ISRA. To investigate issues regarding assets, we asked the participants to rate five statements regarding known issues on information assets [4]. Figure 4 shows the distribution of answers and Table VII displays descriptive statistics, typical of these results is a high variability in the answers.

With regards to Statement 1 (Table VII), the descriptives show that most practitioners agree that assigning monetary value is difficult, with the highest reported median 5 and mean 4.7, with no noticeable difference between groups. The results support the claims regarding information assets in Wangen & Snekkenes (2013) [4]).

The result from ranking Statement 2 regarding risk assessment method adequacy for asset evaluation, shows the sample mean being divided almost in the middle with a median of 3.67. The distribution for statement 2 is also close to normal but being negatively skewed (-0.299), Figure 4, and, therefore, ran significance tests. Our results showed that there was a statistically significant difference (P=0.031%) between expertise groups regarding Statement 2, regarding ISRA method adequacy, Table VIII, showing the Experts being less satisfied with the available asset value estimation methods. Three admin experts also commented on assigning the monetary value to assets, where two commented regarding asset evaluation not always being necessary: (i) *"The value doesn't necessarily be expressed in monetary terms."* (ii)*"... Knowing the value of personal information is not required to be able to protect it from unauthorized collection use of disclosure. The law says to do it."* These two insights show that asset evaluation is not always necessary, especially when the existing security legislation applies then a security classification is sufficient. While the third comment is on the importance of asset evaluation, (iii) *" Asset value can be assigned in various ways, and monetary value is in most cases the hardest one and most often wrongly set. Erroneously set values may in the worst case result in a totally erroneous assessment result. Asset value may have monetary value as one parameter but should be defined by much more than just a monetary number. E.g. if assets protected by law governed requirements are lost in the worst possible way, that may be "end of business," but that most often only relate to a small percentage of the total information assets of the business."*

Zhiwei [11] critiques the asset-based approach, and claims that protection of assets is not a primary goal of organizations, while priority number one should be the protection of the reliability and security of the organizations business processes. Statements 3 and 4 (Table VII addresses Zhiwei's view:

Regarding statement 3, most agreed that Asset protection is the primary goal of the InfoSec program, median = 5 and a mean = 4.37. However, there is a large variability in the results; nine respondents answered three or less showing that a minority disagrees with this statement. Out of this minority, six qualify as experts. The answer to statement 4 regarding the importance of asset security compared to ensuring stable

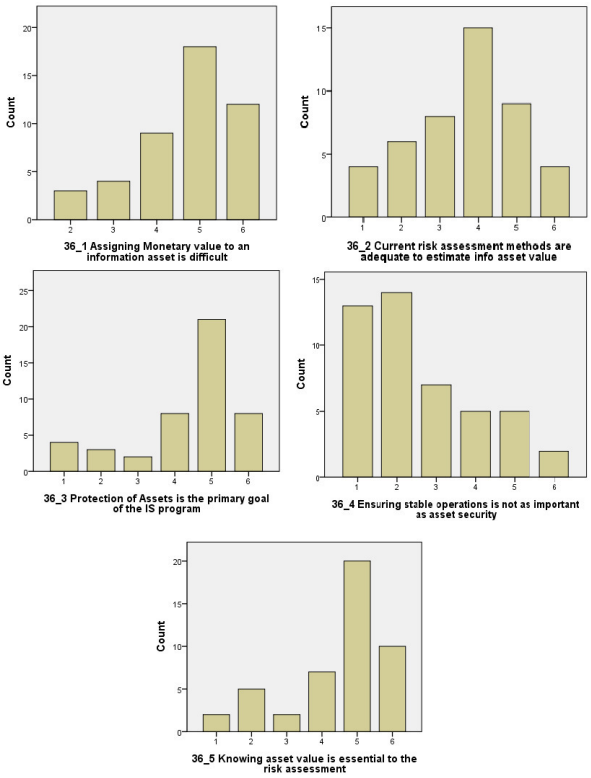


Fig. 4. Statements and rankings regarding Assets (Scale *1 - Strongly disagree* to *6 - Strongly agree*)

operations: The scores was on the low side (median = 2), showing that most of the respondents thought that stable operations are just as (or more) important than asset security. There was a notable difference between expertise groups for both Statement 3 and 4: The competent group consistently valued asset security higher than the proficient and expert group, indicating that protection priorities may be altered with experience in support of Zhiwei, Table VIII.

TABLE VII. PRACTITIONER VIEW ON ISSUES RELATED TO ASSETS. (SCALE 1 - STRONGLY DISAGREE, 6 - STRONGLY AGREE)

	N	Min	Max	Median	Range	Mean	Variance
1. Assigning Monetary value to an information asset is difficult	46	2	6	5	4	4,7	1,328
2. Current risk assessment methods are adequate to estimate info asset value	46	1	6	4	5	3,67	1,958
3. Protection of Assets is the primary goal of the IS program	46	1	6	5	5	4,37	2,149
4. Ensuring stable operations is not as important as asset security	46	1	6	2	5	2,59	2,248
5. Knowing asset value is essential to the risk assessment	46	1	6	5	5	4,48	1,988

TABLE VIII. STATISTICALLY SIGNIFICANT AND NOTABLE DIFFERENCES BETWEEN EXPERTISE CATEGORIES ON ASSETS

	Category	N	Mean	Std. Dev.	95% CI		Min	Max	
Asset Scenario					Lower Bound	Upper Bound			ANOVA, sig
2.	Competent	9	4,44	0,882	3,77	5,12	3	6	
	Proficient	17	3,94	1,298	3,27	4,61	2	6	
	Expert	20	3,1	1,483	2,41	3,79	1	6	
		46	3,67	1,399	3,26	4,09	1	6	.031
		9	Median	Range	Grouped Med				
4.	Competent	9	5	4	4.5		2	6	
	Proficient	17	2	5	1.92		1	6	
	Expert	20	2	3	2		1	4	
		46	2	5	2.29		1	6	

B. Views on common Risk Analysis issues

TABLE IX. DESCRIPTIVE STATISTICS OF ISRA STATEMENTS. (1 - STRONGLY DISAGREE, 6 - STRONGLY AGREE)

	N	Min	Max	Mean	Variance	Median	Skewness	Range
S1.Our ISRAn Methods are mainly Qualitative	46	2	6	4,41	1,537	5	-,414	4
S2.Our ISRAn Methods are mainly Quantitative/Statistical	46	1	6	3,26	2,597	3	,254	5
S3.It is easy to use the ISRAn results to predict the monetary cost of an incident	46	1	6	3,13	2,338	3	,161	5
S4.Our ISRAn method relies heavily on the security expert's predictions	46	1	6	3,87	1,405	4	-,574	5
S5.The resources spent on quantitative/statistical approaches are not worth the results	46	1	6	3,33	1,614	3	,472	5
S6.We find lack of historical data a problem for our risk forecasts/ predictions	46	1	6	4,17	1,614	4	-,341	5
S7.We lack a reliable method for mathematical ISRAn probability calculations	46	1	6	3,74	2,197	3	,087	5
S8.Annual Loss Expectation (ALE) is our preferred approach to calculating impact.	46	1	6	3,02	2,2	3	,474	5
S9.Our consequence/impact estimates of incidents tend to be precise	46	1	6	3,24	1,653	3	,252	5
S10.Consequences of occurred incidents tend to be outliers (extreme)	46	1	6	2,91	1,548	3	,316	5
S11.Causes for severe incidents/ disasters tend to not be thought of in our assessments	46	1	6	2,85	2,043	3	,518	5

TABLE X. DISTRIBUTION OF ANSWERS (X-AXIS) REGARDING ISRA STATEMENTS (Y-AXIS). STATEMENT NUMBERS CORRELATE WITH DESCRIPTIONS IN TABLE IX. (1 - STRONGLY DISAGREE, 6 - STRONGLY AGREE)

Statement nr	1	2	3	4	5	6
S1	0 (0%)	4 (8.7%)	7 (15.2%)	11 (23.9%)	14 (30.4%)	10 (21.7%)
S2	7 (15.2%)	10 (21.7%)	11 (23.9%)	5 (10.9%)	8 (17,4%)	5 (10.9%)
S3	8 (17.4%)	10 (21.7%)	10 (21.7%)	6 (13%)	10 (21.7%)	2 (4.3%)
S4	2 (4.3%)	3 (6.5%)	13 (28.3%)	10 (21.7%)	17 (37%)	1 (2.2%)
S5	2 (4.3%)	10 (21.7%)	18 (39.1%)	6 (13%)	7 (15.2%)	3 (6.5%)
S6	1 (2.2%)	3 (6.5%)	11 (23.9%)	10 (21.7%)	14 (30.4%)	7 (15.2%)
S7	2 (4.3%)	8 (17.4%)	14 (30.4%)	5 (10.9%)	10 (21.7%)	7 (15.2%)
S8	7 (15.2%)	12 (26.1%)	13 (28.3%)	4 (8.7%)	7 (15.2%)	3 (6.5%)
S9	4 (8.7%)	8 (17.4%)	18 (39.1%)	7 (15.2%)	7 (15.2%)	2 (4.3%)
S10	7 (15.2%)	8 (17.4%)	20 (43.5%)	5 (10.9%)	5 (10.9%)	1 (2.2%)
S11	9 (19.6%)	11 (23.9%)	14 (30.4%)	4 (8.7%)	6 (13%)	2 (4.3%)

The qualitative versus quantitative risk assessment is a well-known debate in ISRA [4], the former is mostly subjective knowledge-based and often describes risk using qualitative expressions, such as high, medium, and low. While the quantitative approach is mainly numerical and often based on statistical methods. There are arguments both for and against both approaches [4]. With the described issue at its core, we asked the participants to rank several statements regarding ISRAn practices, Table IX holds the statements with results and the distributions are in Table X. The results were diverse regarding all the statements, with the lowest median at 3 and highest at 5. In the following text, we analyze each statement with regards to descriptive statistics and correlation analysis. There are multiple differences between the three analyzed categories regarding nine of the statements, Table XI, and we analyze these differences together with the statement in question.

The results from Statement (S) 1, shows, with about 75% answering 4 or more, that most respondents consider their approach to be mainly qualitative. Worth noting is the minimum value of 2 in the results documenting that all of the participants consider their ISRAn methods to at least have some level subjectivity. S1 also has the highest median of 5 and lowest variability in the results. Regarding S2, less than half of the respondents consider their approaches to be more quantitative than qualitative, with 28% answering 5 or 6 indicating a mainly quantitative approach. Table XI shows that there is a notable difference between work types in this matter, whereas technical/hands-on practitioners view their approach as more quantitative. S2 regarding quantitative methods is also negatively correlated to S1 at the 0.05 level, Table XII.

In S3, regarding prediction of monetary costs, the median is 3 with a large variability in responses indicating that it is hard to predict the monetary cost of an incident based on ISRAn results. Also, the Expert group rated S3 lower than the other two groups, with the proficient group agreeing most with S3. Meaning that the experts in our sample find it harder to use the ISRAn results to predict the monetary cost of an incident.

The risks of being too reliant on expert predictions are that results can become too opinion-based, vulnerable to several external human factors, for example, emotional state and feelings [12], the Narrative Fallacy [13]), and involve a high degree of guesswork (see [4]). S4, regarding ISRAn reliance on expert predictions, the median is 4, with 87% of the responses being in the 3-5 range. There is notable difference between company sizes (Table XI), where small and medium companies seem more reliant on expert predictions than the enterprise-sized organizations.

Regarding S5, asks if spending resources on quantitative ISRAn are worth the results. The results show that majority (65%) answered 3 or less, while a minority (22%) answered 5 or more. However, there is a notable difference between technical and administrative work type (Table XI). Where the admin respondents consider quantitative risk assessments as a bigger waste of time than the tech respondents, which also corresponds to differences between these groups in S1 and S2.

Lack of historical data is claimed to be a consistent problem in InfoSec [4] and S6 addresses this issue. The median of 4 provides some evidence to support this assertion, there was also a notable difference between expert groups here, whereas the experts ranked this issue higher than the competent and proficient group.

Mathematical probability calculations is an issue with many opinions in the ISRA community [4], S7 and S8 connects to this issue. S7 addresses views on the adequacy of mathematical ISRAn methodology for probability calculations, with the results showing a difference of opinion on existing methods, the median of 3. There was a notable difference between the respondents from Small and Medium companies, ranking this issue higher than those from the Enterprises. The results are similar for S8, regarding Annual loss expectancy (ALE), although the difference is smaller for both total results and between the companies.

S9 addresses risk forecasting accuracy, and the results show that the respondents' general confidence in their predictions is on the low side. There was no notable difference between the expert groups indicating that confidence in precision has not improved with increased experience and expertise. However, there was a difference between company sizes, where the small and medium companies perceive a higher accuracy in their estimates. There is more complexity in larger organizations, which is one of the key challenges for prediction [14] and may be one of the causes.

Both S10 and S11 are connected to unforeseen incidents and causes, both related to Black Swan Risks [13] which are rare outlier risks that carry an extreme impact. Our results indicate that consequences of occurred incidents tend not be outliers and that causes for severe events/disasters are more often known than not. The analysis displays a difference between expert groups, with Experts being confident in their

knowledge about causes of incidents and disasters. From our results we see that most causes are believed to be known, and that Black and Grey Swan-type incident are very seldom. However, rare events and how they drive the InfoSec program is a path for future research.

This section has touched on one of the key challenges in ISRA, which is obtaining quantitative estimates of the probability of occurrence for security incidents, together with a reliable estimate of the consequence in a methodologically sound way. Which is difficult because of several reasons [14], [4], [10], where the factors that limit the forecasting are, for example, complexity, interconnectivity, and active adversaries. These factors do not apply for all InfoSec risks [14] and there is utility in obtaining statistical distributions of InfoSec risks [14]. As our results have shown, there are degrees of subjectivity to every risk assessment and one area to strengthen research is in risk quantification by working on obtaining probability distributions. In addition to combining both the quantitative and qualitative estimates in the risk model.

TABLE XI. NOTABLE DIFFERENCE BETWEEN CATEGORIES (FULL STATEMENTS CORRESPOND TO NUMBERS IN TABLE IX)

Statement	**Expertise**	N	Min	Max	Range	Median	Grouped Median
S3	Comp	9	2	5	3	3,00	3
	Proficient	17	1	6	5	4,00	3,80
	Expert	20	1	5	4	3,00	2,56
S6	Comp	9	2	5	3	4,00	4,17
	Proficient	17	1	6	5	4,00	3,70
	Expert	20	2	6	4	5,00	4,73
S11	Comp	9	1	5	4	4,00	3,75
	Proficient	17	1	6	5	3,00	2,70
	Expert	20	1	5	4	2,50	2,33
	Company Size						
S4	Enterpr	26	1	5	4	3,50	3,62
	Medium	8	3	5	2	4,50	4,33
	Small	12	3	6	3	4,50	4,38
S7	Enterpr	26	1	6	5	3,00	3,07
	Medium	8	2	6	4	5,00	4,60
	Small	12	2	6	4	5,00	4,71
S8	Enterpr	26	1	6	5	2,50	2,50
	Medium	8	2	6	4	2,50	2,67
	Small	12	1	6	5	3,50	3,67
S9	Enterpr	26	1	5	4	3,00	2,75
	Medium	8	2	6	4	3,00	3,25
	Small	12	3	6	3	4,00	3,88
	WorkType						
S1	Technical	17	2	6	4	4,00	4,27
	Admin	29	2	6	4	5,00	4,71
S2	Technical	17	1	6	5	4,00	4,00
	Admin	29	1	6	5	3,00	2,71
S5	Technical	17	1	5	4	3,00	2,73
	Admin	29	2	6	4	3,00	3,44

C. Correlations between statements

Several of the statements have strongly correlating results, Table XII. There is an interesting correlation regarding S2 on quantitative and statistical ISRAn methods: S2, is strongly correlated with S3 and S8, and weakly correlated with S9 and S11. The former correlations indicate that applying quantitative methods makes it easier to convert ISRAn results into monetary costs of incidents. The weak correlation to S9 indicates that working with risk quantification can improve precision and confidence in risk estimates. S3 is also strongly correlated with S8 and S9 further indicating that there are benefits from working with quantification and monetizing risk estimates. S3 is also negatively correlated with statement 1 in Table VII; *Assigning Monetary value to an information asset is difficult*. Further, the correlations test between the two sets of statements also indicates that gathering precise knowledge regarding asset value (36_5) correlates with confidence in consequence estimate precision. Another finding from this table is that prioritizing assets security as more important than stable operations (36_4) correlates with less insight into causes for severe incidents (S11).

Being reliant on expert predictions (S4) correlates strongly with the lack of historical data problem (S6) and lack of mathematical approach (S7) to ISRAn probability calculations. However, expert predictions also correlate with precision (S9), it seems a combination of mathematical models and expertise is then optimal. Lack of historical data (S6) also correlates with S10 and S11, indicating that historical data is necessary to prevent outliers and discover causes.

One Admin expert commented that *"Mathematical probability calculations are not worth anything if the organization does not believe in the probability of an incident occurring. Math alone is not the issue here. It is about the human ability to not just identify risk but accept risk presence (for real and react before the consequence of a corresponding issue hits)"*. Another Admin expert commented that *"There is still a lack of understanding of threat assessment as an input to identifying an actual risk."* The latter statement touches on the intersection between qualitative and quantitative methods since threat assessments are mainly subjective and can be more comprehensive than a purely quantitative approach being limited to observed data.

Consider the complexity and many aspects of loss calculations; one admin proficient commented: *"We consider the impact to business of loss of business (future) / customer impact, loss of reputation / brand impact, legal or regulatory breach and loss of money / financial impact."* Which highlights the many variables that must be considered in such calculations.

TABLE XII. CORRELATIONS BETWEEN ISRAN STATEMENTS. (FULL STATEMENTS CORRESPOND TO NUMBERS IN TABLE IX)

Statements		S2	S3	S5	S6	S7	S8	S9	S10	S11
S1	Pearson	-.367*		.333*	.363*					
	Sig.	,012		,024	,013					
	N	46		46	46					
S2	Pearson	1	.536**				.481**	.345*		.336*
	Sig.		,000				,001	,019		,022
	N	46	46				46	46		46
S3	Pearson		1				.440**	.425**		
	Sig.						,002	,003		
	N		46				46	46		
S4	Pearson				.443**	.385**		.400**	.474**	
	Sig.				,002	,008		,006	,001	
	N				46	46		46	46	
S6	Pearson				1	.414**			.460**	.321*
	Sig.					,004			,001	,030
	N				46	46			46	46
S8	Pearson						1	.428**		.337*
	Sig.							,003		,022
	N						46	46		46

*. Correlation is significant at the 0.05 level (2-tailed).
**. Correlation is significant at the 0.01 level (2-tailed).

D. Application of ISRAn methods and concepts

To obtain an insight into industry practice and adaptation of methods and concepts we compiled a non-exhaustive list of popular risk assessment tools and concepts, and asked how often they applied them in their ISRAn practice. Table XIII displays how the concepts were ranked by the participants. The three most frequently used methods are Business Impact Analysis, Penetration tests, and Security scanners, all with a median of 5. Cascading/correlating risks are the most

frequently applied concept for risk analysis. The items from *Component Testing* down to *Common Mode Failure* have medians between 3-2. The results show that methods for different genres of risk assessment (collected from [15], [13], [16]), such as Fault and Event tree analysis, HAZID, and HAZOP, are not common in ISRAn, where practitioners prefer methods developed specifically for InfoSec. Common concepts such as Black Swan Risks [13] and ALARP (As Low As Reasonably Practicable) [15] are also not widely known and applied by the surveyed practitioners. One admin expert commented on this particular issue: *Fault Tree Analysis, FMEA [Failure Mode and Effect Analysis], Hazop etc. are usually methods used by safety professionals, not information security professionals (I have however used them both but for slightly different purposes) and MTF or MTBF (Mean Time Before Failure) is typically also used in these safety oriented methods. I see the ability to merge methodologies between these areas of expertise for mutual benefit, but as far as I know, the industry does not do that in current operation.*

The same expert also commented on the three of the item's role of tools in reducing uncertainty: - *Different tools are in use for different purposes. I do not see penetration testing/security scanner/component testing as part of risk analysis. It is additional tools relevant to use if the risk evaluators are unable to be certain about probability - such testing can document probability and it also provides low-level insights to mitigation means.*

TABLE XIII. APPLICATION OF TOOLS, METHODS, AND CONCEPTS IN ISRA. (SCALE: 1 - UNFAMILIAR, 2 - VERY SELDOM, 3 - SELDOM, 4 - SOMETIMES, 5 - OFTEN, 6 - VERY OFTEN)

		N	Min	Max	Median	Range	Mean	Variance	Category
1	Business Impact Analysis	46	1	6	5	5	4,63	2.016	Method
2	PenTest	46	1	6	5	5	4,5	1,722	Method
3	Security Scanners	46	1	6	5	5	4,3	2,528	Concept
4	Cascading Risks	46	1	6	4	5	3,39	2.999	Method
5	Component Testing	46	1	6	2,5	5	2,96	3.109	Method
6	Mean Time To Failure	46	1	6	2,5	5	2,8	2,516	Method
7	Event Tree Analysis	46	1	6	2	5	2,93	2,773	Method
8	Fault Tree Analysis	46	1	6	2	5	2,65	2,810	Method
9	ALE/SLE	46	1	6	2	5	2,61	2.866	Method
10	FMEA	46	1	6	2	5	2,57	3.007	Method
11	Attack Trees	46	1	6	2	5	2,48	2,477	Method
12	OCTAVE	46	1	6	2	5	2,17	2,191	Method
13	Monte Carlo Simulations	46	1	6	2	5	2	1,467	Method
14	Common Mode Failure	46	1	6	1	5	2,39	2.955	Concept
15	Bayesian Networks	46	1	5	1	5	2,11	1.566	Method
16	Black Swan Risk	46	1	5	1	4	1,98	1,977	Concept
17	Antifragility	46	1	6	1	5	1,87	1.805	Concept
18	ALARP	46	1	6	1	5	1,7	1,416	Concept
19	CORAS	46	1	5	1	4	1,7	1.372	Method
20	HAZOP	46	1	5	1	4	1,65	1,032	Method
21	HAZID	46	1	5	1	4	1,61	1,088	Method

E. Cost-effectiveness of ISRA methods

As a follow up, we asked the participants which ISRAn method they considered to be most cost-effective, in which we received ten answers. There were no clear answer to this inquiry: Two Admin experts argued for Business Impact Analysis (BIA), as *"at the end of the day the systems that our business use are our main reason to have an IT area"*, and it *"can be done without bringing in external resources"*. BIA contains several tools and methods for reducing uncertainty related to consequences of risks.

Two argued (Admin expert and proficient) for security scanners and penetration tests (pentests), as *"they provide undeniable evidence of vulnerabilities. It is hard for someone to argue with them."* While two respondents (Admin expert and proficient) argued for the use of *Bowtie*-diagrams based on cause, threat, and risk analysis. We do not find Bowtie diagrams extensively described in the ISRA literature, although they are found in the more generic safety-related risk assessment literature, such as [15]. Bowtie are used for both risk analysis, visualization and communication.

F. What is the most important task of the ISRA?

There several tasks that are common when conducting an ISRAn [17], we gathered the common denominators and asked the participants to rate them according to their importance, 1 - Not important to 6 - Very important. Table XIV displays the results, with no notable difference between any groups. The participants ranked all the items highly, with lowest median being 4. The low end of the scale contains importance of knowledge about Stakeholders, Attacker capability, and Uncertainty. Whereas the remainder of the items are rated 5 or higher, meaning they are essential to the process. The respondents ranked Impact/consequences and threat as the most important tasks for the ISRA work.

TABLE XIV. VIEWS ON IMPORTANCE OF TASKS AND ITEMS FOR RISK ANALYSIS. (SCALE: 1 - NOT IMPORTANT, 6 - VERY IMPORTANT)

	N	Min	Max	Median	Range	Mean	Variance
1. Asset	46	1	6	5.5	5	5,15	1.287
2. Threat	46	3	6	6	3	5,33	0,936
3. Guardian/Control	46	3	6	5	3	5,02	1,133
4. Uncertainty	46	1	6	4	5	4,24	1,742
5. Probability/Likelihood	46	3	6	5	3	5,2	0,828
6. Impact/Consequences	46	3	6	6	3	5,37	0,638
7. Stakeholders	46	1	6	5	5	4,5	1,9
8. Attacker Capability	46	2	6	4	4	4,11	1,432
9. Vulnerability	46	3	6	5	3	5,24	0,586
10. Expert Knowledge	46	3	6	5	3	4,96	0,665

V. CHOOSING RISK TREATMENT STRATEGIES

Jaquith [18] claims that for most people, risk management really means risk identification, although these phases are clearly defined in the ISO/IEC vocabulary [1]. Applying ISO/IEC 27005:2011 [3] as a yard stick, the risk identification-phase clearly contains the majority of data collection and analysis. So, we asked the participants to rank the three different ISRA phases on importance. Table XV shows that the phases are almost equally ranked by our sample, with the risk identification scoring highest with a 6 median, otherwise, the difference between the phases are negligible.

TABLE XV. RANK THE PHASES OF THE ISRA PROCESS ACCORDING TO YOUR PERCEIVED IMPORTANCE, SCALE 1 (NOT IMPORTANT) - 6 (VERY HIGH IMPORTANCE)

	N	Min	Max	Median	Range	Mean	Variance
Risk Identification	46	4	6	6	2	5,57	,340
Risk Estimation	46	4	6	5	2	5,15	,532
Risk Evaluation	46	4	6	5	2	5,26	,464

Blakley et.al.[2] claims that the risk treatment strategies applied in IS focus primarily on risk mitigation, while transference, acceptance and avoidance as alternatives are seldom considered. The authors explain that the reason for this is the general approach to ISRM, where the practitioners are geared to imagining and then confirming technical vulnerabilities in information systems, so that steps can be taken to mitigate them. InfoSec activities rarely include any discussion of indemnity or liability transfer, although some organizations do address these issues in an "operational risk" organization

separate from the information security organization. Table XVI displays how the survey participants replied when we asked them how often they recommend the different risk treatment strategies for ISRA (scale 1 - Never, 2 - Very Seldom, 3 - Seldom, 4 - Sometimes, 5 - Often, and 6 - Very Often). Risk mitigation is the option ranked highest with 87% of respondents answering often or very often. This result supports Blakley et.al.'s claims about this strategy. However, the results also show that other strategies are frequently considered. The Blakley et.al. paper was written over a decade ago and the ISRA community may have matured in this area, although this is a field for future research. The *Transference* option is almost normally distributed, while the *Avoidance* option is bimodal with one top at *Sometimes* (39,1%) and one at *Very seldom* (19,6%). The *Acceptance/Retention* option is described by the median with 71% opting for *Sometimes* and *Often* alternatives. A clarification is provided by an admin expert with regards to type of industry: *"When it comes to health information, where regulatory requirements are very clear at placing the responsibility within the business, and a risk could lead to loss of life or health or patient confidentiality, transference is seldom an option."* Whereas another admin expert comment: *"Avoidance is seldom an option. Acceptance is most often already defined at some certain level in the business and is therefore most often not an option for any identified risks above defined threshold of acceptance. Optimisation is most often not prioritized until a result shows all risks identified to be below defined level of risk acceptance or as something to "think about" when all identified risks beyond acceptance threshold is reduced to a level within acceptable threshold."*

TABLE XVI. RESPONDENTS' RECOMMENDATION OF RISK TREATMENT OPTIONS IN ISRA. SCALE 1 (NEVER) TO 6 (VERY OFTEN)

	Valid	Min	Max	Median	Range	Mean	Variance
Transference	46	1	6	4,00	5	3,46	1,631
Mitigation	46	2	6	5,00	4	5,20	,872
Avoidance	46	1	6	4,00	5	3,76	1,608
Acceptance/Retention	46	2	6	4,00	4	4,15	1,065
Optimisation	46	2	6	4,00	4	4,30	1,150

Blakley et.al. also claims that InfoSec as a discipline focus more on reducing the probability of an event than on reducing its consequences. And where the focus is on reducing consequence, it tends to focus much more strongly on quick recovery (for example, by using aggressive auditing to identify the last known good state of the system) than on minimizing the magnitude of a loss through measures to prevent damage from spreading. We asked the participants which they thought more important, reducing the probability or consequence of the risk. Fig. 5 shows that the results are almost 50/50 distributed, no better than random. According our sample, there is no clear preference towards one or the other. With that said, this is often a two part process, where one can treat both probability and consequence of the risk to obtain a reasonable risk level. This issue was also highlighted to some extent by six of the twelve written comments to this question. The type of risk was also highlighted in four answers as a determining factor. One admin expert wrote: *"Proactive approach to risk reduction (i.e. probability) is most often chosen prior to reactive approaches (i.e. impact/consequence) as long as that is a feasible approach compared to cost of reactive approaches. The risk assessment result however, includes recommendations of both types for the business to conclude."* Also highlighting the need for cost/benefit analysis of the proposed risk treatment.

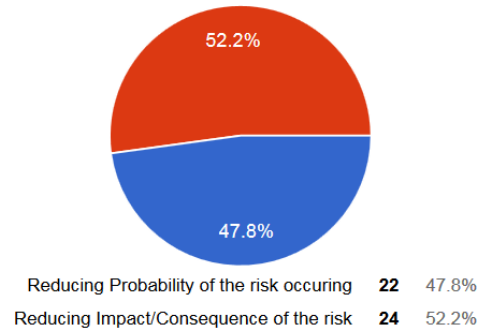


Fig. 5. Results from opting to reduce either probability or consequence

VI. SUMMARY & CONCLUSION

In this section, we first discuss the limitations of this study. Then, we conclude our findings, together with research implications and directions for future work.

A. Limitations

While our choice of online survey allowed us to recruit participants from our target group through specialized web-forums, this approach has some limitations. First of all, our data are self-reported values based on participants perceptions, while not a substitute for behavioral and observational data from real-world scenarios, this self-reported data can still provide valuable insight into day-to-day practices and how practitioners view the research problems. Furthermore, the study design and recruitment process gave us less control of the research participants; the control questions somewhat mitigated this problem, but these were not fool-proof, and circumvention was possible. The sample size was also small, although the online groups and forums exposed the survey to many potential respondents we only managed to recruit forty-six in one month. Based on the many members of these groups, the recruitment strategy was not a success. Many restricting factors could have caused this outcome, for example, activity in the forums, exposure of the survey, and questionnaire length. Although the sample had a good geographical spread and diverse background from the participants, this small sample is sensitive to outliers. The written responses and comments are more anecdotal evidence.

Another limitation of this study is concerning what is not asked for, issues we are not aware of or not present in the questionnaire can not be answered. We partially addressed this issue by adding with comment sections in the questionnaire, but this issue is likely better addressed in open interviews.

B. Conclusion & Future Work

InfoSec risk management and assessment are essential to well-functioning InfoSec program as it determines what to protect and how. In this paper, we have addressed three major areas of practice in ISRM and provided incentives to strengthen

research within them; on the ISRA level, we found that the majority did not differentiate between ISRA methods for different organizational tiers. However, several respondents did distinguish, for example through formality, and handled risks at the higher abstraction levels more formally. As a future direction, we propose to research handling and assessing risk between the organizational tiers, together with risk escalation issues.

Gathering the ISRA team and securing the right knowledge is essential to the assessment; Our results showed that the CISO/CSO and InfoSec personnel most frequently leads and attends risk assessments while various roles in IT department attends based on the scope of the assessment. Knowledge about information assets and business understanding was highlighted as essential, together with knowledge about laws & legislation stressing the importance of legal counsel in the ISRA. Composition and optimization of the ISRA team from the knowledge perspective is a potential path for future research.

Throughout the results, several respondents highlighted the significance of the risk assessors experience for the results, as *any method is only as good as the person executing it.* On qualitative and quantitative approaches, we found that the majority of ISRAn approaches are qualitative. While those who described their work as more technical were more likely to describe their ISRAn approach as quantitative. Our analysis shows that confidence in impact estimates precision tends to be low, however, working with risk quantification is likely to improve accuracy and trust in risk estimates. Which highlights the importance of both the expert and the benefits working with quantification. A path for future work is to research the intersection between these two approaches to optimize the ISRA results.

Related to the precision in impact estimation, we found that Black Swan theory is very seldom applied in ISRA. Possible paths for future work is an analysis of InfoSec risks and how they relate to Black Swans, together with research on rare events and how they drive the InfoSec program. We have provided incentives for strengthening research within obtaining probability distributions for frequencies and consequences for InfoSec, as this is an area that has a potential for producing useful knowledge for decision-makers.

Worth noting is that experts ranked the importance of threat intelligence for ISRA lower than the less experienced groups. On the risk analysis practices, this study documented that asset evaluation is a challenge, with experts considering the existing risk assessment methods as not sufficient to handle this problem. The participants also ranked knowledge about assets as important in multiple instances in the results which make asset evaluation stand out as an issue for future research.

From our list of suggested tools and concepts Business impact analysis, penetration tests, and security scanners are the most frequently applied tools for ISRA. Together with Bowtie-diagrams, these methods and tools are deemed the most cost-effective.

ACKNOWLEDGMENT

The Author thanks professors Einar Snekkenes for discussion, and my colleagues Andrii Shalaginov, Ambika Shrestha Chitrakar, Yi-Ching Lao and Goitom Weldehawaryat for quality assurance. Professor Stewart Kowalski for his knowledge on Likert-scales and analysis. We extend a thanks to all who answered the questionnaire, the anonymous reviewers for their comments, and the support from the COINS Research School for InfoSec.

REFERENCES

[1] *Information technology, Security techniques, ISMS, Overview and vocabulary*, International Organization for Standardization Norm, ISO/IEC 27000:2014. [Online]. Available: http://dx.doi.org/10.3403/30236519

[2] B. Blakley, E. McDermott, and D. Geer, "Information security is information risk management," in *Proceedings of the 2001 workshop on New security paradigms*. ACM, 2001, pp. 97–104.

[3] *Information technology, Security techniques, Information Security Risk Management*, International Organization for Standardization Std., ISO/IEC 27005:2011.

[4] G. Wangen and E. Snekkenes, "A taxonomy of challenges in information security risk management," in *Proceeding of Norwegian Information Security Conference / Norsk informasjonssikkerhetskonferanse - NISK 2013 - Stavanger*, vol. 2013. Akademika forlag, 2013.

[5] S. Fenz, J. Heurix, T. Neubauer, and F. Pechstein, "Current challenges in information security risk management," *Information Management & Computer Security*, vol. 22, no. 5, pp. 410–430, 2014.

[6] G. Wangen, "An initial insight into infosec risk management practices," in *Proceeding of Norwegian Information Security Conference / Norsk informasjonssikkerhetskonferanse - NISK 2015 - Aalesund*, vol. 2015. Open Journal Systems, 2015.

[7] A. G. Kotulic and J. G. Clark, "Why there aren't more information security research studies," *Information & Management*, vol. 41, no. 5, pp. 597–607, 2004. [Online]. Available: http://dx.doi.org/10.1016/j.im.2003.08.001

[8] G. A. Churchill Jr, "A paradigm for developing better measures of marketing constructs," *Journal of marketing research*, pp. 64–73, 1979.

[9] G. Locke and P. Gallagher, "800-39 nist sp, managing information security risks - organization, mission, and information systems view," National Institute of Standards and Technology: U.S. Department of Commerce, Tech. Rep., 2008.

[10] S. Fenz and A. Ekelhart, "Verification, validation, and evaluation in information security risk management," *Security Privacy, IEEE*, vol. 9, no. 2, pp. 58–65, 2011.

[11] Y. Zhiwei and J. Zhongyuan, "A survey on the evolution of risk evaluation for information systems security," *Energy Procedia*, vol. 17, pp. 1288–1294, 2012.

[12] G. F. Loewenstein, E. U. Weber, C. K. Hsee, and N. Welch, "Risk as feelings." *Psychological bulletin*, vol. 127, no. 2, p. 267, 2001.

[13] N. N. Taleb, *The Black Swan: The Impact of the Highly Improbable*, 2nd ed. Random House LLC, 2010.

[14] G. Wangen and A. Shalaginov, *Risks and Security of Internet and Systems: 10th International Conference, CRiSIS 2015, Mytilene, Lesbos Island, Greece, July 20-22, 2015, Revised Selected Papers*. Cham: Springer International Publishing, 2016, ch. Quantitative Risk, Statistical Methods and the Four Quadrants for Information Security, pp. 127–143. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-31811-0_8

[15] T. Aven, W. Røed, and H. S. Wiencke, *Risikoanalyse (Norwegian Ed)*. Prinsipper og metoder, med anvendelser. Oslo: Universitetsforlaget, 2008.

[16] N. N. Taleb, *Antifragile: things that gain from disorder*. Random House LLC, 2012.

[17] G. Wangen, C. Hallstensen, and E. Snekkenes, "A framework for estimating information security risk assessment method compcomplete - core unified risk framework," in *[Under Revision]*. .., 2016.

[18] A. Jaquith, *Security metrics: replacing fear, uncertainty, and doubt*. Addison-Wesley Upper Saddle River, 2007.

5th International Symposium on Frontiers in Network Applications, Network Systems and Web Services

SYMPOSIUM SoFAST-WS focuses on modern challenges and solutions in network systems, applications and service computing. The Symposium builds upon the success of Frontiers in Network Applications and Network Systems (FINANS'2012) and 4th International Symposium on Web Services (WSS' 2012) held in 2012 in Wroclaw, Poland. These two events are now integrated into one event to fully exploit the synergy of topics and cooperation of research groups.

The topics discussed during the symposium include different aspects of network systems, applications and service computing. The primary objective of the symposium is to bring together researchers and practitioners analyzing, developing and administering network systems, with particular emphasis on Internet systems. Authors are invited to submit their papers in English, presenting the results of original research or innovative practical applications in the field.

TOPICS

- Architecture, scalability and security of Open API solutions,
- Technical and social aspects of Open API and open data,
- Service delivery platforms—architecture and applications,
- Telecommunication operators API exposition in Telco 2.0 model,
- The applications of intelligent techniques in network systems,
- Mobile applications,
- Network-based computing systems,
- Network and mobile GIS platforms and applications,
- Computer forensic,
- Network security,
- Anomaly and intrusion detection,
- Traffic classification algorithms and techniques,
- Network traffic engineering,
- High-speed network traffic processing,
- Heterogeneous cellular networks,
- Wireless communications,
- Security issues in Cloud Computing,
- Network aspects of Cloud Computing,
- Control of networks,
- Standards for Web services,
- Semantic Web services,
- Context-aware Web services,
- Composition approaches for Web services,
- Security of Web services,
- Software agents for Web services composition,
- Supporting SWS Deployment,
- Architectures for SWS Deployment,
- Applications of SWS to E-business and E-government,
- Supporting Enterprise Application Integration with SWS,
- SWS Conversational Protocols and Choreography,
- Ontologies and Languages for Service Description,
- Ontologies and Languages for Process Modeling,
- Foundations of Reasoning about Services and/or Processes,
- Composition of Semantic Web Services,
- Innovative network applications, systems and services.

EVENT CHAIRS

- **Furtak, Janusz,** Military University of Technology, Poland
- **Grzenda, Maciej,** Orange Labs Poland and Warsaw University of Technology, Poland
- **Legierski, Jarosław,** Orange Labs Poland, Poland
- **Luckner, Marcin,** Warsaw University of Technology, Poland
- **Szmit, Maciej,** Orange Labs Poland, Poland

PROGRAM COMMITTEE

- **Benslimane, Sidi Mohammed,** University of Sidi Bel-Abbès, Algeria
- **Chojnacki, Andrzej,** Military University of Technology, Poland
- **Cocucci, Osvaldo,** Orange Labs Products & Services, France
- **Fernández, Alberto,** Universidad Rey Juan Carlos, Spain
- **García-Domínguez, Antonio,** University of York, United Kingdom
- **Gibert, Philippe,** Orange Labs Products and Services, France
- **Kaczmarski, Krzysztof,** Warsaw University of Technology, Poland
- **Katakis, Ioannis,** National and Kapodistrian University of Athens, Greece
- **Kiedrowicz, Maciej,** Military University of Technology, Poland
- **Korbel, Piotr,** Lodz University of Technology, Poland

- **Kowalczyk, Emil,** Orange Labs, Poland
- **Kowalski, Andrzej,** Orange Labs, Poland
- **López Nores, Martín,** University of Vigo, Spain
- **Maamar, Zakaria,** Zayed University, United Arab Emirates
- **Macukow, Bohdan,** Warsaw University of Technology, Poland
- **Misztal, Michal,** Military University of Technology, Poland
- **Nowicki, Tadeusz,** Military University of Technology, Poland
- **Richomme, Morgan,** Orange Labs, France
- **Wrona, Konrad,** NATO Consultation, Netherlands
- **Zieliński, Zbigniew,** Military University of Technology, Poland
- **Żorski, Witold,** Military University of Technology, Poland

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1011–1014
DOI: 10.15439/2016F283
ACSIS, Vol. 8. ISSN 2300-5963

IoT gateway – implementation proposal based on Arduino board

Artur Grygoruk
R&D Center Orange Poland
ul. Obrzeżna 7
02-691 Warsaw, Poland
Email: artur.grygoruk@orange.com

Jarosław Legierski
R&D Center Orange Poland
ul. Obrzeżna 7
02-691 Warsaw, Poland
Email: jaroslaw.legierski@orange.com

Warsaw University of Technology
Faculty of Electronics and Information Technology
ul. Nowowiejska 15/19 00-665 Warsaw, Poland

Warsaw University of Technology, Faculty of Mathematics and Information Science,
ul. Koszykowa 75, 00-662 Warsaw, Poland

Abstract — The paper presents proposal of practical implementation simple IoT gateway based on Arduino microcontroller, dedicated to use in home IoT environment. Authors are concentrated on research of performance and security aspects of created system. By performed load tests and denial of service attack were investigated performance and capacity limits of implemented gateway.

I. Introduction

IoT gateway concept is one of the most important aspect in Internet of Things idea. This network element is presented as a proxy between sensing network and application layers. Many small and autonomous devices and sensors urgently needs this component for communication with higher layers of the network.

In contemporary world of digital communication one of the major aspects is data transmission protection. The ways of implemented protection mechanisms depends on infrastructure details and characteristics of services dedicated to end users. IoT Gateway is a very interesting approach from this point of view and very often a single point of failure for IoT infrastructure. IoT gateway installed in single instance can be observed as Single Point of Failure [2] and is really vulnerable to all threats which are based on network traffic volumetric attack.

II. Existing Solutions

There can be defined two types of IoT gateway implementations. The first one: gateway installed in form of dedicated software located e.g. in typical wireless router or in smartphone as an application [2]. The second implementation is based on a gateway which is using a dedicated hardware. IoT gateway must meet requirements such as: hardware low cost, easy extensibility and application-layer support. The fact is that standardized to different network platforms IoT gateway doesn't exist. Each type of IoT device and each vendor uses own IoT gateway implementation e.g. on smartwatches market each supplier can offer client buying his own IoT gateway application which is installed in smartphone device. This approach is generally different than presented in Wifi segment where each computer, tablet or smartphone can use one common Wifi gateway.

Because of low hardware cost, availability and possibility of modifying the hardware and software in very easy way, the authors of this paper created a prototype of IoT gateway based on Arduino platform. Arduino is an open-source electronics single board microcomputer based on easy implementable hardware and software. In the literature we could find many examples of using Arduino (mostly used as sensor node) and different most advanced platform such as Raspberry Pi which were used to build the IoT Gateway: [4],[5],[6], [7]. In the literature is presented only one similar IoT gateway implementation based on Arduino. In [3] authors presents the concept of Arduino board based IoT gateway dedicated specially to the medical purposes. This gateway implemented on Arduino Yun is dedicated for monitoring vital parameters of human body using different sensors such as: heart rate sensor, blood pressure, pulse oximetry or body temperature.

III. IoT home gateway concept

IoT Gateway is a single place where a lot of sensors and other components can communicate with applications using standard protocols included in wireless technology like mobile networks or WiFi. Sensing domain elements are connected to one root component which is a point of communication with rest part of devices [1]. IoT home gateway [2] can be defined as an element connecting simple sensors installed in dedicated network often connected wireless e.g. using BLE (Bluetooth Low Energy) technology with home network layer.

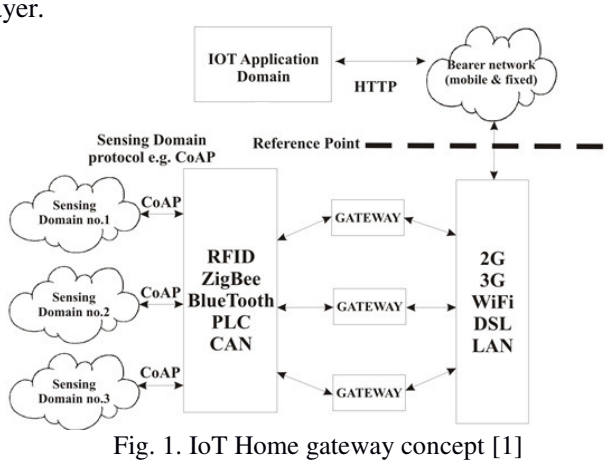


Fig. 1. IoT Home gateway concept [1]

As it was presented on Figure 1. IoT gateway fulfils a role of point between sensors and network domain, where tiny sensors can communicate with other subnetworks. For IoT networks it is recommended to use protocols which packets have low overhead and stateless communication method e.g. Constrained Application Protocol (CoAP). CoAP is an application part protocol to present readings from sensors in unreliable transmission. It uses only UDP datagrams to send information very fast and with low latency. It is great protocol for reading sensor data and controlling actuators to drive motors.

IV. SYSTEM ARCHITECTURE

On Figure 2. high level system architecture was presented. Arduino based on IoT gateway (2) collects an information from different sensors from sensing domain (1). In tested environment: Passive Infra Red sensor, sound sensor, pressure and smoke sensors were used.

Web Application (3) hosted on board presents information from sensors and exposes them for third party systems. In implemented gateway two different Ethernet Modules (4) (Ethernet ENC28j60 and Wiznet550 with tcp offload capability) were used to provide connection to network domain. As IoT gateway system core element three boards: Arduino UnoR3, Mega2560 and Leonardo (2) were tested. As third party application http client (6) e.g. web browser was connected to network domain (5).

Moreover, the IoT Gateway designed architecture provides a connection to web service which presents in browser the information returned by each sensor.

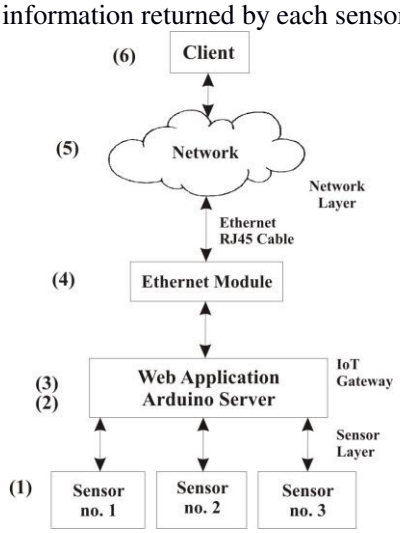


Fig. 2. Iot Gateway System Architecture

However, it should be emphasized that presented architecture is designed to perform IoT Gateway tests according to performance and security aspects. Two main most probably issues are Flash Crowd and (D)DoS (eng. Distributed Denial of Service) attacks. The first one occurs when too many legitimate users want to get access to the service in the same moment. IoT Gateway is vulnerable due to the fact that Arduino has low capacity of RAM and CPU parameters. The second one is a malicious attack using thousands of network machines to block a service.

A) *Arduino environment*

Two web applications which fulfil a crucial role of plugin for sensing domains, have been developed and installed on Arduino Leonardo, Mega2560 and UnoR3 supported by different Ethernet Modules (hardware Wiznet550 and software ENC28j60). Configuration of Ethernet Network Module which is presented on Fig. 3. was provided by ICSP pin with settings and configuration on board side. User can send and read information from sensors by using http client and dedicated web application

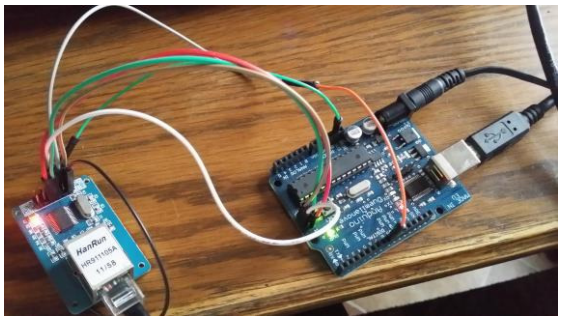


Fig. 3. IoT gateway prototype with ENC28j60 card.

B) *Web Application*

Web application was hosted on Arduino board. It was a standard web server which exposed information from connected sensors. It displayed a single html page with presented a few changing values of sensor readings after refreshing the web page content from web browser.

V. MEASUREMENTS

A) *Test environment*

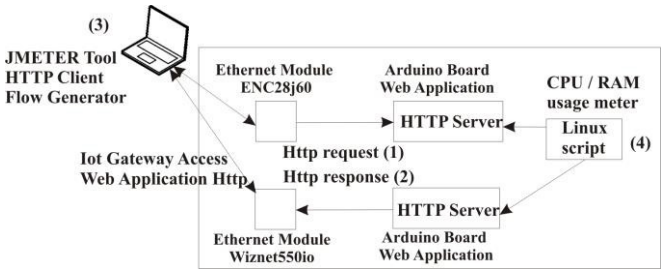


Fig. 4. Load test concept

Fig. 4. presents test environment used for IoT gateway performance tests. Based on Apache JMeter tool (3) http traffic to Arduino Gateway was generated (1). IoT gateway returned web page (2) with information from sensors (with http 200 message). In other cases no response or http 404 message was generated. CPU/RAM meter (4) based on Linux script running on dedicated Laptop was used to measure Arduino device performance.

Moreover, the second part of the research was performed with usage of Hping3 security tool. It gives a lot of possibilities to generate a huge traffic simulating Distributed Denial of Service. TCP SYN Flood and Http Slowloris attacks were used to block an access to a web service. Unfortunately, it could be observed that service is very sensitive to receiving big part of http traffic. It was unresponsive to next requests sent by clients. It had to be restarted to provide access after web logic failure.

B) Load test – error rate

Based on test scenario described in details in point A of this paper load tests for three Arduino boards and two Ethernet cards were performed. Figures 5-10 presents results of load tests (observed packet error rate in network traffic).

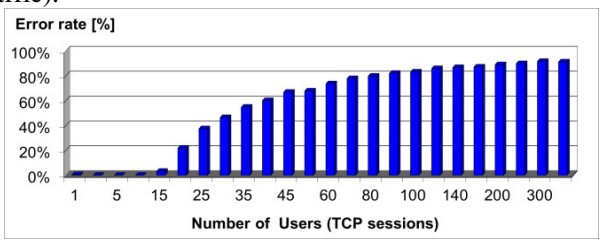


Fig. 5. Error rate (Arduino Leonardo + Wiznet550)

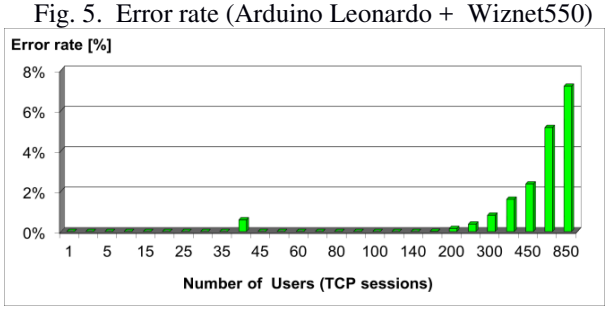


Fig. 6. Error rate (Arduino Leonardo + Enc28j60)

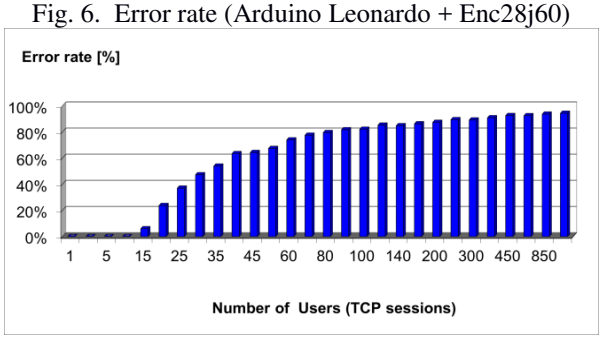


Fig. 7. Load Error rate (Arduino Mega2560 + Wiznet550)

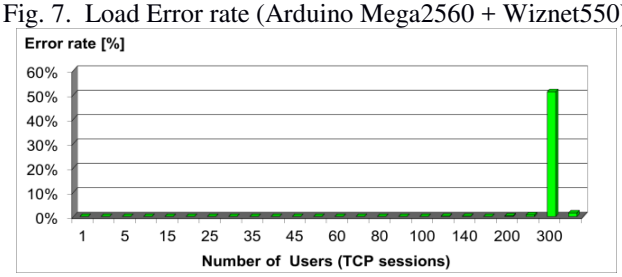


Fig. 8. Error rate (Arduino Mega2560 + Enc28j60)

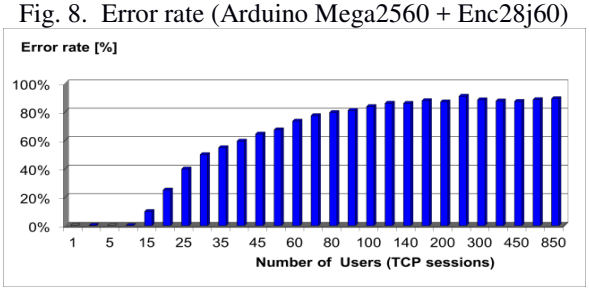


Fig. 9. Error rate (Arduino UnoR3 + Wiznet550)

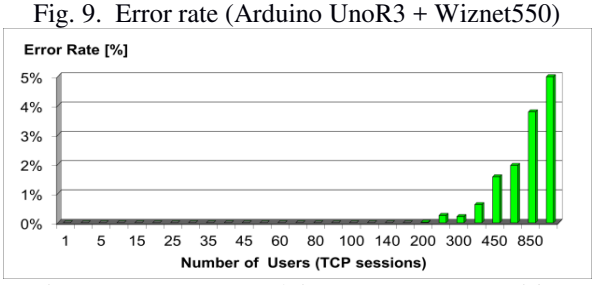


Fig. 10. Error rate (Arduino UnoR3 + Enc28j60)

Moreover, based on the result of load test for Wiznet550 network adapter module maximal number of 15 simultaneous sessions was defined. Large number of sessions results in showing high error rate of receiving packets. The better result was observed for ENC28J60 Ethernet card and for this component the lower error rate about 5-8% for 300-800 TCP sessions was detected.

A) Load test – average response time

Besides the error rate during the load tests average response time was measured.

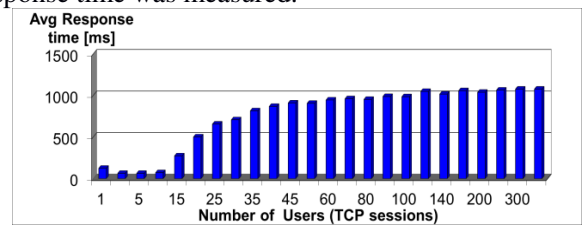


Fig. 11. Avg. response time (Arduino Leonardo+Wiznet550)

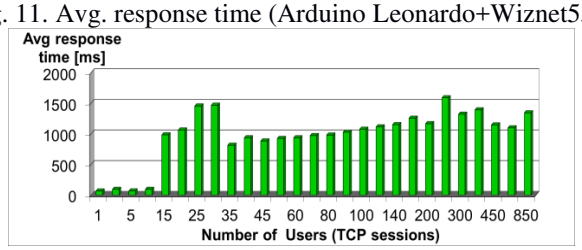


Fig. 12. Avg response time (Arduino Leonardo + Enc28j60)

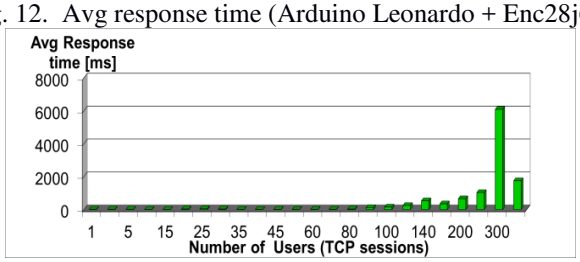


Fig. 13. Avg response time (Arduino Mega2560+Enc28j60)

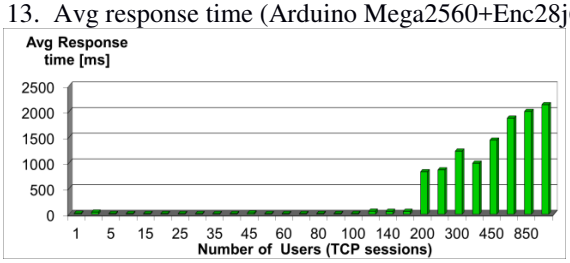


Fig. 14. Avg response time (Arduino UnoR3 + Enc28j60)

Based on requirements it should be emphasized that for the real time communication systems maximum acceptable response time should not exceed 100 ms and this value were reached for 10 simultaneous TCP sessions for Arduino Leonardo with Wiznet550 card (Fig 11). For others Arduino boards and Wiznet550 adapter we can observe similar values. For Enc28j60 network card we can observe better performance (avg response time 94 ms for 10 TCP sessions for Arduino Leonardo, time 125 ms for 140 TCP sessions for Arduino Mega2560 and 56 ms for 160 TCP sessions for UnoR3).

Table 1 and 2 presents observed CPU and RAM usage. It can be observed that percentage usage of CPU for Arduino board is higher about 10% when Enc28j60 Ethernet module was connected to device. It is a result of receiving and parsing packets on core of application web server. When Wiznet550 tcp off load connector is used a

big part of packet processing is handled by this element. It protects presented gateway against change of web server logic into block state.

Table. 1 Load test result – CPU usage [%]

Number of Threads	Arduino Leonardo Wiznet	Mega2560 Wiznet	UnoR3 Wiznet	Arduino Leonardo ENC28j60	Arduino Mega2560 ENC28j60	Arduino UnoR3 ENC28j60
1	4,500	1,065	6,882	15,900	10,317	-
5	1,667	1,008	6,981	22.367	10,138	9,991
10	9,333	0,935	6,036	15,367	9,608	9,802
20	6,133	0,852	7,233	16,800	1,276	10,020
30	3,467	0,898	7,226	10,233	1,123	9,945
40	3,500	0,797	7,638	13,133	0,999	9,836
50	12,467	1,113	7,326	9,467	1,126	9,797
100	10,500	0,906	5,638	16,200	1,042	10,235
200	6,833	0,805	6,112	6,300	1,137	-
300	8,700	0,808	6,730	13,233	0,949	-

Table 2. Load test result – RAM usage [%]

Number of Threads	Arduino Leonardo Wiznet	Arduino Mega2560 Wiznet	Arduino UnoR3 Wiznet	Arduino Leonardo ENC28j60	Arduino Mega2560 ENC28j60	Arduino UnoR3 ENC28j60
1	35,900	14,627	38,833	35,850	24,578	-
5	37,500	14,700	38,851	34,900	24,827	38,474
10	38,600	14,157	38,985	34,500	24,941	38,565
20	38,100	14,780	38,822	35,133	14,479	38,531
30	37,900	14,793	38,731	35,300	14,497	38,539
40	38,400	14,791	38,682	34,867	14,699	38,483
50	37,500	14,792	38,827	34,200	14,690	38,507
100	39,400	14,788	39,314	34,800	14,722	38,197
200	38,900	14,800	39,310	39,300	19,700	-
300	38,733	14,800	39,191	38,900	19,669	-

RAM percentage usage results presented for Arduino boards don't depend on kind of Ethernet module.

Sometimes the usage WIZnet550io module resulted in the error XML Parse Exception. It is a result of blocking a part of data on the same link which is used by legitimate user and by attackers. Service doesn't have to be suspended in case of receiving too many requests to web application. Packet error rate (PER) has always been the biggest value for Ethernet module WIZnet550io. In contrast to Wiznet550io Ethernet ENC28J60 requires a large number of users to return permanent failure. It leads to rejection of the packets due to insufficient capacity of Arduino board resources at very high network traffic.

VI. FUTURE WORK

In the future, the authors of this publication are planning to focus on improving security aspects. The most important is implementation of encryption for data transmission between the IoT gateway and http client [8]. Nowadays a lot of Man in The Middle attacks can be performed successfully because of sending data via plain text. Implementation of SSL protocol and certificate on web server side should prevent against this situation.

Due to the fact that only limited resources are available such as: CPU, memory and number of I/O ports the implementation of similar IoT environment based on more advanced hardware platform should be performed. For example the usage of Intel Galileo should allow to implement some additional features such as Web application or firewall features inside the board because of better hardware capabilities in comparison with Arduino board. The use of Intel Galileo allows to keep compatibility of sensors layer with the platform Arduino and allows to focus on the software development.

VII. SUMMARY

Presented in this paper Arduino boards are offered as a standalone device without network adapter. For load tests Arduino boards were equipped with Wiznet550io module and ENC28j60 network adapter. Better results of performance were observed for ENC28j60 from 10 TCP sessions for Arduino Leonardo to 160 TCP sessions for Arduino UnoR3.

As a device with very limited hardware Arduino can host services which can't be stable if too many users want to get access to the resources in the same time. Arduino board can be used as the IoT gateway only in very small IoT environment. ENC28j60 Ethernet module could better cope with Http request avalanche when in the same time Wiznet550io module was really poor to protect IoT gateway. However, it should be emphasized that there is a need to build new solution for Iot gateway protection.

Prototype of IoT Gateway was made as part of the Open Middleware 2.0 Community by Orange Labs program [9].

REFERENCES

[1] Hao Chen, Xueqin Jia and Heng Li, "A brief introduction to IoT gateway," *Communication Technology and Application (ICCTA 2011), IET International Conference on*, Beijing, 2011, pp. 610-613. doi: 10.1049/cp.2011.0740

[2] Thomas Zachariah, Noah Klugman, Bradford Campbell, Joshua Adkins, Neal Jackson, and Prabal Dutta, "The Internet of Things Has a Gateway Problem," Electrical Engineering and Computer Science Department University of Michigan Ann Arbor, MI 48109

[3] Boopala krishnan. N, Siva Sankara Sai. S and S. B. Mohanthy, "Real Time Internet Application with distributed flow environment for medical IoT," Green Computing and Internet of Things (ICGCIoT), 2015 International Conference on, Noida, 2015, pp. 832-837.

[4] S. M. Kim, H. S. Choi and W. S. Rhee, "IoT home gateway for auto-configuration and management of MQTT devices," Wireless Sensors (ICWiSe), 2015 IEEE Conference on, Melaka, 2015, pp. 12-17.

[5] http://thenewstack.io/tutorial-prototyping-a-sensor-node-and-iot-gateway-with-arduino-and-raspberry-pi-part-1/ [08.05.2016]

[6] A. Herutomo, M. Abdurohman, N. A. Suwastika, S. Prabowo and C. W. Wijiutomo, "Forest fire detection system reliability test using wireless sensor network and OpenMTC communication platform," Information and Communication Technology (ICoICT), 2015 3rd International Conference on, Nusa Dua, 2015, pp. 87-91.

[7] A. E. Boualouache, O. Nouali, S. Moussaoui and A. Derder, "A BLE-based data collection system for IoT," New Technologies of Information and Communication (NTIC), 2015 First International Conference on, Mila, 2015, pp. 1-5.

[8] P. Wawrzyniak, Ł. Wronkowski, D. Kuniszewski, A. Cackowski, P. Czapliński i K. Szymański "Send It Safe – A Novel Application for Secure Key Exchange Using Telecommunications Open Middleware APIs," Frontiers in Network Applications, Network Systems and Web Services (SoFAST-WS'14), Federated Conference on Computer Science and Information Systems FedCSIS 2014, Warszawa 2014

[9] Open Middleware 2.0 Community portal – http://www.openmiddleware.pl [08.05.2016]

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1015–1019
DOI: 10.15439/2016F194
ACSIS, Vol. 8. ISSN 2300-5963

Time-Dependent Queue-Size Distribution in a Finite-Buffer Model with Server Setup Times

Wojciech M. Kempa
Silesian University of Technology,
Institute of Mathematics,
ul. Kaszubska 23, 44-100 Gliwice, Poland,
Email: wojciech.kempa@polsl.pl

Dariusz Kurzyk
Silesian University of Technology,
Institute of Mathematics,
ul. Kaszubska 23, 44-100 Gliwice, Poland,
Institute of Theoretical and Applied Informatics,
Polish Academy of Sciences,
ul. Bałtycka 5, 44-100 Gliwice, Poland,
Email: dariusz.kurzyk@polsl.pl

***Abstract*—Transient queue-size distribution in a finite-buffer system with Poisson arrivals and generally distributed processing times is investigated. In the evolution of the system the server needs randomly distributed setup times preceding the service initialization in each new busy period. Applying the paradigm of embedded Markov chain and the formula of total probability, a Volterra-type system of integral equations for the transient queue-size distribution, conditioned by the number of packets being accumulated in the buffer before the opening of the system, is built. The solution of the corresponding system written for Laplace transforms is obtained algebraically in the compact explicit form. Numerical examples are attached as well.**

I. Introduction

FINITE-BUFFER queues are widely used in modelling real-life processing systems, in which the phenomena like buffering of waiting packets (jobs, customers, calls, etc.), delays of different nature, and packet losses caused by buffer overflows or temporary suspensions of the service process occur. In particular, e.g. the incoming/outcoming stream of packets in a node of computer network (like, e.g. IP router in the Internet) can be efficiently modelled by using one of finite queueing systems. One of the most important challenges of wireless communication (e.g., based on Wi-Fi IEEE.802.11 standard or wireless sensor networks), is the problem of energy saving. In practice, a mechanism for temporary switching off the radio transmitting/receiving, e.g. at a time at which the stream of packets directed to this node becomes less intensive or when the accumulating buffer is empty, is being implemented. After such a time of unavailability, to start the processing normally, a node needs some period of time (called a setup time, usually random) to achieve full readiness to work. During the setup time the service process is blocked and the arriving packets accumulate in the buffer queue. Such a mechanism can be observed, e.g. in manufacturing systems or in the GSM standard transmission in which the node is being switched on just before sending the identification frame by a BS (=Base Station).

In [10] a steady-state threshold strategy for jobs behavior in a model with setup times is obtained. One can find the study of a Markovian system with server setups preceding the first processing in each busy period in [6]. Different results on queueing models with setup times applied in the analysis of WSNs' operation are derived, e.g. in [11] and [12]. In [12] the model of a sleep/wakeup protocol in the IEEE 802.15.4 can be found (see also [7], [9] and [13] for some other results concerning the energy saving problem in WSNs). The IMS session re-setup delay in WiMAX/LTE heterogeneous networks is modelled by an appropriate queueing system in [1]. The $M/G/1$-type queue with vacation policy and server setup times is used in [8] for modelling the BS sleeping mode in cellular networks. In [2] a model of data center with servers leaving their idle periods "via" setup times is investigated.

The non-stationary study of the queue-size distribution in the $M/G/1$-type queueing system with random batch arrivals, N-policy and server setup times can be found in [3] (see also [4], where departure process is investigated).

In the paper we investigate the finite-buffer $M/G/1$-type queueing system in which the first service beginning each new busy period is preceded by a generally-distributed setup time, during which the processing is still blocked and the server acquires full readiness for the service process. Applying the approach based on the idea of embedded Markov chain and the total probability law, a system of integral equations for the transient queue-size distribution, conditioned by the initial buffer state, is built. The solution of the corresponding system written for Laplace transforms is obtained by using the linear algebraic approach and given in a closed form, utilizing a certain functional sequence defined recursively.

II. Model Description and Auxiliary Results

In the article we deal with a single-server finite-buffer queueing model, in which packets occur according to a Poisson process with intensity λ and are being processed individually with a general-type CDF (=cumulative distribution function) $F(\cdot)$ of the service time, according to the FIFO discipline. The total number of packets present in the system simultaneously is bounded by a non-random value K, i.e. we have $K-1$ places in the buffer queue and one place "in processing". As it is usually assumed, packets arriving during the buffer overflow period, i.e. when the server is busy with processing and the buffer is saturated, are being lost. In general, it is allowed for

978-83-60810-90-3/$25.00

the buffer to contain a number of packets being waiting before the opening of the system at time $t = 0$. Each busy period, which starts together with the first arrival after the idle time, is preceded by a random setup time with a CDF $G(\cdot)$ of a general type. The setup time is a period during which the processing is blocked and the server "acquires" full readiness for processing. It is assumed that the system being empty before the start also initializes a setup time at the first arrival epoch. We assume, moreover, that all interarrival, processing and setup times in the evolution of the system are independent.

Let us denote by $X(t)$ the number of packets present in the system at time t, including a packet being processed at this time (if any). Define the transient distribution function of $X(t)$, conditioned by the initial level of buffer saturation, in the following way:

$$Q_n(t,m) \stackrel{def}{=} \mathbf{P}\{X(t) = m \,|\, X(0) = n\}dt, \tag{1}$$

where $t > 0$, $0 \le n \le K$ and $m \ge 0$. We are interested in finding the closed-form representation for the LT (=Laplace transform) of $Q_n(t,m)$, i.e. for the functional

$$\widetilde{q}_n(s,m) \stackrel{def}{=} \int_0^\infty e^{-st} Q_n(t,m)dt, \quad \mathrm{Re}(s) > 0. \tag{2}$$

In the analytical approach we use the following result from linear algebra which can be found in [5]:

Lemma 1. *Let (α_k), $k \ge 0$, and (ϕ_k), $k \ge 1$, with the assumption $\alpha_0 \ne 0$, be two given number sequences. Each solution of the following system of linear equations:*

$$\sum_{k=-1}^{n} \alpha_{k+1} x_{n-k} - x_n = \phi_n, \quad n \ge 0, \tag{3}$$

can be written in the form

$$x_n = CR_{n+1} + \sum_{k=0}^{n} R_{n-k}\phi_k, \quad n \ge 0, \tag{4}$$

where C is a constant independent on n, and (R_k) is the sequence (which is called in [5] a potential) associated with the sequence (α_k) as in the following relationships:

$$R_0 = 0, \quad R_1 = \alpha_0^{-1},$$
$$R_{k+1} = \alpha_0^{-1}\big(R_k - \sum_{i=0}^{k} \alpha_{i+1} R_{k-i}\big), \quad k \ge 1. \tag{5}$$

We end this section with stating some additional notations. We use the abbreviation $\overline{G}(x) \stackrel{def}{=} 1 - G(x)$ and the nomenclature $I\{\mathbb{A}\}$ for the indicator of the random event $\mathbb{A}$. Besides, let us denote by $f(\cdot)$ and $g(\cdot)$ the Laplace-Stieltjes transforms of CDFs $F(\cdot)$ and $G(\cdot)$, respecively.

III. System of Equations for Conditional Queue-Size Distribution

Let us start with the case of the system being empty at time $t = 0$. Thus, the evolution of the system begins with an idle period during which the service station "waits" for packets. Simultaneously with the arrival occurrence a setup time begins. Let us note that we can distinguish three mutually excluding situations (random events):

(1) the first packet arrives before time t and the setup time also ends before t (we denote this event by $A_1(t)$);
(2) the first packet enters before t but the setup time completes after t ($A_2(t)$);
(3) the first packet arrives after time t ($A_3(t)$).

Introduce the following notation:

$$Q_0^{(i)}(t,m) = \mathbf{P}\{\big(X(t) = m\big) \cap A_i(t) \,|\, X(0) = 0\}, \tag{6}$$

where $t > 0$, $m \ge 0$ and $i = 1, 2, 3$. It is obvious that the formula of total probability leads to the following relationship:

$$Q_0(t,m) = \mathbf{P}\{X(t) = m \,|\, X(0) = 0\} = \sum_{i=1}^{3} Q_0^{(i)}(t,m) \tag{7}$$

and, moreover,

$$\widetilde{q}_0(s,z) = \sum_{i=1}^{3} \int_0^\infty e^{-st} Q_0^{(i)}(t,m)dt. \tag{8}$$

Considering the random event $A_1(t)$, we get the following equation:

$$Q_0^{(1)}(t,m) = \int_{x=0}^{t} \lambda e^{-\lambda x}dx$$
$$\times \int_{y=0}^{t-x} \Bigg[\sum_{i=0}^{K-2} \frac{(\lambda y)^i}{i!} e^{-\lambda y} Q_{i+1}(t-x-y,m)$$
$$+ Q_K(t-x-y,m) \sum_{i=K-1}^{\infty} \frac{(\lambda y)^i}{i!} e^{-\lambda y}\Bigg] dG(y). \tag{9}$$

Let us note that the first summand on the right side of (9) corresponds to the situation in which there is at least one free place in the buffer at the completion epoch of the setup time. The second summand presents the case in that the buffer becomes saturated during the setup time.

Similarly, for $A_2(t)$, we obtain

$$Q_0^{(2)}(t,m) = \int_0^t \lambda e^{-\lambda x}\overline{G}(t-x)$$
$$\times \Bigg\{ I\{1 \le m \le K-1\} \frac{\big[\lambda(t-x)\big]^{m-1}}{(m-1)!} e^{-\lambda(t-x)}$$
$$+ I\{m = K\} \sum_{i=K-1}^{\infty} \frac{\big[\lambda(t-x)\big]^i}{i!} e^{-\lambda(t-x)}\Bigg\} dx. \tag{10}$$

Let us comment (10) briefly. The first summand under the integral on the right side of (10) relates to the case when the number of packets, measured at time t, is less than the maximal value K. The second one concerns the situation $m = K$, so during the time $(t-x)$ (the time between the first arrival epoch and t) at least $K-1$ packets must occur. However, $K-1$ packets only are physically buffered. The remaining ones will be lost due to the buffer saturation.

Finally we have, obviously,

$$Q_0^{(3)}(t,m) = I\{m = 0\}e^{-\lambda t}. \tag{11}$$

From (9)–(11), referring to (7), we get

$$Q_0(t,m) = \int_{x=0}^{t} \lambda e^{-\lambda x} dx \times \int_{y=0}^{t-x} \Bigg[\sum_{i=0}^{K-2} \frac{(\lambda y)^i}{i!} e^{-\lambda y} Q_{i+1}(t-x-y,m) + Q_K(t-x-y,m) \sum_{i=K-1}^{\infty} \frac{(\lambda y)^i}{i!} e^{-\lambda y} \Bigg] dG(y) + \int_0^t \lambda e^{-\lambda x} \overline{G}(t-x) \Bigg\{ I\{1 \le m \le K-1\} \times \frac{\big[\lambda(t-x)\big]^{m-1}}{(m-1)!} e^{-\lambda(t-x)} + I\{m=K\} \sum_{i=K-1}^{\infty} \frac{\big[\lambda(t-x)\big]^i}{i!} \times e^{-\lambda(t-x)} \Bigg\} dx + I\{m=0\} e^{-\lambda t}. \tag{12}$$

Investigate now the case of the system being non-empty at the start moment (i.e. $1 \le n \le K$). Due to the fact that successive departure epochs are Markov (renewal) moments during the operation of the $M/G/1$-type system, then, by virtue of the continuous version of the total probability law, applied with respect to the first departure moment after $t = 0$, we obtain the following system of integral equations for $1 \le n \le K$:

$$Q_n(t,m) = \int_0^t \Bigg[\sum_{i=0}^{K-n-1} \frac{(\lambda x)^i}{i!} e^{-\lambda x} Q_{n+i-1}(t-x,m) + Q_{K-1}(t-x,m) \sum_{i=K-n}^{\infty} \frac{(\lambda x)^i}{i!} e^{-\lambda x} \Bigg] dF(x) + \overline{F}(t) \Bigg[I\{n \le m \le K-1\} \frac{(\lambda t)^{m-n}}{(m-n)!} e^{-\lambda t} + I\{m=K\} \sum_{i=K-n}^{\infty} \frac{(\lambda t)^i}{i!} e^{-\lambda t} \Bigg], \tag{13}$$

It is easy to note that the first summand under the integral on the right side of (13) describes the case in which the buffer does not become saturated before the first departure time $0 < x < t$, while the second one relates to the opposite situation. In the last summand the first service completes after t.

Observe that the following identity is true (compare (9)):

$$\int_{t=0}^{\infty} e^{-st} dt \int_{x=0}^{t} \lambda e^{-\lambda x} dx \int_{y=0}^{t-x} \frac{(\lambda y)^i}{i!} e^{-\lambda y} \times Q_j(t-x-y,m) dG(y) = a_i(s) \widetilde{q}_j(s,m), \tag{14}$$

where

$$a_i(s) \stackrel{def}{=} \frac{\lambda}{\lambda+s} \int_0^{\infty} \frac{(\lambda y)^i}{i!} e^{-(\lambda+s)y} dG(y). \tag{15}$$

Similarly, defining (see (10))

$$\overline{a}_i(s) \stackrel{def}{=} \frac{\lambda}{\lambda+s} \int_0^{\infty} \frac{(\lambda u)^i}{i!} e^{-(\lambda+s)u} \overline{G}(u) du, \tag{16}$$

and taking into consideration (14)–(15), we rewrite the equation (12) in the following form:

$$\widetilde{q}_0(s,m) = \sum_{i=0}^{K-2} a_i(s) \widetilde{q}_{i+1}(s,m) + \widetilde{q}_K(s,m) \sum_{i=K-1}^{\infty} a_i(s) + \beta(s,m), \tag{17}$$

where

$$\beta(s,m) \stackrel{def}{=} I\{1 \le m \le K-1\} \overline{a}_{m-1}(s) + I\{m=K\} \sum_{i=K-1}^{\infty} \overline{a}_i(s) + I\{m=0\} \frac{1}{\lambda+s}. \tag{18}$$

Similarly, denoting

$$\alpha_i(s) \stackrel{def}{=} \int_0^{\infty} e^{-(\lambda+s)x} \frac{(\lambda x)^i}{i!} dF(x); \tag{19}$$

$$\gamma_n(s,m) \stackrel{def}{=} \int_0^{\infty} e^{-(\lambda+s)t} \overline{F}(t) \Big[I\{n \le m \le K-1\} \times \frac{(\lambda t)^{m-n}}{(m-n)!} + I\{m=K\} \sum_{i=K-n}^{\infty} \frac{(\lambda t)^i}{i!} \Big] dt, \tag{20}$$

where $\mathrm{Re}(s) > 0$, we transform (13) as follows:

$$\widetilde{q}_n(s,m) = \sum_{i=0}^{K-n-1} \alpha_i(s) \widetilde{q}_{n+i-1}(s,m) + \widetilde{q}_{K-1}(s,m) \sum_{i=K-n}^{\infty} \alpha_i(s) + \gamma_n(s,m). \tag{21}$$

Let us apply to (17) and (21) the following substitution:

$$\widetilde{u}_n(s,m) \stackrel{def}{=} \widetilde{q}_{K-n}(s,m), \quad 0 \le n \le K. \tag{22}$$

After this operation, we get from (21) the following system:

$$\sum_{i=-1}^{n} \alpha_{i+1}(s) \widetilde{u}_{n-i}(s,m) - \widetilde{u}_n(s,m) = \phi_n(s,m), \tag{23}$$

where $0 \le n \le K-1$, and

$$\phi_n(s,m) \stackrel{def}{=} \alpha_{n+1}(s) \widetilde{u}_0(s,m) - \widetilde{u}_1(s,m) \sum_{i=n+1}^{\infty} \alpha_i(s) - \gamma_{K-n}(s,m). \tag{24}$$

In the same manner, inserting (22) into (17), we obtain

$$\widetilde{u}_K(s,m) = \sum_{i=0}^{K-2} a_i(s) \widetilde{u}_{K-i-1}(s,m) + \widetilde{u}_0(s,m) \sum_{i=K-1}^{\infty} a_i(s) + \beta(s,m). \tag{25}$$

IV. Main Results for Transforms

Let us note that (23) has the same form as (3), but with coefficients $\alpha_i(\cdot)$ and $\phi_i(\cdot,\cdot)$, $i \geq 0$, depending on s and (s,m), respectively. So, there is possible to solve (23) by utilizing the result (4). Moreover, due to the fact that the number of equations in (23) is finite comparing to (3), the representation for $C = C(s,m)$ can be found explicitly, considering the last equation (25) as a special-type boundary condition. According to (4), we obtain for $n \geq 0$

$$\widetilde{u}_n(s,m) = C(s,m)R_{n+1}(s) + \sum_{i=0}^{n} R_{n-i}(s)\phi_i(s,m), \quad (26)$$

where the sequence $\big(R_k(s)\big)$ is defined in (5) (with $\alpha_k(s)$ instead of α_k). Substituting $n=0$ into (26), we get

$$\widetilde{u}_0(s,m) = C(s,m)R_1(s). \quad (27)$$

Taking $n=1$ in (26) and referring to (24) and (27), we obtain

$$\widetilde{u}_1(s,m) = C(s,m)R_2(s) + R_1(s)\Big(\alpha_1(s)R_1(s)C(s,m) - \widetilde{u}_1(s,m)\sum_{i=1}^{\infty}\alpha_i(s) - \gamma_K(s,m)\Big) \quad (28)$$

that leads to

$$\widetilde{u}_1(s,m) = \eta(s)\Big[C(s,m)\big(R_2(s) + \alpha_1(s)R_1^2(s)\big) - R_1(s)\gamma_K(s,m)\Big], \quad (29)$$

where $\eta(s) \stackrel{def}{=} \frac{f(\lambda+s)}{f(s)}$. Since using (27) and (29) the functionals $\phi_i(s,m)$, $i \geq 0$, can be found, if only the formula for $C(s,m)$ is known, the key problem is in finding the explicit representation for $C(s,m)$. By using the representations (24) and (26), we can present (25) in the following form:

$$\widetilde{u}_K(s,m) = \Delta_1(s)C(s,m) + \kappa_1(s,m), \quad (30)$$

where we denote

$$\Delta_1(s) \stackrel{def}{=} \sum_{i=1}^{K-1} a_{K-i-1}(s)\Big[R_{i+1}(s) + \sum_{j=0}^{i} R_{i-j}(s) \times \Big(R_1(s)\alpha_{j+1}(s) - \eta(s)\big(R_2(s) + \alpha_1(s)R_1^2(s)\big)\sum_{r=j+1}^{\infty}\alpha_r(s)\Big] + R_1(s)\sum_{i=K-1}^{\infty} a_i(s) \quad (31)$$

and

$$\kappa_1(s,m) \stackrel{def}{=} \sum_{i=1}^{K-1} a_{K-i-1}\sum_{j=0}^{i} R_{i-j}(s)\Big[R_1(s)\gamma_K(s,m)\eta(s) \sum_{r=j+1}^{\infty}\alpha_r(s) - \gamma_{K-j}(s,m)\Big] + \beta(s,m). \quad (32)$$

Similarly, let us take $n=K$ in (26) and apply the formulae (24), (27) and (29). In consequence we obtain

$$\widetilde{u}_K(s,m) = \Delta_2(s)C(s,m) + \kappa_2(s,m), \quad (33)$$

where

$$\Delta_2(s) \stackrel{def}{=} R_{K+1}(s) + \sum_{i=0}^{K} R_{K-i}(s)\Big[\alpha_{i+1}(s)R_1(s) - \eta(s)\big(R_2(s) + \alpha_1(s)R_1^2(s)\big)\sum_{j=i+1}^{\infty}\alpha_j(s)\Big] \quad (34)$$

and

$$\kappa_2(s,m) \stackrel{def}{=} \sum_{i=0}^{K} R_{K-i}(s)\Big(\eta(s)R_1(s)\gamma_K(s,m) \times \sum_{j=i+1}^{\infty}\alpha_j(s) - \gamma_{K-i}(s,m)\Big). \quad (35)$$

Comparing the right sides of (30) and (33), we express $C(s,m)$ as follows:

$$C(s,m) = \frac{\kappa_2(s,m) - \kappa_1(s,m)}{\Delta_1(s) - \Delta_2(s)}. \quad (36)$$

Now the formulae (24), (26), (27), (29) and (36) lead to the following main theorem:

Theorem 1. *The representation for the LT of the conditional transient queue-size distribution in the $M/G/1/K$-type model with generally distributed setup times is following:*

$$\widetilde{q}_n(s,m) = \int_0^{\infty} e^{-st}\mathbf{P}\{X(t) = m \,|\, X(0) = n\}dt = \frac{\kappa_2(s,m) - \kappa_1(s,m)}{\Delta_1(s) - \Delta_2(s)}\Bigg\{R_{K-n+1}(s) + \sum_{i=0}^{K-n} R_{K-n-i}(s) \times \Big[\alpha_{i+1}(s)R_1(s) - \eta(s)\Big(R_2(s) + \alpha_1(s)R_1^2(s)\Big) \times \sum_{j=i+1}^{\infty}\alpha_j(s)\Big]\Bigg\} + \sum_{i=0}^{K-n} R_{K-n-i}(s)\Big(\eta(s)R_1(s)\gamma_K(s,m) \times \sum_{j=i+1}^{\infty}\alpha_j(s) - \gamma_{K-i}(s,m)\Big), \quad (37)$$

where the formulae for $\alpha_i(s)$, $\gamma_i(s,m)$, $R_i(s)$, $\Delta_1(s)$, $\kappa_1(s,m)$, $\Delta_2(s)$ and $\kappa_2(s,m)$ are given in (19), (20), (5), (31), (32), (34) and (35), respectively.

Remark IV.1. *Let us note that from the formula (37) the stationary queue-size distribution π_m, $m = 0,...,K$, can be found by using the Tauberian theorem, namely for any $n \in \{0,...,K\}$*

$$\pi_m = \lim_{t\to\infty}\mathbf{P}\{X(t) = m\} = \lim_{s\downarrow 0} s\cdot\widetilde{q}_n(s,m). \quad (38)$$

V. Numerical examples

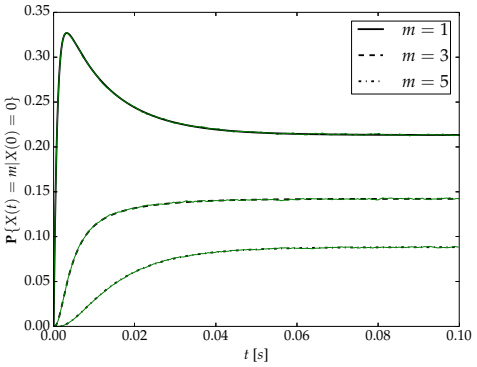
Let us take into consideration a node of the wireless network with buffer of size 6 packets, with the arrival stream of packets of average sizes 200 B, entering with intensity 600 Kb/s. Adjusting the Poisson arrival process, we have the rate $\lambda = 375$ packets per second. Assume that packets are being transmitted with speed 720 Kb/s, where the processing times have 2-Erlang distribution, that gives the intensity 450 packets

per second and the value $\mu = 900$ of the parameter of this distribution. Hence, the load of the node is at the high level ($\rho = 0.833$). Moreover, we consider the mechanism of exponentially distributed setup with three different means: 0 (without setup - "pure" system), 5 and 50 [ms].

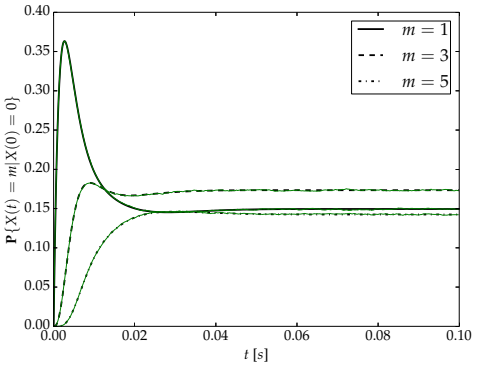
The evolution of the transient probabilities $\mathbf{P}\{X(t) = m \,|\, X(0) = 0\}$ for $m = 1, 3, 5$ are presented in Fig. 1. Evidently, due to different durations of setup times, the transient evolutions and stationary values of the proper probabilities are different in cases (a)–(c). As one can observe, a short setup time with mean 5 [ms] allows the system to stabilize faster than in the absence of such a period (compare cases (a) and (b)). From the other side, if the setup time is relatively long in comparison to the arrival/service rates (case (c)), the time of stabilization elongates in comparison to the system without setup time. It is worth mentioning that all analytical results are confirmed by process-based discrete-event simulations (DES).

VI. Summary

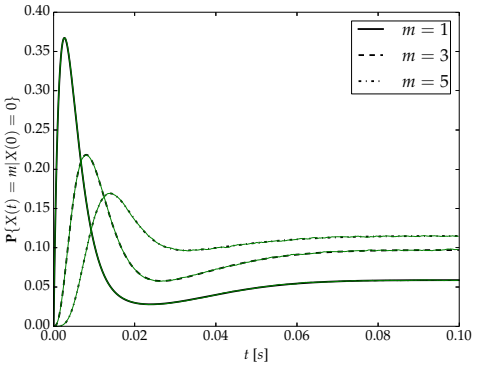
In the article a finite-buffer queueing model with Poisson arrivals and generally distributed processing times and server setup times is considered. The closed-form representation for the LT of the transient queue-size distribution conditioned by the initial buffer state is found, from which the stationary distribution can be obtained directly by using the Tauberian theorem. Numerical examples are attached as well.



(a) no setup time



(b) setup time mean 5 ms



(c) setup time mean 50 ms

Fig. 1. Subfigures (a), (b) i (c) present the probabilities $\mathbf{P}\{X(t) = m|X(0) = 0\}$ for $m = 1$ (solid line), $m = 3$ (dashed line) and $m = 5$ (dot dashed line), for the case of no setup time (subfigure (a)), a setup time with mean 5 [ms] (subfigure (b)) and with mean 50 [ms] (subfigure (c)). Bold black lines and thin green lines correspond with analytical and DES results, respectively

References

[1] E. P. Edward, "A Novel Seamless Handover Scheme for WiMAX/LTE Heterogeneous Networks," *Arab. J. Sci. Eng.,* vol. 41, iss. 3, 2016, pp. 1129–1143.

[2] J. N. Hu and P. D. Tuan, "Power Consumption Analysis for data Centers with Independent Setup Times and Threshold Controls," *AIP Conference Proceedings,* vol. 1648, 2015.

[3] W. M. Kempa, "On Transient Queue-Size Distribution in the Batch Arrival System with the N-policy and Setup Times," *Math. Commun.,* vol. 17, iss. 1, 2012, 285–302.

[4] W. M. Kempa and D. Kurzyk, "Transient Departure Process in $M/G/1/K$-type Queue with Threshold Server's Waking Up," *Proc. of the 23rd International Conference on Software, Telecommunications and Computer Networks (SoftCOM 2015),*Split - Bol (Island of Brac), Croatia, 2015, pp. 32–36, DOI: 10.1109/SOFTCOM.2015.7314127.

[5] V. S. Korolyuk, *Boundary-value problems for compound Poisson processes,* Naukova Dumka, Kiev, 1975.

[6] Q. Ma, "Analysis of a Clearing Queueing System with Setup Times," *RAIRO-Oper. Res.,* vol. 49, iss. 1, 2015, pp. 67–76, DOI: http://dx.doi.org/10.1051/ro/2014035.

[7] J. Miček and J. Kapitulík, "WSN Sensor Node for Protected Area Monitoring," *Proc. of FedCSIS 2012,* 803–807.

[8] Z. S. Niu and X. Y. Guo and S. Zhou and P. R. Kumar, "Characterizing Energy-Delay Tradeoff in Hyper-Cellular Networks With Base Station Sleeping Control," *IEEE J. Sel. Area Comm.,* vol. 33, iss. 4, 2015, 641–650, DOI: 10.1109/JSAC.2015.2393494.

[9] M. Salayma and A. Al-Dubai and I. Romdhani and M. B. Yassein, "Battery Aware Beacon Enabled IEEE802.15.4: An Adaptive and Cross-Layer Approach," *Proc. of FedCSIS 2015,* 1267–1272, DOI: 10.15439/2015F118.

[10] W. Sun and P. F. Guo and N. S. Tian, "Equilibrium Threshold Strategies in Observable Queueing Systems with Setup/Closedown Times," *Cent. Europ. J. Oper. Re.,* vol. 18, iss. 3, 2010, pp. 241–268, DOI: 10.1007/s10100-009-0104-4.

[11] Q. T. Sun and S. F. Jin and C. Chen, "Energy Analysis of Sensor Nodes in WSN Based on Discrete-Time Queueing Model with a Setup," *Proc. of 22nd Chinese Control and Decision Conference,* Xuzhou, China, vols. 1–5, 2010, pp. 4114-4118, DOI: 10.1109/CCDC.2010.5498425.

[12] W. Y. Yue and Q. T. Sun and S. F. Jin, "Performance Analysis of Sensor Nodes in a WSN With Sleep/Wakeup Protocol," *Lect. Notes Oper. Res.,* vol. 12, 2010, pp. 370–377.

[13] M. Zieliński and F. Mieyeville and D. Navarro and O. Bareille, "A Low Power Wireless Sensor Node with Vibration Sensing and Energy Harvesting Capability," *Proc. of FedCSIS 2014,* 1065–1071.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1021–1030
DOI: 10.15439/2016F219
ACSIS, Vol. 8. ISSN 2300-5963

A new authentication management model oriented on user's experience

Mariusz Sepczuk , Zbigniew Kotulski,
Institute of Telecommunications
Warsaw University of Technology
Warsaw, Poland
Email: {msepczuk, zkotulsk} @tele.pw.edu.pl

***Abstract*— Authenticating users connecting to online services, social networks or m-banking became an indispensable element of our everyday life. Reliable authentication is a foundation of security of Internet services but, on the other hand, also a source of users' frustration due to possible account blocking in case of three fails. In this paper we propose a model of authentication service management which helps in keeping a balance between the authentication security level and positive users' perception of this procedure. The proposed procedure allows a user more than three attempts of authentication by switching after two failures to a more secure authentication protocol keeping a balance between QoP and QoE measures. Finally, the procedure determines an optimal path of authentication using a decision tree algorithm.**

***Index Terms*— authentication, Quality of Experience, Quality of Protection.**

I. INTRODUCTION

Diversity of Internet services at the present time grows faster and faster. In particular, the variety of manners in which the services are provided, from a wired environment (e.g., LAN) to a wireless environment (e.g. WiFi, mobile environment), is observed. Many users become more and more demanding about services' usability. Thus, any services, especially newly designed, should be developed taking into account users' satisfaction factor.

One of the main issues in all Internet services is security protection. Nowadays, there are few user- friendly and at the same time secure services. It is well known that for most services the high level of protection makes their usability is declined. So, it is important to find a balance between security and usability of a service. Of course, that idea depends on kind of a protection mechanism which is considered. Examples of such security mechanisms are authentication solutions. The authentication is an act of reliable entity identification. Within this process two problems can be considered: a choice of the specific authentication solution and its influence on user's behavior. The choice of the proper identification mechanism isnot a simple problem, because many factors can have a significant impact on it. Even if such a mechanism is selected, in most cases it is not considered how a person feels using it. Therefore, an appropriate authentication solution should provide both an adequate security level and sufficient users' satisfaction.

In this paper we propose a service model which can be used to proper management of the authentication mechanisms based on users' satisfaction.

The rest of the paper is organized as follows: Section 2 presents a connection between QoP and QoE measures characterizing Internet services. Section 3 briefly discuses an impact of security context and contextual data on security management while Section 4 presents basic known results on contextual security and user-friendly authentication mechanisms. .Section 5 contains main theoretical result of the paper which is an authentication management model oriented on user's experience. Finally, Section 6 presents results of a simulation which confirms correctness of a created model and Section 7 concludes the paper and outlines the future work.

II. RELATION BETWEEN QOS, QOE AND QOP

In Internet services, to measure Quality of Service (QoS) [1] many parameters like jitter, network latency, throughput, etc. are used. Based on their value a service and a network parameters should be correctly modify to ensure the best quality. However, not always changing a QoS parameter is enough to provide a good quality service. Sometimes to provide high quality of a service not all parameters should have the best values. In most cases it is expensive to set the best values of QoS parameters. Thus, investigations concerning users' experience were conducted. As a result of the research a Quality of Experience (QoE) factor was designed [2, 3]. This implies that QoS parameters should be set based on a users' QoE value [4, 5].

In the area of security the Quality of Protection (QoP) measure is a counterpart of QoS [6, 7]. The term defines a minimum protection level that should be provided to a secure Internet service. For example, it is obvious that different level of protection ensures an authentication mechanism which used a hash function SHA-1 than those with a hash function SHA-2/256 [8]. So, it is natural to measure a level of protection which is required. But, as it is for QoS, not always a security mechanism applied meets users' requirements. Sometimes the mechanism is too

978-83-60810-90-3/$25.00

difficult to use (e.g., a multi-factor authentication can be a barrier for elder people), sometimes it is too annoying (e.g., a continuous request for fingerprinting due to a device read/scan problems).

To summarize above considerations, the relationship between QoS, QoP and QoE can be presented in a form of the graph (see Fig 1). The QoS parameter has an impact on security services (a security level) and at the same time on users' satisfaction. Once again, a proper security mechanism should be provided with respect to a user's expectations.

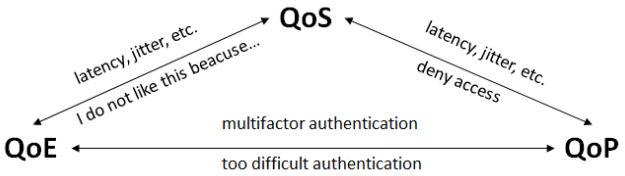


Figure 1. Relationship between QoS, QoP and QoE

An open problem is to answer a question which mechanism should be provided in particular conditions. For this issue the solution can be the idea of context-based systems in which a state of the system is dynamically changed with respect of environmental conditions.

III. Security context

The term "context" concerns all data which can be used to adapt state of a system to particular conditions. They come from many sources [9].There are many descriptions about context. In [10] Schilit et. al. show context-aware systems as systems able to adapt to dynamically and continuously changing GPS coordinates , type of devices, people relations and time. The author describes three features of context data: where you are?, with whom you are?, what is your neighborhood? Furthermore, he emphases that contextual information is not only a localization, but also other useful information. Recently this definition had been used by many authors. In general they claim that contextual data are data which are answers on a question that starts with: Who?, Where?, What?, When?. Beside context-aware system, description of the context-aware application can be found in [11]. More universal definition was proposed by Wrona and Gomez [12]. According to them the contextual data are information which can describe a state of an entity. This definition is better than the previous one because it includes all data which can be contextual information.

Some contextual data have common features. For example, day, part of a year and day of birth apply to determine the time while GPS coordinates, the UK and the Earth describe a localization. For this reason it is naturally to divide contextual data into categories. Two context categories were are already defined: time and position. Beside these, many more exist like: access device, operating systems, environment, neighborhood, etc.

All contextual information can be divided in three classification groups: storage of context, retrieval of context and its dynamism. First group includes all aspects of store data (in database, in Hidden Markov Model, in file, etc.). Second group describe different aspects of data retrieval like its history, presentation, the way in which was gathered, etc. The last third group consists data connected with a specific environment- some environment could be dynamically changed over time and some could be more static- every information changed very rarely or not changed at all.

As it was shown in the literature (see e.g., [13, 14, 15]) contextual information can be used to improve Internet systems work. Using contextual data, QoP and QoE measures combination cause that a model of providing the best authentication mechanism adequate for specific requirements can be created.

IV. User authentication and Quality of Experience: related works

As mentioned earlier, not many works apply to problem of providing an appropriate authentication mechanism based on user's context and respecting the user's QoE. Security solutions which include contextual data aspects are more common. Several approaches for the context security have been described in [16, 17, 18, 19]. The authors in their papers present an idea how contextual information can be used in an authentication mechanism, but they do not consider user's Quality of Experience. Another example of using contextual data in security is access control approach. Based on sensitivity or importance of data (e.g. patient health history, personal data, etc.), proper access to them should be provided to avoid information leakage. This idea is shown in [20, 21, 22, 23]. However, the aspects of security should be considered as well as user's satisfaction. In most cases it is difficult to obtain an answer to the question which security features have impact on user's experience. In [24] the authors present the results of a survey of over 300 users to determine their understanding of the security feature in selected applications. The experiment includes some areas of difficulty with many security features showing usability challenges for users. Similar considerations includes the paper [25]. Based on conclusions from [24, 25] was created a few papers which apply to experiments of balancing QoE and QoP. In [26, 27] is described provisioning of QoE and QoP in Mobile Networks and Wireless Networks. Authors focus on ensuring an appropriate level of QoE and QoE of cryptographic algorithms used in a mobile environment. The paper [28] shows long – term QoP in mobile networks with QoE aspects, too. More experimental results can be found in [29]. The paper contains research about impact of an authentications mechanism on users perceive logins. Computed exponential QoE-QoP relationship can be served to assessing used identification mechanism in the domain of user acceptability. The extension of these research and more detailed description can be found in [30, 31]. Based on [29] was created an experiment shown in [32]. The author tries to find QoE – QoP dependence in popular SaaS cloud products. A similar approach, but with mobile devices, is shown in [33], where security barriers survey was described.

Other, more holistic idea of connection security and user experience can be found in [34]. The paper shows a framework of

criteria for the evaluation of authentication schemes in IMS, focus on security, user-friendliness and simplicity. Very interesting description contains [35] where authors explain how contextual information can be used to provide secure and user oriented mechanisms. The paper [36] is an example of using the framework from the paper [35] for constructing an adaptable authentication protocol.

V. Authentication Model Using User's Context and QoE

A described solution is an example of a model which helps in authentication mechanisms management. The model was created to provide a correct authentication mechanism based on user's context (e.g. position, time, neighborhood, etc.) and on the knowledge about user's experience in using a particular authentication service. One of the most important elements of the model is a table which contains a list of authentication mechanisms (A_1,A_2,…,A_n) and their QoP (QoP_1,QoP_2,…,QoP_n) and QoE (QoE_1,QoE_2,…,QoE_n) measures (see TABLE I). The values of such parameters (measures) are usually based on experts' knowledge and data obtained from experiments. The table is used to select the best authentication mechanism from many possible choices.

TABLE I
ASSIGNMENT OF AUTHENTICATION MECHANISMS AND THEIR QOP AND QOE VALUES

LP	Kind of mechanism	QoP	QoE
1	A_1	QoP_1	QoE_1
2	A_2	QoP_2	QoE_2
3	A_3	QoP_3	QoE_3
…	…	…	…
n	A_n	QoP_n	QoE_n

A user who wants to use a service at first must authenticate himself. Based on current user's context the required minimal value of Quality of Protection level is calculated according to formula (1):

$$f(c_1, c_2, c_3, \ldots, c_m) = QoP_c \tag{1}$$

where c_1,c_2,…,c_m are context factors.

Having the boundary QoP_c value it is possible to select some authentication mechanisms for which the QoP value is greater than or equal to QoP_c. The best authentication mechanism can be chosen using a decision tree structure (see Fig. 2). The tree is spanned on j states. Each state represents a situation in which a user tries to authenticate himself with a particular authentication protocol. In every state the user has two trials to identify himself with a selected protocol (of course, twice means that first trial was failed and now is a second trial). In the third trial, when the first and second trials were incorrect, he or she can still use the same protocol or, if it is possible, change an authentication protocol to some more convenient in a specific situation. If he or she decides to change the protocol, the new one is with a higher value of OoP. The required higher value of QoP implicates less probability of a successful attack in three trials of the new protocol than in a case of a single try in the previous one. A user can choose a new authentication protocol option until the cumulative value of QoP in a current state does not achieve the boundary value QoP_c.

The main goal of created model is to choose the best path of the tree. The best path means a scenario where a user uses the last authentication protocol in the state j and he finally authenticates correctly with the highest probability. Moreover, the path should contain the authentication protocols which are relatively simple, secure and at the same time with the high value of the QoE measure. A choice of this path could be made by calculations based on a decision tree algorithm. In the next paragraph we will describe the probability of a successful authentication in every branch of the decision tree.

As it shown in Figure2, in an authentication decision tree there are two types of states: state 1 without changing an authentication protocol (but with three possible trials of the same protocol) and the states from 2 to n where a user changes this protocol (after two trials of the same protocol he or she tries with a new one in the third trial). Furthermore, every trial of an authentication protocol can be successful; the event R_{ij} means that in state i (i=2…n), in its step j (j=1,2,3) an authentication is right or successful , the event F_{ij} means that in state i, in its step j the authentication fails. Moreover, the event B_i was defined as correct authentication in state i after using all options of a path. So, for state 1 the probability of correct authentication according to (2) is equal:

$$\begin{aligned} P(B_1) &= P(R_{11}) + P(R_{12} \mid F_{11})P(F_{11}) + \\ &P(R_{13} \mid F_{11} \cap F_{12})P(F_{11} \cap F_{12}) \end{aligned} \tag{2}$$

The state 2 and next states are different from state 1, thus, for state 2 the probability of correct authentication according to (3) is equal:

$$\begin{aligned} P(B_2) &= P(R_{21} \mid F_{11} \cap F_{12})P(F_{11} \cap F_{12}) + \\ &P(R_{22} \mid F_{11} \cap F_{12} \cap F_{21})P(F_{11} \cap F_{12} \cap F_{21}) + \\ &P(R_{23} \mid F_{11} \cap F_{12} \cap F_{21} \cap F_{22})P(F_{11} \cap F_{12} \cap F_{21} \cap F_{22}) \end{aligned} \tag{3}$$

Analogously, it is possible to calculate the probability of correct authentication formula for each states from 2 to n. A general formula for the states from 2 to n is equal:

$$\begin{aligned} P(B_n) &= P\left(R_{n1} \mid \left(\bigcap_{j=1}^{n-1} F_{j1} \cap F_{j2}\right)\right) P\left(\bigcap_{j=1}^{n-1} F_{j1} \cap F_{j2}\right) + \\ &P\left(R_{n2} \mid \left(\bigcap_{j=1}^{n-1} F_{j1} \cap F_{j2}\right) \cap F_{n1}\right) P\left(\left(\bigcap_{j=1}^{n-1} F_{j1} \cap F_{j2}\right) \cap F_{n1}\right) + \\ &P\left(R_{n2} \mid \left(\bigcap_{j=1}^{n-1} F_{j1} \cap F_{j2}\right) \cap F_{n1} \cap F_{n2}\right) P\left(\left(\bigcap_{j=1}^{n-1} F_{j1} \cap F_{j2}\right) \cap F_{n1} \cap F_{n2}\right) \end{aligned} \tag{4}$$

Finally, a formula for the probability of correct authentication in a decision tree is equal according to (5):

$$P(B_n)=\begin{cases} P(R_{11})+P(R_{12}\mid F_{11})P(F_{11})+P(R_{13}\mid F_{11}\cap F_{12})P(F_{11}\cap F_{12}) & ,for \quad n=1 \\ P\left(R_{n1}\mid\left(\bigcap_{j=1}^{n-1}F_{j1}\cap F_{j2}\right)\right)P\left(\bigcap_{j=1}^{n-1}F_{j1}\cap F_{j2}\right)+P\left(R_{n2}\mid\left(\bigcap_{j=1}^{n-1}F_{j1}\cap F_{j2}\right)\cap F_{n1}\right)\cdot & \\ P\left(\left(\bigcap_{j=1}^{n-1}F_{j1}\cap F_{j2}\right)\cap F_{n1}\right)+P\left(R_{n3}\mid\left(\bigcap_{j=1}^{n-1}F_{j1}\cap F_{j2}\right)\cap F_{n1}\cap F_{n2}\right)P\left(\left(\bigcap_{j=1}^{n-1}F_{j1}\cap F_{j2}\right)\cap F_{n1}\cap F_{n2}\right) & ,for \quad n=2,\ldots,n \end{cases} \tag{5}$$

TABLE II
EVENTS AND THEIR IMPACT ON QOE AND QOP MEASURES

	QoE			QoP	
	Description	Parameter	Value	Description	Parameter
Increase	A user has still possibility of authentication using the same or a new mechanism	α	≈0,15	An authentication mechanism was changed	QoP_j
Increase	A user finally authenticate himself	β	≈0,25		
Decrease	With every user try his or her satisfaction is lower	γ	≈0,1	First or second trial was failed	$\frac{1}{\lvert t_j\rvert}$

To obtain the best authentication path it is necessary to calculate the probability of correct authentication for every branch of the tree in a state number n (we denote it as $P(B_{nm})$ where m means a number of the branch). From the set of received probabilities the maximal value is selected ($\max\{P(B_{n1}), P(B_{n2}), P(B_{n3}),\ldots, P(B_{nm})\}$) and this value indicates a path with authentication mechanisms which should be used to deliver proper levels of protection (QoP) and user's satisfaction (QoE).

Beside the probability of a successful authentication, the level of QoP and QoE measures for every branch of the tree should be calculated. These two values define together a new parameter called Quality of User Security Service (QoUSS). Usually in literature information about parameter QoSS (Quality of Security Services) can be found. This factor describes security based on QoS of an Internet service [37, 38]. The QoUSS measure includes information about both the security level and, what is important, the satisfaction level of a used service; it is defined by a function:

$$f(QoP, QoE) = QoUSS \tag{6}$$

The argument QoE in that formula means final user satisfaction after correct authentication in a last state and QoP means the resultant level of protection in the final state.

Before we define the expressions for calculating QoE and QoP measures suitable in our model, let us describe example cases which can have impact on these two values. The TABLE II includes events which affect increase or decrease of the QoUSS arguments.

We postulate that the values of parameters moderating QoP and QoE included in TABLE II should be small, because they must not impact a resultant value of the measures. They are considered as correction parameters, so we assume $\alpha, \beta, \gamma \in (0,1)$. In TABLE II we proposed some intuitively assumed values of these parameters to reflect users' emotions connected with successes and fails of their authentication. More realistic parameters should be dedicated to specific authentication mechanisms and they must be obtained from gathering experimental data. Moreover, the value QoP_j is a minimal protection level of a new authentication protocol and $|t_j|$ is the number of all possible trials of the authentication protocol in step number j.
Thus, we propose the QoP formula as:

$$QoP_{jFIN} = QoP\left(1-\frac{1}{|t_j|-1}\right). \tag{7}$$

The proposed formula for QoE is more complicated, so it will be briefly described.
Again, like in the case of calculating the probability of a successful authentication, all states can be divided on two QoE types:

- The first state when an user authenticates himself,
- The second and next states when an user authenticates himself.

We propose a general formula of MOS dependency in an exponential form:

$$QoE = A\cdot\exp\{Z\} \tag{8}$$

where A is a constant value allowing tune the model to users' behavior, e.g., $A\in(0{,}01;1)$.

Such a shape of this function is to provide adequate sensitivity of the measure in critical areas of minimal and maximal scorings. The argument Z depends on a state in a decision tree, and the scaling constant A is determined by the MOS scale (which is from 1 to 5). Each authentication protocol has a particular QoE value. For the first failed try a user can be a little confused that he does not authenticate himself (the QoE decreases with γ_1) and at the same time the user feels good that he or she can still try with next attempt (the QoE value increases with α_1). For the second failed try user is more confused (decrease with

γ_2) but still can try authenticate himself (increase with α_2). Finally, a user authenticates himself so his satisfaction increases with the value (β).
In most cases MOS dependency has an exponential distribution, so in the first state final QoE value is equal:

$$QoE_{jFIN=1} = \begin{cases} 1 & ,if\ A \cdot \exp\{Z\} \le 1 \\ A \cdot \exp\{Z\} & ,if\ 1 < A \cdot \exp\{Z\} < 5 \\ 5 & ,if\ A \cdot \exp\{Z\} \ge 5 \end{cases} \quad (9)$$

where $Z = QoE_1 - \gamma_1 - \gamma_2 + \alpha_1 + \alpha_2 + \beta$

For the second state (and each next one) the value of QoE depends on QoE value from the previous state. QoE value on the beginning of a new state, which is connected with the previous is equal:

$$QoE_{j-1FIN} = \begin{cases} 1 & ,if\ A \cdot \exp\{Z\} \le 1 \\ A \cdot \exp\{Z\} & ,if\ 1 < A \cdot \exp\{Z\} < 5 \\ 5 & ,if\ A \cdot \exp\{Z\} \ge 5 \end{cases} \quad (10)$$

where $Z = QoE_1 - \gamma_1 - \gamma_2 + \alpha_1 + \alpha_2 + \alpha_3$

The value of α_3 is reflects the result of changing the authentication protocol. Basically founding connection between two values of QoE and calculation one average value is a difficult issue. Thus, reasonable is to assume the worst case in which choosing value is lesser. In presented case the lesser value is chosen between a value from the previous state and the QoE value for the present authentication protocol (for the present state):

$$QoE_j = \min\{QoE_{j-1FIN}, QoE_{jFIN}\} \quad (11)$$

For such a value of QoE calculations are performed based on formula (8) in case of an authentication. When the user finally do not authenticate correctly, his/her QoE decrees to 0 (but with flow of time this value can grow because the user thinks about this situation and agrees that this mechanism is secure and protects him against crackers).

Finally, for each branch of the tree the following 3-tuple was calculated: **$(P(B_n), QoE_{FINpath}, QoP_{FINpath})$**. Probability of choosing particular path includes QoE and QoP values. But it may be that paths have values like in TABLE II.

TABLE III
EXAMPLE OF RESULTS OF THE DECISION TREE ALGORITHM

Path number	1	2	3	4
QoE	3	3,5	4,5	3
QoP	3	4,5	3,5	4
Probability	0,5	0.9	0,7	0,6

It would seem that the path number 2 is the best one when considering the probability. However it is not so obvious. The path number 3 has a higher value of QoE, but a lesser value of QoP. Due to this fact there is need to use multi-objective optimization to choose the best path.
Let us assume that all results from the decision tree algorithm are in TABLE IV.

TABLE IV
ALL RESULTS FROM THE DECISION TREE ALGORITHM

Path number	1	2	3	4	…	n	**Weight**
QoE	qoe_1	qoe_2	qoe_3	qoe_4	…	qoe_n	$\mathbf{w_1}$
QoP	qop_1	qop_2	qop_3	qop_4	…	qop_n	$\mathbf{w_2}$
Probability	p_1	p_2	p_3	p_4	…	p_n	$\mathbf{w_3}$

To choose which path is the best weight sum method should be used. In general below conditions must met:

Maximize: $f(u) = \sum_{i=1}^{n} w_j \cdot K_i(u)$

Subject to: $u \in U$,
where the weights w_i, i=1,…,n corresponding to objective function satisfy the following conditions:

$$\sum_{i=1}^{n} w_1 = 1, \quad w_i \ge 0, \quad i = 1,\ldots,n,$$

and $K_i(u)$ is the objective function and U is feasible design space.
In general the maximized formula must be satisfied:

$$u_k \succ u_l \Leftrightarrow \sum_{i=1}^{n} w_i \cdot K_i(u_k) > \sum_{i=1}^{n} w_i \cdot K_i(u_l)$$

It means that decision u_k is better than decision u_l when sum of multiplications of weight and objective function of decision u_k is greater than for the decision u_l.

In presented case the function f(u) for each path is presented in TABLE V

TABLE V
VALUE OF FUNCTION F(U) IN MULTI –OBJECTIVE OPTIMIZATION

Path number	1	2	3	4	…	n	**Weight**
QoE	qoe_1	qoe_2	qoe_3	qoe_4	…	qoe_n	$\mathbf{w_1}$
QoP	qop_1	qop_2	qop_3	qop_4	…	qop_n	$\mathbf{w_2}$
Probability	p_1	p_2	p_3	p_4	…	p_n	$\mathbf{w_3}$
f(u)	$qoe_1 \cdot w_1 + qop_1 \cdot w_2 + p_1 \cdot w_3$	$qoe_2 \cdot w_2 + qop_2 \cdot w_2 + p_2 \cdot w_3$	$qoe_3 \cdot w_1 + qop_3 \cdot w_2 + p_3 \cdot w_3$	$qoe_4 \cdot w_1 + qop_4 \cdot w_2 + p_4 \cdot w_3$		$qoe_n \cdot w_1 + qop_n \cdot w_2 + p_n \cdot w_3$	

Of course to perform optimization values of should be normalized. What is also important that calculation are made with assumption that the most important should be path with the highest QoE value than path with QoP value and a finally probability of path. It means that $w_1 > w_2 > w_3$.
In considered example f(u) has the following values (see TABLE VI):

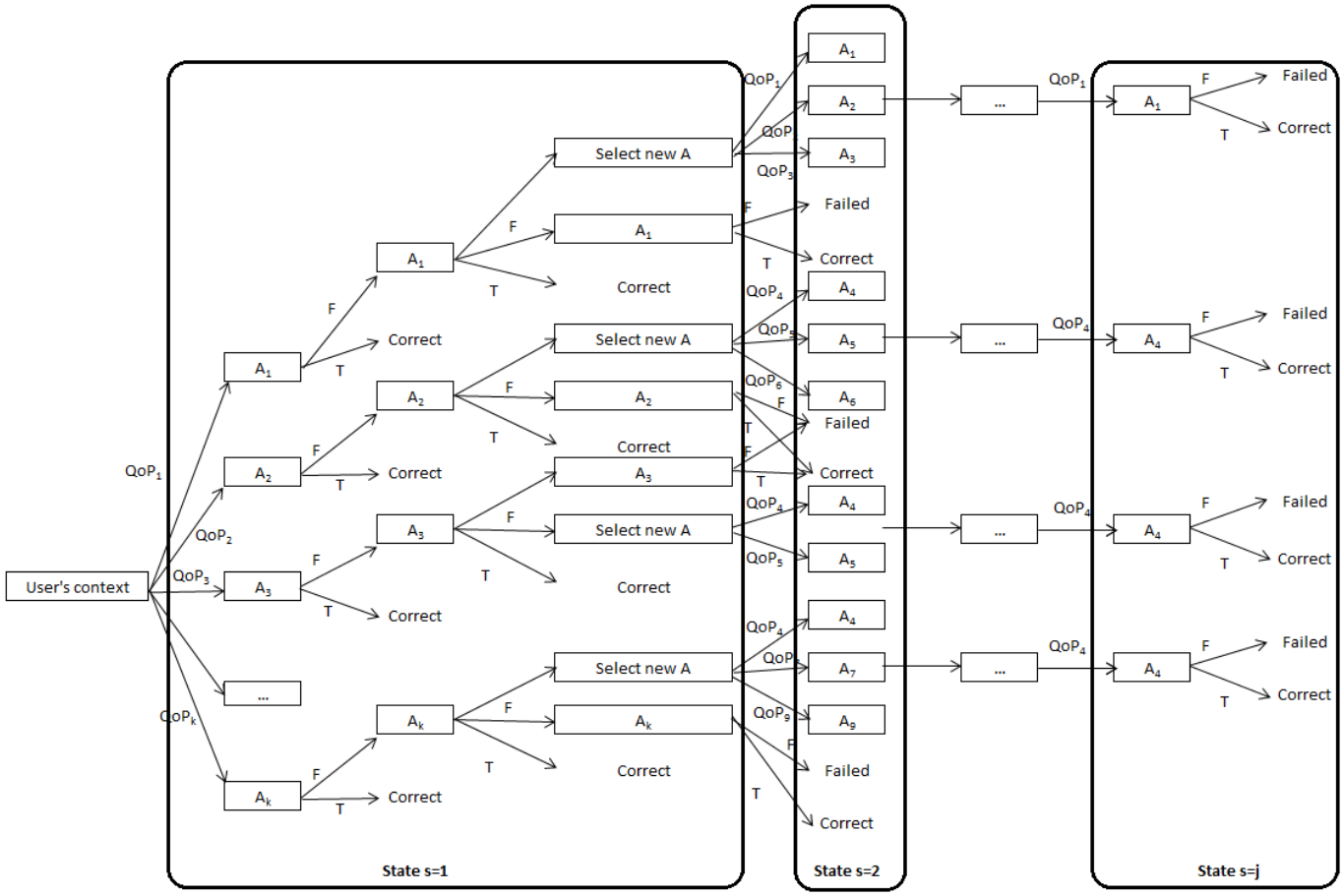


Figure 2. A decision tree for the authentication mechanism

TABLE VI
THE VALUE OF FUNCTION F(U) IN THE CONSIDERED EXAMPLE

Path number	1	**2**	**3**	4	**Weight**
QoE	0,6	**0,7**	0,9	0,6	**0,45**
QoP	0,6	**0,9**	0,7	0,8	**0,35**
Probability	0,5	**0,9**	0,7	0,6	**0,2**
f(u)	0,58	**0,81**	0,79	0,67	

Based on TABLE VI, the best path with authentication mechanism is the second one, but a few more than third path.

VI. Discussion

Now some facts about α, β, γ impact on the formula (7) will be considered. The discussion assumes two scenarios:

- A user correctly authenticates himself in the third step in stage number 1 (Scenario 1),
- A user correctly authenticates himself in the third step in stage number 2 (Scenario 2).

For these scenarios TABLES VIA, VIB, VIC (Scenario 1), VIIA, VIIB, VIIIA, VIIIB, VIIIC (Scenario 2) and corresponding to them charts were created. In Scenario 1 and Scenario 2 we assume the factor A=0,075. Moreover, if the value of QoE_{FIN} after calculations is greater than 5, automatically it is corrected to 5.

TABLE VIA
PARAMETERS FOR SCENARIO 1 – INCREASE OF Γ

QoE_1	α_1	α_2	β	γ_1	γ_2	QoE_{1FIN}
3,5	0,15	0,16	0,25	0,1	0,1	3,56
3,5	0,15	0,16	0,25	0,2	0,25	2,77
3,5	0,15	0,16	0,25	0,25	0,3	2,51
3,5	0,15	0,16	0,25	0,3	0,33	2,32
3,5	0,15	0,16	0,25	0,35	0,4	2,05

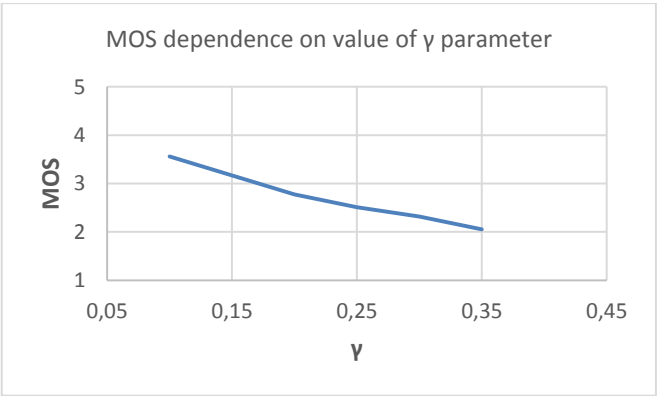


Figure 3. MOS dependence on value of γ factor in Scenario 1

TABLE VIB
PARAMETERS FOR SCENARIO 1 – INCREASE OF A

QoE_1	α_1	α_2	β	γ_1	γ_2	QoE_{1FIN}
3,5	0,15	0,16	0,25	0,1	0,15	3,39
3,5	0,2	0,22	0,25	0,1	0,15	3,78
3,5	0,25	0,3	0,25	0,1	0,15	4,30
3,5	0,31	0,35	0,25	0,1	0,15	4,81
3,5	0,35	0,4	0,25	0,1	0,15	5,00

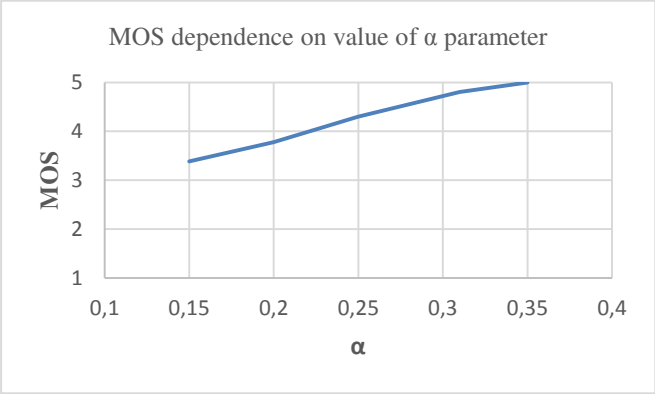


Figure 4. MOS dependence on value of α factor in Scenario 1

TABLE VIC
PARAMETERS FOR SCENARIO 1 – INCREASE OF B

QoE_1	α_1	α_2	β	γ_1	γ_2	QoE_{1FIN}
3,5	0,15	0,16	0,25	0,1	0,15	3,39
3,5	0,15	0,16	0,3	0,1	0,15	3,56
3,5	0,15	0,16	0,35	0,1	0,15	3,74
3,5	0,15	0,16	0,4	0,1	0,15	3,93
3,5	0,15	0,16	0,45	0,1	0,15	4,14

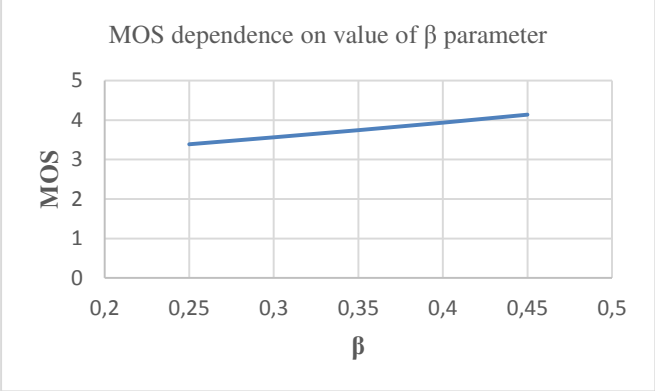


Figure 5. MOS dependence on value of β factor in Scenario 1

Based on Scenario 1 with a correct authentication in third step increase of α, β, γ parameters will be considered with their impact on a final QoE value. Figure 3 shows a situation in which γ increases and QoE value decreases. According to formula (7) it is a proper behavior. The γ factor is a parameter of QoE, which informs about decreeing of user's satisfaction. The greater factor is, the worse final QoE value is. Figure 4 shows the case in which α parameter increases. As it was shown in Figure 4, the greater α values are, the greater final QoE value is. But, wrong choice of α parameters results in a too high final value of QoE: for greater α the value of QoE is greater than 5 (but of course in MOS scale it will be corrected to maximum 5).Thus, α parameter should be determined on a proper, not too high level. Finally, Figure 5 presents β parameter impact on the final QoE value. As it was shown, QoE value increases with β increasing. And it is a proper situation, because β is a factor which value describes user's satisfaction when he/she correctly authenticates him/herself. Moreover, Figure 5 shows that β has not got such an impact on QoE as α has.

Comparing all these three factors we can notice that a significant influence has the α parameter. As a result the final QoE value can be overestimated.

TABLE VIIA
PARAMETERS FOR SCENARIO 2 – PART 1

QoE_1	α_1	α_2	α_3	γ_1	γ_2	QoE_{1FIN}
3,5	0,15	0,16	0,17	0,1	0,15	3,13
3,5	0,15	0,16	0,17	0,2	0,25	2,56
3,5	0,15	0,16	0,17	0,25	0,3	2,32
3,5	0,15	0,16	0,17	0,3	0,33	2,14
3,5	0,15	0,16	0,17	0,35	0,4	1,90

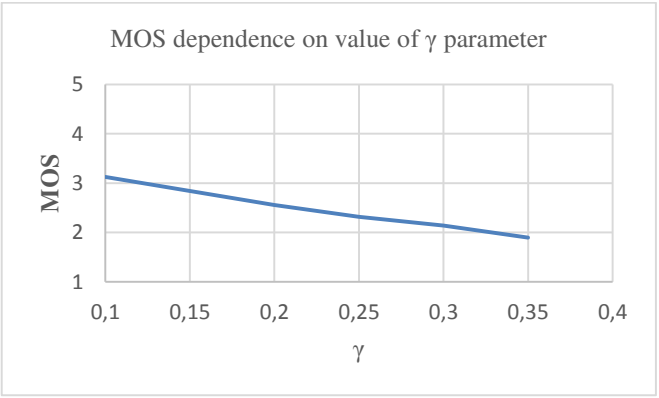


Figure 6. MOS dependence on value of γ factor in Scenario 2 - part 1

TABLE VIIB
PARAMETERS FOR SCENARIO 2 – PART 1

QoE_2	α_1	α_2	β	γ_1	γ_2	QoE_{2FIN}
3,5	0,15	0,16	0,17	0,1	0,15	3,13
3,5	0,2	0,22	0,24	0,1	0,15	3,74
3,5	0,25	0,3	0,33	0,1	0,15	4,66
3,5	0,31	0,35	0,37	0,1	0,15	5,00
3,5	0,35	0,4	0,43	0,1	0,15	5,00

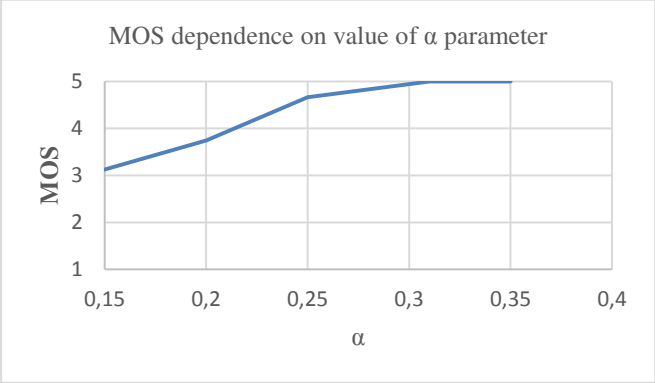


Figure 7. MOS dependence on value of α factor in Scenario 2 - part 1

Considering part 1 of Scenario 2 it can be noticed that α and β factors have similar tendency like in the Scenario 1 – an increase of γ results in decrease of QoE and an increase of α results in too fast QoE increase.

TABLE VIIIA
PARAMETERS FOR SCENARIO 1 – PART 2: INCREASE OF Γ

QoE_1	α_1	α_2	β	γ_1	γ_2	QoE_{1FIN}
3,5	0,15	0,16	0,25	0,1	0,15	3,39
3,5	0,15	0,16	0,25	0,2	0,25	2,77
3,5	0,15	0,16	0,25	0,25	0,3	2,51
3,5	0,15	0,16	0,25	0,3	0,33	2,32
3,5	0,15	0,16	0,25	0,35	0,4	2,05

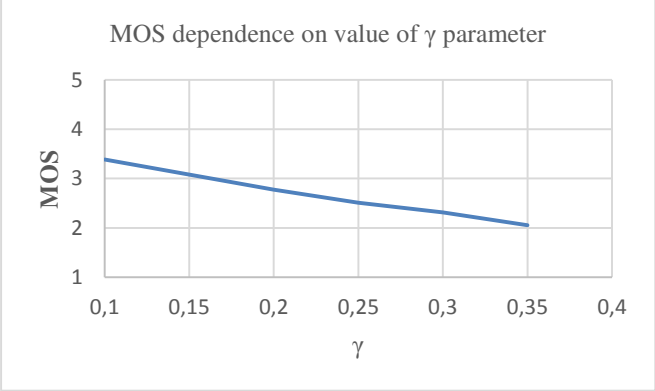


Figure 8. MOS dependence on value of γ factor in Scenario 2 - part 2

TABLE VIIIB
PARAMETERS FOR SCENARIO 1 – PART 2: INCREASE OF A

QoE_1	α_1	α_2	β	γ_1	γ_2	QoE_{1FIN}
3,5	0,15	0,16	0,25	0,1	0,15	3,39
3,5	0,2	0,22	0,25	0,1	0,15	3,78
3,5	0,25	0,3	0,25	0,1	0,15	4,30
3,5	0,31	0,35	0,25	0,1	0,15	4,81
3,5	0,35	0,4	0,25	0,1	0,15	5,00

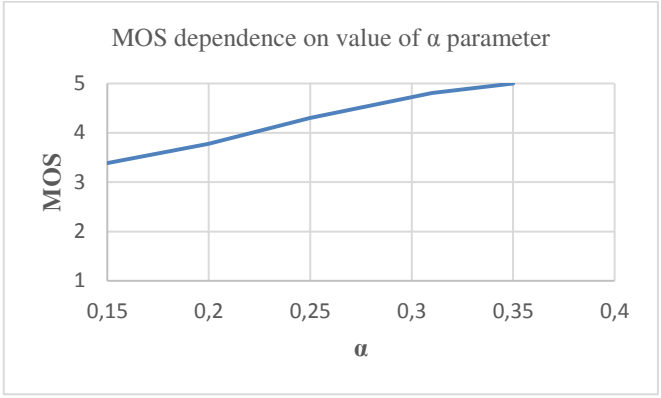


Figure 9. MOS dependence on value of α factor in Scenario 2- part 2

TABLE VIIIC
PARAMETERS FOR SCENARIO 1 – PART 2: INCREASE OF B

QoE_1	α_1	α_2	β	γ_1	γ_2	QoE_{1FIN}
3,5	0,15	0,16	0,25	0,1	0,15	2,89
3,5	0,15	0,16	0,3	0,1	0,15	3,03
3,5	0,15	0,16	0,35	0,1	0,15	3,19
3,5	0,15	0,16	0,4	0,1	0,15	3,35
3,5	0,15	0,16	0,45	0,1	0,15	3,52

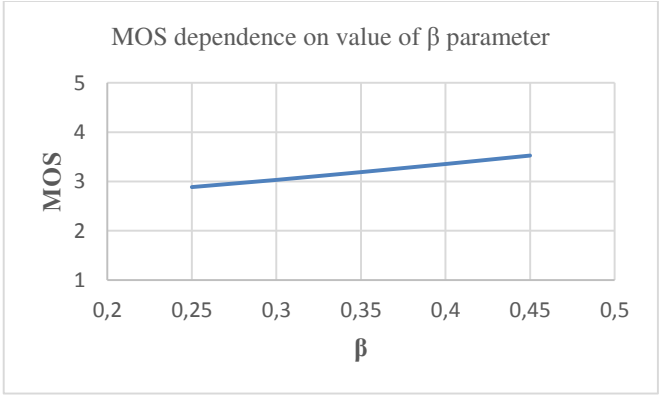


Figure 10. MOS dependence on value of β factor in Scenario 2 - part 2

Considering part 2 of Scenario 2 we can notice that it is similar to Scenario 1 – α parameter has a meaningful impact on the final QoE value.

VI. Conclusions and Future Work

In this paper a model of selecting the best authentication mechanism has been proposed. The model consists of two stages. The first stage constructs a decision tree algorithm, which is used to calculate three parameters characterizing the authentication: a probability of choosing particular path describing the steps of authentication, a final QoP value and a final QoE value. The second stage concerns multi-objective optimization of the measures established in the first stage. Based on a weighted sum method applied to QoP and QoE measures for the authentication mechanism the best path is chosen. Moreover, a discussion about an impact of QoE parameters on a final value of QoE was

conducted. The assumptive final QoP and QoE formulas fulfill requirements applied to acceptable values. In spite of all, experiments which verify created formulas should be performed.

To illustrate the results a simple numerical example has been presented. In our future work the research will be concentrated on conducting a real-world experiment of a reasonable scale authentication mechanism based on the created model. This will make possible to compare theoretical and experimental results to verify the model and tune its numerical parameters.

REFERENCES

[1] Wang, Z., Crowcroft, J., "Quality-of-service routing for supporting multimedia applications", IEEE JSAC, vol. 14, no. 7, pp. 1228-1233, 1996 DOI: 10.1109/49.536364

[2] "Qualinet White Paper on Definitions of Quality of Experience" Output from the fifth Qualinet meeting, Novi Sad, March 12, 2013 Version 1.2

[3] Reichl, P., Egger, S., Möller, S., Kilkki, K., Fiedler, M., Hossfeld, T., Tsiaras, Ch., Asrese, A., "Towards a comprehensive framework for QoE and user behavior modelling", Seventh International Workshop on Quality of Multimedia Experience (QoMEX), 2015 DOI: 10.1109/QoMEX.2015.7148138

[4] Hossfeld, T., Fiedler, M., Tran-Gia, P., "A Generic Quantitative Relationship between Quality of Experience and Quality of Service", IEEE Network Special Issue on Improving QoE for Network Service, March 2010 DOI:10.1109/MNET.2010.5430142.

[5] Ciszkowski, T., Mazurczyk, W., Kotulski, Z., Hossfeld, T., Fiedler, M., Collange, D., "Towards Quality of Experience-based Reputation Models for Future Web Service Provisioning", Telecommunication Systems, Vol.51, No.4, pp.283-295, (2012) DOI: 10.1007/s11235-011-9435-2.

[6] Gerstel, O., Sasaki, G., "Quality of Protection (QoP): a quantitative unifying paradigm to protection service grades", in: SPIE Proc. OptiComm 2001, vol. 4599, (2001a), pp. 12–23 DOI: 10.1117/12.436060.

[7] "Quality of Protection Security Measurements and Metrics", Editors: Gollmann, Dieter, Massacci, Fabio, Yautsiukhin, Artsiom (Eds.), Springer 2006 DOI: 10.1007/978-0-387-36584-8.

[8] Księżopolski, B., Kotulski, Z., "Adaptable security mechanism for dynamic environments", Computers & Security, Vol.26, No.3, pp.246-255, (2007) DOI: 10.1016/j.cose.2006.11.002.

[9] Siewruk G., Średniawa M., Grabowski S., Legierski J. , "Integration of context information from different sources: Unified Communication, Telco 2.0 and M2M", Proceedings of the 2013 Federated Conference on Computer Science and Information Systems pp. 851–858

[10] Schilit, B., Adams, N., Want, R., "Context-Aware Computing Applications", Proceeding WMCSA '94 Proceedings of the 1994 First Workshop on Mobile Computing Systems and Applications, pp.: 85 – 90 DOI: 10.1109/WMCSA.1994.16.

[11] Pascalau E., Nalepa G.J., Kluza K., "Towards a Better Understanding of Context-Aware Applications", Proceedings of the 2013 Federated Conference on Computer Science and Information Systems pp. 959–962

[12] Wrona, K., Gomez, L., "Context-aware security and secure context- awareness in ubiquitous computing environments", XXI Autumn Meeting of Polish Information Processing Society, Conference Proceedings, pp.: 255 – 265

[13] Alves, P., Ferreira, P., Radiator, "Efficient message propagation in context-aware systems", in Journal of Internet Services and Applications, 2014, DOI: 10.1186/1869-0238-5-4.

[14] Orynczak, G., Kotulski, Z., "Context-Aware Secure Routing Protocol for Real-Time Services", in: Cryptography and Security Systems, Volume 448 of the series Communications in Computer and Information Science pp 193-207, Springer 2014 DOI: 10.1007/978-3-662-44893-9_17.

[15] Wenning, B-L., "Context-Based Routing", in Dynamic Networks, Springer 2010 DOI: 10.1007/978-3-8348-9709-1.

[16] Goel, D., Kher, E., Joag, S., Mujumdar, V., Griss, M., Dey, A. K., "Context-Aware Authentication Framework", Proc. First Annual Conference on Mobile Computing, Applications, and Services (MobiCASE 2009), pp. 26-29 DOI: 10.1007/978-3-642-12607-9_3.

[17] Lenzini, G., "Trust-Based and Context-Aware Authentication in a Software Architecture for Context and Proximity-Aware Services", in Architecting Dependable Systems VI Volume 5835 of the series Lecture Notes in Computer Science, 2009, pp. 284-307 DOI: 10.1007/978-3-642-10248-6_12.

[18] Park, S., Han, Y., Chung, T., "Context-Aware Security Management System for Pervasive Computing Environment", in Modeling and Using Context Volume 4635 of the series Lecture Notes in Computer Science pp 384-396, August 2007, pp. 20-24 DOI: 10.1007/978-3-540-74255-5_29.

[19] Hayashi, E., Das, S., Amini, S., Hong, J., Oakley, I., "CASA: Context- Aware Scalable Authentication", Proc. of the Ninth Symposium on Usable Privacy and Security, ISBN 978-1-4503-2319-2, July 2013 DOI: 10.1145/2501604.2501607.

[20] Kulkarni, D., Tripathi, A., "Context-aware role-based access control in pervasive computing systems", Proc. of the 13th ACM symposium on Access control models and technologies (SACMAT '08), June 2008, pp. 113-122 DOI: 10.1145/1377836.1377854.

[21] Chun-Dong, W., Ting, L., Li-Chun, F., "Context-Aware Environment-Role-Based Access Control Model for Web Services", in Multimedia and Ubiquitous Engineering, ISBN 978-0-7695-3134-2, April 2008, pp. 288 – 293 DOI: 10.1109/MUE.2008.77.

[22] Khan, M., F., F., Sakamura, K., "Context-aware access control for clinical information systems" , in Innovations in Information Technology (IIT), ISBN 978-1-4673-1100-7, March 2012, pp. 123 – 128 DOI: 10.1109/INNOVATIONS.2012.6207715

[23] Krawczyk, H., Lubomski, P., "User Trust Levels and Their Impact on System Security and Usability", Proc. of

22nd International Conference Computer Networks, ISBN 978-3-319-19418-9, May 2015, pp. 82 – 91 DOI: 10.1007/978-3-319-19419-6_8

[24] Furnell, S., M., Jusoh, A., Katsabas, D., "The challenges of understanding and using security: A survey of end - user", in Computers and Security vol. 25 issue 1, February 2006, pp. 27 – 35 DOI: 10.1016/j.cose.2005.12.004.

[25] Furnell, S., "Usability versus complexity – Striking the balance in end – user security ", in Network Security, December 2010, pp. 13 – 17 DOI: 10.1016/S1353-4858(10)70147-1.

[26] Wu, D., Zhang, H., Wang, H., Wang, C., Wang, R., Xie, Y., "Quality of protection – driven data forwarding for intermittently connected wireless networks", in IEEE Wireless Communications vol. 22 issue 4, August 2015, pp. 66 – 73 DOI: 10.1109/MWC.2015.7224729.

[27] Li, H., Liu, D., Dai, Y., Luan, T., H., „Engineering searchable encryption of mobile cloud networks: when QoE meets QoP", in IEEE Wireless Communication vol. 22 issue 4, August 2015, pp. 74 – 80 DOI: 10.1109/MWC.2015.7224730.

[28] Wang, W., Zhang, Q., "Toward long-term quality of protection in mobile networks: a context-aware perspective", in IEEE Wireless Communications vol. 22 issue 4, August 2015, pp. 34 – 40 DOI: 10.1109/MWC.2015.7224725.

[29] Lorentzen, C.; Fiedler, M.; Johnson, H.; Shaikh, J.; Jrstad, I., "On user perception of web login — A study on QoE in the context of security", in Telecommunication Networks and Applications Conference (ATNAC), Auckland, Oct. 31 2010-Nov. 3 2010, pp. 84 – 89 DOI: 10.1109/ATNAC.2010.5680262

[30] Lorentzen, Ch., "User Perception and Performance of Authentication Procedures", Thesis, Blekinge Institute of Technology, School of Computing, 2011.

[31] Lorentzen, Ch., "On User Perception of Authentication in Networks", PhD. Thesis, Blekinge Institute of Technology 2014.

[32] Sepczuk, M., "Security oriented on user's perception in cloud computing", in Przegląd Telekomunikacyjny, no. 8-9, 2013, pp. 1245 – 1251.

[33] Crawford, H., Renaud, K., "Understanding user perceptions of transparent authentication on a mobile device", in Journal of Trust Management vol. 1 issue 1, June 2014 DOI: 10.1186/2196-064X-1-7.

[34] Eliasson; Ch., Fiedler, M.; Jørstad, I., "A Criteria-Based Evaluation Framework for Authentication Schemes in IMS", ARES '09. International Conference on Availability, Reliability and Security, 2009 DOI: 10.1109/ARES.2009.166.

[35] Kotulski, Z., Sepczuk, M., Sitek, A., Tunia, M. A., „Adaptable Context Management Framework for Secure Network Services", in Annales UMCS Informatica, vol. 14, no.2, September 2014, pp. 7 – 30 DOI: 10.2478/umcsinfo-2014-0013.

[36] Sepczuk, M., "Authentication Mechanism Based on Adaptable Context Management Framework for Secure Network Services", in Annales UMCS, Informatica, vol. 14, no.2, September 2014, pp. 31-44 DOI: 10.2478/umcsinfo-2014-0010.

[37] Irvine, C., Levin, T., "Quality of Security Service", Proceeding NSPW '00 Proceedings of the 2000 workshop on New security paradigms, Pages 91 - 99, doi: 10.1145/366173.366195.

[38] EL Yamany, H., F., Capretz, M., Allison, D., S., "Quality of Security Services for Web Services within SOA", in Congress on Services – I, July 2009, pp.: 653 – 660 DOI: 10.1109/SERVICES-I.2009.95.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1031–1040
DOI: 10.15439/2016F87
ACSIS, Vol. 8. ISSN 2300-5963

On Constructing Persistent Identifiers with Persistent Resolution Targets

Oliver Wannenwetsch
Gesellschaft für wissenschaftliche Datenverarbeitung Göttingen (GWDG),
Göttingen, Germany
oliver.wannenwetsch@gwdg.de

Tim A. Majchrzak
Department of Information Systems
University of Agder,
Kristiansand, Norway
timam@uia.no

***Abstract*—Persistent Identifiers (PID) are the foundation referencing digital assets in scientific publications, books, and digital repositories. In its realization, PIDs contain metadata and resolving targets in form of URLs that point to data sets located on the network. In contrast to PIDs, the target URLs are typically changing over time; thus, PIDs need continuous maintenance – an effort that is increasing tremendously with the advancement of e-Science and the advent of the Internet-of-Things (IoT). Nowadays, billions of sensors and data sets are subject of PID assignment. This paper presents a new approach of embedding location independent targets into PIDs that allows the creation of maintenance-free PIDs using content-centric network technology and overlay networks. For proving the validity of the presented approach, the Handle PID System is used in conjunction with Magnet Link access information encoding, state-of-the-art decentralized data distribution with BitTorrent, and Named Data Networking (NDN) as location-independent data access technology for networks. Contrasting existing approaches, no green-field implementation of PID or major modifications of the Handle System is required to enable location-independent data dissemination with maintenance-free PIDs.**

***Index Terms*—Persistent Identifier; Information Centric Networks; Named Data Networking; Magnet Link; URN; Handle System; Digital Object Identifier; Overlay Network**

I. Introduction

The concept of *Persistent Identifier (PID)* is essential for referencing, citing and linking (digital) resources using a durable and reliable identifier. PIDs are used to ensure the long-term valid access to possibly moving digital resources that suffer from changing URLs and storage locations in networks. PIDs contain an adjustable target Uniform Resource Locator (URL). To reflect changing and volatile data locations, the target URL of a PID must be updated to the currently valid storage location. By employing organizational and technical measurements, different PID systems allow building, using and maintaining a long-term existing (digital) identifier that is backed by distributed systems, replication schemes, and policies. With these measurements in place, PID infrastructure operating organizations are able to offer PID systems that are resilient against failure, and even catastrophic scenarios. Ideally, the range of PID infrastructure resilience includes scheduled downtimes of server and networks, major infrastructure problems caused by power failures, and hardware and software problems, as well as ultra critical events of complete data center losses caused by fires, explosions or natural disasters.

Besides the measurements that protect PIDs on infrastructure level, PIDs have to be protected on the content-side as well. The content of the PID has to be intact and readable with given encoding schemes. Furthermore, the metadata have to be addressed with a given metadata scheme that explains the semantics of the metadata. Then, the content has to reflect the current state of the data object the PID is linking to. This includes the *up-to-dateness* of metadata sets stored in PIDs, which are often encoded as *key-value* pairs. Particularly important is the correctness of the PID metadata field `target URL`, which points to the digital object addressed by the PID for long-term access (c.f. Figure 1). Contrasting the infrastructure protection of PID, this content validation is not done by the PID infrastructure operating organization. It is a task of the data owners that registered the PID for their data or the subsequent organization that has the task of curating the data and its associated PIDs. Only those organizations have the necessary understanding on the data and its metadata to check and update PIDs for assuring long-term access. Moreover, they are aware of the current location of the data linked by PID. Thus, with regular control and adjustment of the target URLs to the current data location, well-established PID systems can guarantee persistency of identifiers physically, while data owner organizations have to accept the burden of regularly checking and updating metadata and target URLs. Only with collaboration PIDs are able to provide long-term access [1].

While PIDs solve the problem of changing data locations by constant efforts from PID infrastructure providers and data hosting organizations, the network research community has come up with numerous concepts for creating location-independent data access. In these concepts that access information for data attached to a network is not based on the data location, but it is based on the content of the data [2]. Hence, in the case of changing data location the access information remain stable. Two efforts that realize location-independent access is the state-of-the-art decentralized data distribution technology *BitTorrent* and *Named Data Networking (NDN)* as next-generation Internet technology. Although both technologies allow stable location-independent access to data, they do not provide persistent access like PID.

Our approach presented in this paper combines the very stable concept of the *Handle* PID system with the advantages of

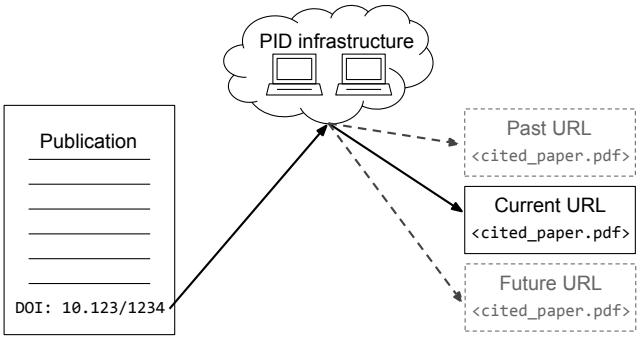


Figure 1: Adjustment of PID target URLs to reflect current data location.

location-independent access with its stable access information scheme. By this, we significantly lower the efforts of PID maintenance, as regular checks for data hosting organizations concerning locations and data availability will become obsolete. PIDs that are enabled for location-independent access remain valid as long as data is available online in BitTorrent or Named Data Networking (NDN) networks. To store the location-independent access information in existing Handle PID data structures, we extend the – yet not standardized – approach of *Magnet Link Schemes*, which is very successful in peer-to-peer communities [3]. In detail, our novel approach has the benefits of using an unmodified version of the Handle PID system that allows a practical implementation of our approach in existing Handle PID systems. Besides the advantages of interoperability, it does not come with a significant impact on PID resolution. For the location-independent access with BitTorrent and NDN only the small overhead of PID resolution is added.

Furthermore, we extend the Magnet Link scheme into the domain of NDN and provide a transport container format that includes besides the NDN data name also the necessary cryptographic access information for verifying data access authenticity. Thus, we can also improve Web browser-based NDN applications that use HTML-links for interconnecting NDN Web resources by the Magnet Link scheme. The latter has been originally designed for interconnecting non HTTP-resources in hypertext document contexts.

This paper is structured as follows. First, we present existing efforts in Section II. Second, we summarize the existing approaches of location-based data access through PID in Section III. Then, we line out state-of-the-art techniques for location-independent access in Section IV. In Section V, we introduce the Magnet Uniform Resource Identifier (URI) scheme format as container for storing location-independent access information. After that, we extend the Magnet URI scheme for the usage in the domain of content-centric networks and PID in Section VI. A proof-of-concept implementation is illustrated in Section VII together with an evaluation of performance in Section VIII that is combined with a discussion of the results. Finally, we draw a conclusion in Section X.

II. Related Work

To our knowledge, the concept of building care-free persistent identifier targets using content-centric technology has not been subject of extensive study. In the literature several related concepts can be identified.

The concept of bridging different content-centric network systems through a centralized Uniform Resource Name (URN) system has been initially drafted by Sollins in 2012 [4]. Her concept utilizes foundations of PID principles for creating an identification system for different Information Centric Networks (ICN) families and their related data objects that meets the requirements of scalability, longevity, evolvability, and security. Sollins's identification system abstracts different object naming schemes from ICN families such as Data Oriented Network Architecture (DONA) [5], Network of Information (NETINF) [6] and Publish-Subscribe Architecture (PURSUIT) [7]. Although, PID principles are used for location-independent data access, the publication by Sollins does not suggest access through an existing well-introduced PID system that is provided in our work, but rather uses a greed-field approach for location-independent data access.

The realization of complex secure naming schemes for content-centric data has been covered by Dannewitz et al. in 2010 [8]. They demand name persistency without incorporating the concept of PIDs. In their publication, Dannewitz et al. clarify that basic security functionality must be attached directly to the data and its naming scheme, because the identity of network locations cannot be used as a trust base for data authenticity. Our approach follows this principle for secure location-independent data access and facilitates directly attached PID security mechanisms. By this, location-independent access through PID is shifted into the requirements formulated by Dannewitz et al., and our approach enables authentic data access through PIDs.

In the context of semantic digital archives for archiving data of Personal Information Manager (PIM) applications, Haun and Nürnberger proposed a PID schema for accessing objects in file systems using an URN-like Magnet Links scheme [9]. They link the congruent attributes of the Magnet Link scheme to the attributes provided by some PID systems such as global uniqueness, persistence and scalability for the application in offline data archives serving data from archive medium such as file systems on Write Once Read Multiple (WORM) medium. In contrast, our approach relies on currently employed PID systems and incorporates location-independent data access in a distributed online environment using the full-featured Handle PID system.

III. Location-based PID Targets

Today's data dissemination is dominated by end-to-end connections, URLs, and DNS-backed domain names. When data is moved from one host to the other, it results in broken URLs and inaccessible content. To ease these problems, PIDs are used for long-term data access by providing a long-living identifier. When using a PID, the identifier is embedded into a

medium such as scientific publications, books, or Web sites. To access the digital resources, *behind* the PID a resolution service is employed that uses the PID to provide a currently valid network location (target URL). Figure 1 illustrates how the target URL is adjusted to the current location when the data behind the PID is moved from one host to the other. Thus, PIDs reflect the current location of data and the identifier on the medium can remain unchanged.

The location-dependent data access through target URLs stored in PIDs also forms the chain of PID resolution. For this we have a look at the resolution chain presented in Figure 2. It depicts the fact that data access through PID with Hypertext Transport Protocol (HTTP) relies on different infrastructures and involves five levels from the PID resolution up to the data download. Even if the chain is shortened, e.g. by directly linking PID targets to IP-addresses instead of DNS-based host names, the problem remains identical: If PID-tagged data is moved from one host to the other, the PID access chain needs to be adjusted on one or even on multiple levels to provide valid PID resolution. If the adjustment is not done or partially incorrect, location-depended data access is impossible through PID. Defects can occur on every level. On level ❶ the PID HTTP resolution service can be temporary out of order. The target URL does not reflect the current location of the data in stage ❷. On level ❸ the Domain Name System (DNS) resolution can fail if a domain has been expired or DNS resolution fails due to misconfiguration. Level ❹ and ❺ are related to the network, but their functionality is also required for successful data access through PID. To detect broken PID resolution a check of every PID target and, thus, a successful resolution is necessary for judging the integrity. The check is done in many cases by evaluating the HTTP status codes like `404 - not found` that are provided by the data repositories software [10]. This can only be achieved by regularly checking *all* PIDs of an organization. This is very time consuming since typically robot or spider programs crawl all PIDs of a data owner. The crawling programs are programmed and operated by repository owners and not PID infrastructure providers. They provide an optional data quality assurance service.

The adjustment of target URLs has impact on different dimensions and is shared unevenly between the users – the PID operators and data repository owners. The costs and efforts behind URL adjustments have been accepted as part of the PID operation. They are considered *inevitable*, such as energy leaks in today's electrical grid infrastructure. The adjustment of target URLs is a shared effort on the side of data owners and dependent on the number of PIDs a data owner has registered.

It can be questioned whether the proliferation of e-Science already increases the effort necessary for PID maintenance. We thus have visualized statistics from DataCite (one of the largest PID infrastructure providers) in Figure 3. The assignment of new PIDs (*line*) massively increased, following a super-linear pattern [11]. An aggregation of the DataCite Statistics for successful DOI PID resolutions shows also a massive increase in PID-tagged data sets (*bar*) [12] [13]. With a massive increase of PID numbers, the efforts for maintaining PID targets will

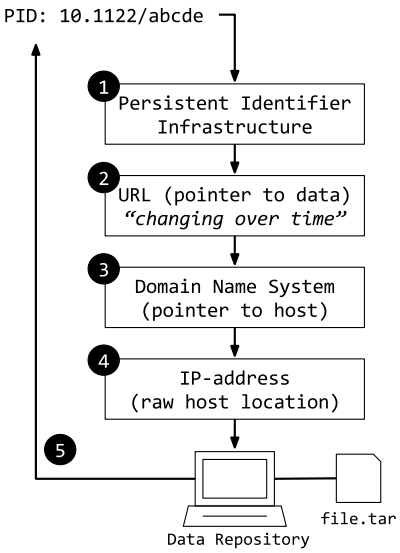


Figure 2: Data access through PID requires a working chain of services relying on unstable URLs.

increase identically, as every assigned PID needs to be checked for validity to comply with PID infrastructure policy. It is not a question, whether the PID systems are scaling out sufficiently, but rather the data hosters are able to verify their location-dependent PIDs with a reasonable effort regularly. PIDs with location-independent targets decouple the growing number of PIDs from the efforts of maintaining PID target URLs.

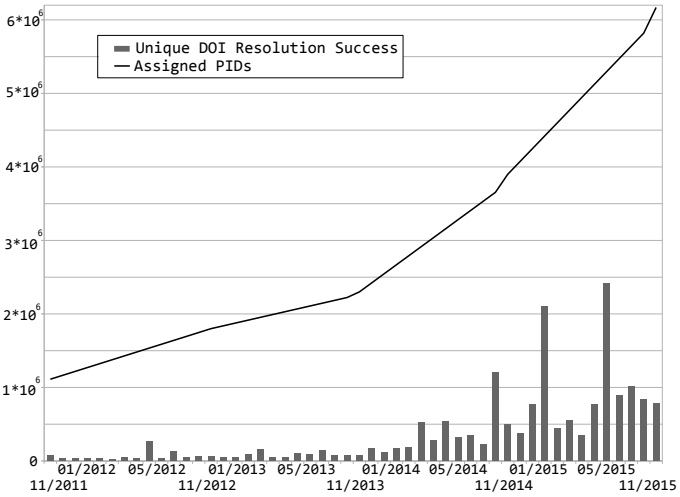


Figure 3: PID assignment and unique successful resolution for the DataCite DOI infrastructure between 11/2011 and 11/2015 based on data from [11] [12].

IV. Location-Independent Access

For location-independent data access, we propose two different techniques that allow access based on the content and not on the network location. We can thus show that our approach of location-independent persistent PID resolution targets works with various location-independent access technologies.

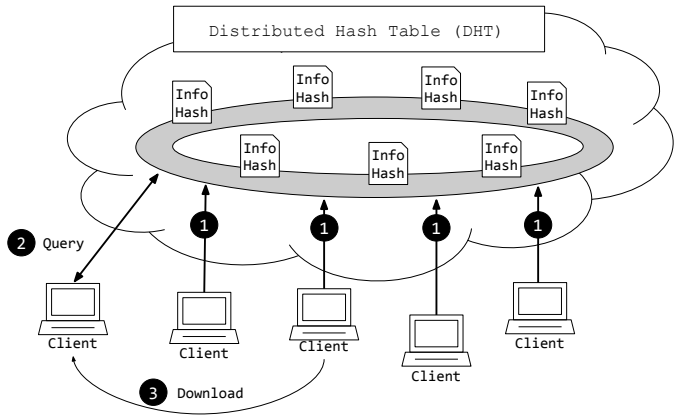


Figure 4: Accessing data from a DHT-controlled swarm using an infohash in BitTorrent.

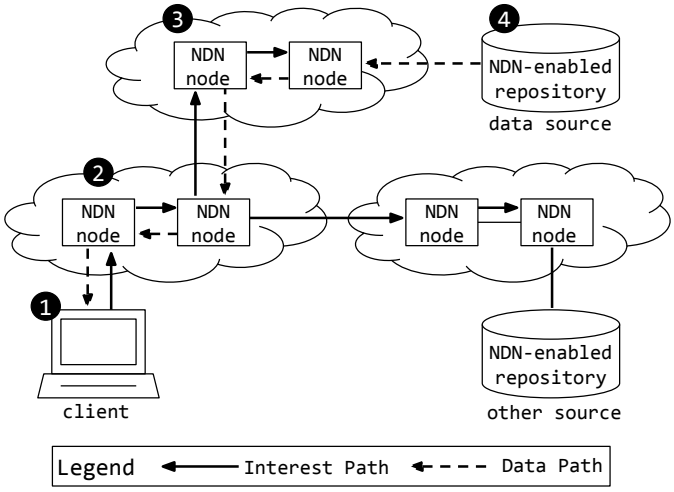


Figure 5: Accessing data using data names in NDN networks.

We propose BitTorrent, a well-established location-independent access technology. It works on top of today's location-based networks with Transmission Control Protocol (TCP) and User Datagram Protocol (UDP) [14]. In contrast to existing location-based data repositories that are subject of PID target resolution, BitTorrent technology uses a peer-to-peer approach supporting parallel downloads. With its latest features of Distributed Hash Table (DHT) and Peer Exchange (PEX), BitTorrent does not require central infrastructure to discover other network peers and localize files [15] [16]. Thus, data can be accessed with BitTorrent from every connected peer, as long as data is available online. For addressing data sets, BitTorrent uses *infohashes* that are computed as SHA-1 checksums on the content of the file. Every peer that possesses the infohash can download the data set from the BitTorrent *swarm* that consists of the peers offering the data set for download. The swarm arrangement and the *overlay network* for the specific file is computed for every download. In Figure 4, the BitTorrent file access is depicted. In step ❶, every node is sending its network address and the infohashes of the files ready for upload to the DHT. Then, a client can look up the infohash in the DHT (step ❷) to locate other peers that are able to serve the file that belongs to the infohash (or at least parts of the file). Through connecting to the peers in step ❸, the client is retrieving the file. This can be done simultaneously by parallel peer connection.

In contrast to BitTorrent, Named Data Networking (NDN) is a current research topic of location-independent data access using information-centric principles [17]. NDN is also featured in the location-independent PID approach presented in this paper to support a next-generation Internet technology. In NDN, data sets are enumerated through *Data Names* that form a hierarchical name space [17]. The working principle of NDN is shown in Figure 5. To access data from a client (step ❶) in the *Named Data* space, an interest data package is sent through the network. Based on the data name driven routing principles, the NDN network directs the interest through the network (step ❷ and ❸). If a NDN node is found that owns a named data set (step ❹), a data package is sent back along the interest path to reach the node, which stated the data request. Hence, NDN abstract from the network location and the data source remains opaque [17].

V. Magnet URI Scheme

For embedding location access information into PID, the Magnet URI scheme is used as transport container, which is a work-in-progress specification for *Magnet Links* [18]. Magnet Links can store extensive information on accessing resources in networks, like HTTP download URLs, mirror server information, or peer-to-peer access data. By applying this principle, Magnet Links are usable for describing digital resources and its content. But as a descriptive access format, Magnet Links need a storage medium to be present on. This could be a Web site with HTML content, an E-Mail message content, or – like in our use case – a Persistent Identifier. To provide persistent access to digital data encoded in Magnet Links for a time of years, or possibly decades, a medium is needed that provides these properties. Hence, persistency is not archived by the Magnet Links but provided by the PID System that ensures long living existence of access information through its infrastructure, policies and replication partnerships. If Magnet Links are stored on *perishable* media like Web sites, they do not provide any advantage over common URL access information. The conjunction with PID provides the additional benefit of long-term access from data hosted at suddenly moving data. Besides the encoding of access information, one design goal of Magnet Links is to integrate features made available by local utility programs seamlessly into the storage medium like a Web site, by following best practices among the Internet Engineering Task Force (IETF) specifications for URN [19]. Although its lack of specification, Magnet Links are supported by numerous peer-to-peer tools and are the de facto standard in large file sharing communities, such as BitTorrent [20]. In BitTorrent-enabled Magnet Links, access information for downloading

files from a decentralized peer-to-peer infrastructure are stored together with optional metadata and suggested file names (c.f. Tab. I). To initiate a download with a Magnet Link from a Web browser, a pseudo-protocol handler for the Magnet Link URL format `magnet:?xt=urn:<System>:<Access Information>` is registered. It passes the information to a location independent download client.

Table I: Magnet URI scheme keys

Key	Name	Purpose
as	Acceptable Source	location dependent download URL
dn	Display Name	file name
kt	Keyword Topic	search key word
tr	Address Tracker	optional tracker information for BitTorrent
xl	Exact Length	size in bytes
xt	Exact Topic	location independent access information in URN-format

VI. Location-Independent Data Access through PID

For creating a persistent resolution target in a Handle PID that is based on the content of linked data set and not the network location like the target URL, we leverage the principles of location-independent data access with Magnet URI-encoded access information. In the first step in Subsection VI-A, we extend the Magnet URI scheme into the domain of NDN applications to support this cutting-edge data access technology. By this, we can encode all access information for BitTorrent and NDN as well as all systems listed in Table II into one uniform scheme.

Then, in the second step in Subsection VI-C, we propose an approach of embedding the Magnet Link URI scheme into PIDs of an unmodified Handle System. Starting from Subsection VI-E onward, we examine the PID resolution of PIDs with enabled location-independent access data.

A. Magnet URI Scheme Extension for NDN

Besides the already established usage of the Magnet URI scheme in peer-to-peer overlay networks that particularly allow BitTorrent location-independent data access, a next-generation access technology for supporting location-independent data access is integrated in our approach. We extend the Magnet URI schema into the domain of content-centric networks, enabling support for Named Data Network data access through Magnet Links. Thus, the Magnet URI schema is extended to store a NDN data name that identifies a digital object within a NDN network. With the data name, the NDN network can transport data from a source node holding the data back to the client node requesting the information through an interest [17]. The location of the data is not important, as long as it is attached to the network through a reachable NDN node. The data name is encoded through an extended `xt` key that holds besides the data name also a checksum of the data name to detect data corruption. The schema is `uid:ndn<DATANAME>.<CHECKSUM>`. Unlike current host-based networks that use a Public Key Infrastructure (PKI) to verify host identities through SSL certificates, NDN data cannot verified through a trustful data location. As a result, the Magnet Link also has to include the verification information needed to assure that the received content *is* the requested content. This is done by adding a cryptographic signature of the NDN content in a separate `<SIGNATURE>` field, which is part of the second `xt` key extension. Thus, verification needs to be done on content level using a public PKI, which is in the scope of current NDN research. To get the certificate needed to verify the data, the NDN access information for the certificate need to be added to the Magnet Link, too. To obtain the certificate with the public key, we propose the `xt` key `uid:ndnsec<SIGNATURE>.<CERT_DATANAME>` that allows the download of information for content verification. By this, access NDN access information can be encoded into a Magnet Link and also the genuineness of the obtained data can be verified through security information embedded into the Magnet Link. The extension we provide for the Magnet URI scheme supporting NDN is depicted in bold letters in Table II.

B. Embedding Magnet Links in Handle PIDs

For this integration of Magnet Links into the Handle PID System maximum compatibility is paramount, as data dissemination has a very slow change momentum, owed to billions of PID-tagged data sets. Hence, the usage of Magnet Links for location independent data access and its impact on the adaption in the Handle System is investigated. By design, Handle supports hierarchical data types, identified by UTF-8 named fields. The data itself is organized as indexed, typed key-value pairs that store sequences of octets, which are preceded by its length in a 4-byte unsigned integer. Like other PID systems, the Handle System provides a URL data type `0.TYPE/URL` [21]. Supplemental services such as PID HTTP resolvers, also known as Handle proxies, use the URL semantic to resolve PIDs into target URLs by using HTTP-forwarding with HTTP status code `303`.

Table II: Magnet Links Scheme Adopters (proposed schemes for Named Data Networking in bold font)

System	URN	Value
Gnutella2	sha1	file hash (SHA-1)
BitTorrent	btih	unique file identifier
Gnutella2	tiger	file hash (Tiger Tree Hash)
Kazaa	kzhash	file hash (proprietary)
NDN Access	**ndn**	**DataName and Checksum (SHA256)**
NDN Verification	**ndnsec**	**content signature & public key data name (NDN specs.)**

Unlike URLs, Magnet Links do not specify the data location but can be considered as URN-like data classification. Hence, the Handle PID data type `0.TYPE/URL` does not fit semantically for Magnet Links, because URL is a subset of URI [22]. As a result, an own data type `MAGNET` needs to be registered at a Handle PID server that should be capable of holding Magnet Links. To retrieve data from the PID via location-independent technology, a Magnet Link can be placed into the Handle. As Magnet Links fit into the UTF-8 encoding of Handle values, they can be placed without any further encoding.

For Instance, a valid DHT-enabled Magnet Link containing BitTorrent information for retrieving a file can be generated and stored in the MAGNET field of a Handle PID. Additionally, a Magnet Link-wrapped NDN Data Name can be placed into the MAGNET using URL escaping according to NDN name specifications [23].

C. Data Access Through PID with Persistent Resolution Targets

Handle PIDs that are equipped with a Magnet Link can be resolved like any other PIDs using the native Handle protocol. This can be done either by using that protocol on top of TCP or UDP, or via a HTTP-based proxy that answers resolution requests as a Web service. The *resolving* process for PIDs with persistent resolution targets works similar to the resolution process of location-based PIDs regarding the initial steps done within the Handle infrastructure. Hence, Figures 2 and 6 share the first initial step ❶, where the PID is resolved by the Handle infrastructure. In this step the Global Handle Registry (GHR) determines the Local Handle System (LHS), which is responsible for a specific local sub-namespace (Handle prefix). Then, the LHS looks up the requested PID in its database and returns the requested values to the client. For resolving using location-based data access, the value with the type 0.TYPE/URL is returned from the database and for resolving a with a persistent target location independent data access information in form of MAGNET is returned.

The new data access chain depicted in Figure 6 is different from the location-based data access using the target URL shown in Figure 2. With the Magnet Link of the PID acquired through the resolution process in step ❷ the data access is now handled by the overlay network in BitTorrent, or by the NDN network (step ❸). The PID resolving process is then a single redirection that leads to a starting download right after resolving, instead of multiple redirections using an entire chain of services for data access. These connections rely on peer connections based on IP-addresses for BitTorrent and node connections for NDN. Thus, the number of steps is reduced to three; also fewer layers and infrastructure are required for accessing the data. No central infrastructure is involved and the entire chain of location-based infrastructure is not needed anymore. The only requirements for data access is a running PID infrastructure and employing BitTorrent or NDN software for sharing the data online.

D. Creating and Resolving PIDs with Persistent Resolution Targets in Web Environments

For a convenient and smooth resolution of PIDs in the context of the World Wide Web, querying and resolving of Handle PIDs is realized by using a proxy service that accepts requests in HTTP(S) and resolves and maintains PID with the native protocol [21]. The Handle System maintainer, the Corporation for National Research Initiatives (CNRI), provides a Handle Proxy Servlet that offers HTTP-based resolving. The official Handle Proxy needs a small extension to resolve PIDs smoothly with HTTP into location-independent access information. This is done by changing the resolving mechanism

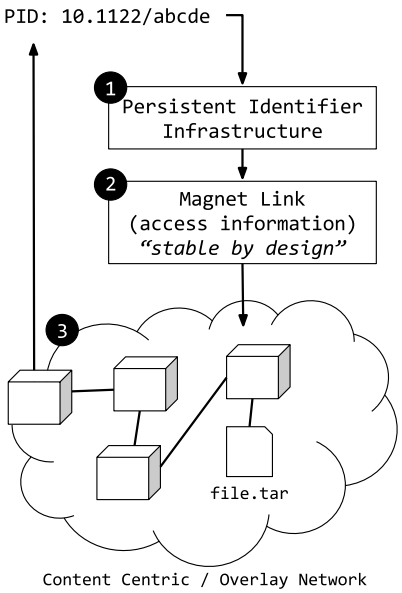


Figure 6: Location independent access through PID relies on stable content-based access information.

from 0.TYPE/URL to MAGNET for the default value. However, non-modified Handle Proxies also work, when explicitly querying for MAGNET or using URL rewriting appending query parameters on the HTTP request.

For offering a Web-based creation of PIDs that features a Magnet Link as persistent resolution target, a Web service is a reasonable choice. For every system that is offered for location-independent access technology, the Web service consumes either the native access information and encodes them into a Magnet Link or directly Magnet Links. Then, the Web service is registering a Handle PID at the LHS using the native Handle protocol and updates the Handle with a MAGNET containing the Magnet Link. In case of BitTorrent, the Web service consumes torrent files that contain along with the checksum and the name of the torrent also the infohash. These information are parsed and embedded into the Magnet Links as persistent PID resolution target using the current Magnet URI scheme description. For NDN data access, the Web service consumes the data name and the SHA-256 checksum of the data set. Furthermore, the Web service allows attaching NDN verification information to PID by extending the Magnet Link in the PID containing the cryptographic signature, as well as the data name to obtain the public certificate of the owner. The NDN access information are encoded into Magnet Link using our proposed Magnet URI scheme (c.f Subsection VI-A) in order to be embedded into the PID.

E. Data Access Through PID with Persistent Resolution Targets using a Web browser

Although, the Handle System can remain unmodified and the Handle proxy is only subject of optional minor modifications, the automatic resolution on HTTP clients provides solvable challenges. To provide a smooth resolution, a HTTP client like

a Web browser needs to support multiple protocols through asking a protocol handler that determines the behaviour for resources outside HTTP sphere, like `mailto:` resources that are forwarded to the E-Mail program. For `magnet:` resources, the Magnet Link protocol handler is invoked [24]. Based on information stored in the Magnet Link value `Exact Topic (xt)`, the Magnet Link Handler selects the appropriate application and passes all information necessary for the download. Then the application is doing the heavy data download via the suggested access technology stored in the Magnet Link.

For invoking the process automatically in the Web browser, following steps are applied. When resolving HTTP-related URLs, the proxy-based resolver responds with HTTP status "`303 – See Other`". According to the HTTP standard RFC 7231, a `303` response to a `GET` request indicates that the server does have a representation of the target resource which can be transferred over HTTP [25]. In the case of normal location-based forwarding, the target `URL` field of the PID is resolved into a target server target URL, as the proxy server does not possess the data. In the case of PID holding location-independent access information, the `MAGNET` field is resolved into a forwarding to the overlay network of NDN name space. This HTTP-forwarding indicates that the Handle proxy server does not posses the data and it is not reachable via HTTP. Hence, the use of Magnet URIs in PID for HTTP-based resolution is in line with the HTTP standard and especially supports it with embedding additional information into the Magnet Link for a self descriptive data access.

VII. Implementation

A. Server Side

For verifying our approach, we set up an entire stack of software component for verification. On the server side, the stack consists of an unmodified Handle System (c.f. VII-A1) and a custom Web service called *PID-Burner* (c.f. VII-A2), which is able to create, update and resolve Magnet Link enabled PIDs based on our approach.

1) Local Handle System: The LHS that hosts the Handle PIDs under a specific prefix does not require any modifications in the source code to run our approach for storing and serving PIDs with Magnet Links. By default, the Handle System supports a list of preconfigured data types that are available as standard type set in every Handle System of a specific version. The list of pre-configures data types contains types for realizing typical PID scenarios with `EMAIL` and `URL` for location-based access. But also special scenarios like target URL forwarding based on users' geolocations. It should also contain a `URN` data type according to the Handle System documentation [26], but it – unfortunately – is not part of the preconfigured data types in the most recent version 8.1.0 [27].

As none of these preconfigured data types are suitable or working for storing Magnet Links, we use the well-designed extensible type system of the Handle System architecture to register a new data type `MAGNET`. By this, we only add a configuration item to the LHS that has no impact on the existing type system and runs on Handle legacy systems, too.

2) PID-Burner - Creating, Maintaining and Resolving of PIDs with Persistent Resolution Targets: The PID-Burner Web service allows creating, updating and resolving PIDs with Magnet Links embedded as persistent resolution targets. It is implemented from scratch, but incorporates libraries and frameworks for PID management, as well as BitTorrent libraries and initially created libraries for Handling NDN access information and processing Magnet Links with the extensions proposed. The PID-Burner engine is implemented in Python using the *Bottle Web framework* from creating a Representational State Transfer (REST) interface [28]. Besides the REST interface it offers a JavaScript-based user interface for the Web browser. As back-end for interacting with the Handle PID service the EPIC-API v2 Web service from European Persistent Identifier Consortium (EPIC) is incorporated to create and update PIDs [29]. For processing BitTorrent access information contained in torrent files, *libtorrent* Python bindings are used for extracting the necessary access information like the infohash, the file name and checksum [30]. NDN access information processing is done with a custom library, as well as the generation of Magnet Links.

For creating and updating PIDs with Magnet Links, the user can upload a torrent file that contains the BitTorrent access information. These torrent files can be created from original files in BitTorrent programs like Transmission [31]. NDN access information are uploaded as Java Script Object Notation (JSON) data structures and can store the data name and the checksum. The NDN data names has to be determined by the user depending on its NDN network topology. The checksum can be computed using any checksumming tool like OpenSSL. JSON is used for NDN access information encoding due to the lack of standardized access container formats in NDN that are comparable to torrent container files. The optional cryptographic verification information are attachable to the PID using NDN access information containing the cryptographic signature of the data and the NDN data name to retrieve the X.509 certificate of the data signer.

For resolving Magnet Links enabled PIDs with HTTP, the Web service implements a resolution functionality that is almost identical to the original Handle HTTP proxy by CNRI [32]. As described in Subsection VI-C, the resolution process does not depend on target URLs, but rather uses the PID value stored in the `MAGENT` data field of the PID. Hence, if a Web client asks the PID-Burner service for resolving, the Magnet Link with the access information is returned as HTTP status `303 – See Other` for existing PIDs. If the PID does not contain a Magnet Link, the resolution is done against the URL value of the PID and PID-Burner behaves identical to the Handle HTTP proxy service. The behaviour allows a maximum on compatibility towards the original Handle System also on HTTP resolution.

B. Client Side

End-users who are interested in using Magnet Link-enabled PIDs for data access need client software that is able to

process access information for location-independent access. For BitTorrent-based access a client software is needed that is able to process Magnet Links. To simulate end-user environment we are using a Gnome 3.16.2 Desktop on Fedora 22 together with Chromium 47.0.2526.106 as Web browser running on Intel i5-2400 with 8GB RAM. As BitTorrent Software, we use an unmodified version of Transmission 2.92 [31] and use public BitTorrent infrastructure available to every Internet user (DHT bootstrap servers) for file download.

For NDN access, we provide an own Magnet Link adapter that process the NDN access information and passes them to the NDN download tools. This NDN Magnet Link adapter has been implemented in Python and is necessary to parse NDN Magnet Links the based on our proposes schema extension. For NDN data hosting and downloading, we use the experimental NDN Repo NG tool set that consists of NDN server and download applications for exchanging data over a NDN network [33]. The Repo NG tool set is running in a private testbed at Gesellschaft für wissenschaftliche Datenverarbeitung Göttingen mbH (GWDG) that consists of six NDN nodes.

VIII. Evaluation and Discussion

All approaches related to PID systems have to provide interoperability at its highest degree to comply with the slow changing momentum of the Handle Infrastructure.

First, the Handle System is a very large distributed infrastructure with shared responsibilities. The services consist of 1 000 servers in 75 countries, which are operated by hundreds of organisations. It currently holds over >100 million PIDs, owned by over 12 thousand registrants in 2015 [34] [35]. Hence, approaches that demand a fundamental change in the system have the challenges of convincing a large community. In contrast, our approach presented in the paper operates on-top of the infrastructure and has no impact on existing PID infrastructure. If LHS operators are interested in implementing PIDs with persistent resolution targets they just have to add a service that is able to resolve, maintain and update `Magnet` Links within PIDs as described in VII-A2.

Second, for data repository owners that use PID for registering their data and HTTP-based file distribution, our approach opens up new perspectives on data dissemination. With our approach they can combine the advantages of location-independent access, peer-to-peer networks and PID into a single concept. For this, they have to create access information for their existing files and provide a location-independent upload point such as a BitTorrent Upload server.

Although our concept offers many advantages, we have to investigate its performance. The evaluation has to be split into two parts; the first is the evaluation of the PID access. This is done by comparing the distribution of string lengths in PID target URLs against existing Magnet Link resources, checking whether Magnet Links will increase the size of PIDs. If Magnet Links increase the size, resolution performance will decrease. This can be explained with higher data transmission volumes and larger data sets that are to be handled by software stacks.

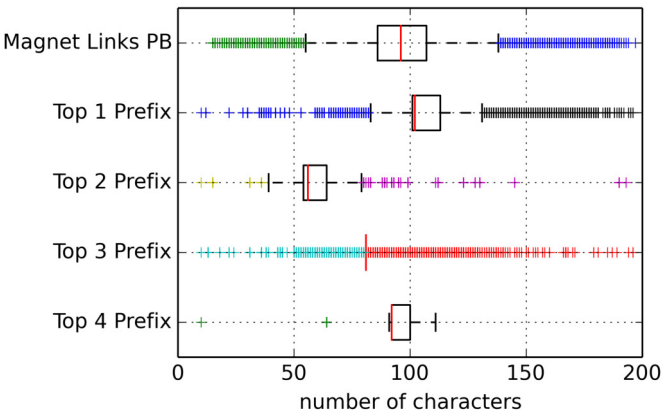


Figure 7: Distribution of string length for Pirate Bay Magnet Links and PID target URLs for frequently resolved Handle prefixes (95% data shown in plot, x-axis limited at 200 chars)

In Figure 7, the string length distribution for BitTorrent Magnet URLs aggregated from the *Pirate Bay* Web site is plotted as box-plot in row one. The Pirate Bay data set (row one) is chosen as a benchmark data set, as it forms one of the largest data accumulation for heterogeneous files exchanged by BitTorrent. For the string length analysis of the Pirate Bay data sets the tracker information have to be removed in order to provide clean analysis; furthermore, the tracker information are not necessary to perform a download using DHT techniques in BitTorrent. It consists of 1 643 194 Magnet Links in total that consist of BitTorrent access information in the form of infohashes and file names [20].

In comparison to the access information in Magnet Links, we now compare the string lengths of PIDs (row two to five). For this we use real-world data from network users who resolved PIDs at Handle Servers hosted at GWDG. The data set for PID resolving data consists of 1.294.668 target URLs in total for a time span between June 2014 and August 2014. The five Handle prefixes with the top-most resolution at the time span have been selected.

The box plots in Figure 7 show that real-world Magnet Link collections share a comparable string length distribution with Handle PID target URLs. To investigate the relation between PID size and the PID resolution performance, we measured the resolutions performance with PIDs of a defined size (c.f. Figure 8). The measurements were done at the LHS, hosting the Handle Prefix 11022 at GWDG with suppressed caching support to measure raw resolution times. It can be observed that the resolution time is slightly increasing for a number of milliseconds for extreme PID sizes with 32 768 characters. With the average PID size derived from the Pirate Bay data set of 97 characters (c.f. with the 2^7 bar in Figure 8), the impact of Magnet Link usage in PID is not perceivable to users. As a result, the PID replication and resolution can be expected similar to the existing target URL-based approach. Embedding Magnet Links in PIDs has no significant impact on the size

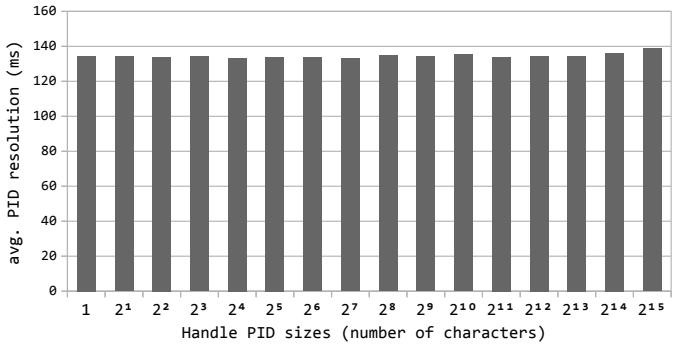


Figure 8: Average resolution time of Handle PIDs with different target URL lengths measured at the GWDG LHS for the Handle Prefix 11022.

of PIDs and underlines the practical usability of the concepts presented in this paper. Hence, Magnet Link enabled PIDs will not resolved into location-independent targets slower than current state-of-the-art Handle PIDs. The PID resolution time t_r can be assumed to be identical.

The second part of the evaluation is the data download speed provided from the location-independent data access. For BitTorrent and NDN the time for bootstrapping t_b the node and initiate a data download from the swarm is time intensive. Bootstrapping includes in the case of BitTorrent joining the DHT. The latter is very fast, although state-of-the-art DHT joining is accelerated through hard-coded bootstrapping servers. For NDN bootstrapping includes learning the node environments and the name routes. Different NDN bootstrapping algorithms are subject of current research [36].

After successful bootstrapping the advantages can speed up data transfer. If a peer is found that offer data access the download speed is at least comparable to existing location-based access that uses HTTP for download. If more than one peer is found, the bandwidth can be utilized to provide simultaneous data transfers, too. In this case the data volume v is split into n parts and the longest transfer $max()$ time of chunk is setting the overall transmission time as most time consuming partial transfer. Hence, the transfer duration d_{tr} of the first transfer through PID can be estimated as

$$d_{tr} = t_r + t_b + max\left(\frac{v_n}{v/sec}\right) \quad (1)$$

This parallelization results in higher download bandwidth and short transmission duration. Thus, for large download volumes the data access starting from PID resolution to download completion is faster. For small download volumes the completion time is longer in comparison to traditional data access through PID. Traditional location-based approaches using HTTP with no multi-source download capabilities have a transfer duration of

$$d_{tr} = t_r + \sum_{n=0}^{N} \frac{v_n}{v/sec} \quad (2)$$

where the duration is the sum of all partial volumes that are downloaded in serial.

Besides the performance aspects, it is observable that the PID usage of the PID infrastructure usage has an impact on the string length distribution of the target URLs. The top 4 prefix has a dense character distribution that is caused by the fixed pattern of the target URL, where only parts of the URL are varying. This is caused by the repository software that is managing the PIDs and uses IDs with similar length for ever data set. In contrast, the top four prefix shows a sparse distribution of target URL string length caused by almost non-systematic target PIDs. This distribution can be found by PID System operator that offer PID services to a large group of institutions like EPIC [29].

IX. Future Work

Despite our contribution to the PID efforts in the Handle System challenges provide open research questions for future work. The support for non-URL targets in HTTP-based Handle PID resolution could be moved into the existing Handle stack. By this, resources can be directly linked in the typical workflows that end-user facilitate in their Web browsers. Thus, location-independent access through PIDs in the Handle system could work with click, as simple as opening a PDF file. Fortunately, the Handle PID is very well designed and implemented by CNRI, and the source code is available.

A number of challenges arises from the usage of Magnet Links. The Magnet URI scheme originates in the file sharing community and has evolved in the past decade. Despite being a community effort, it is widely used by the most frequented file sharing search engines. For usage in the research data community, Magnet URI community contributions and drafts have to be collected and a standardisation effort needs to be started, e.g., as an initial IETF draft. The relation between Magnet URI and URN will facilitate the standardisation efforts and help to improve the reputation of Magnet Links.

Moreover, we propose using Magnet Links in the NDN community for encoding of full access information. Magnet Links could be one appropriate container format for transmitting NDN access information, which is closer to the original idea of object identification, based on URN, rather than location identification done with current NDN concepts based on URL. Similar approaches have been provided for other content-centric network like Open NetINF and proposed at IETF [37].

X. Conclusion

The integration of location independent data access in persistent identifiers is feasible without major modification of PID infrastructure. The Magnet URI scheme is suitable container for storing application independent access information inside Handle PIDs although it currently lacks IETF standardisation. As illustrated in our evaluation, their usage has no major implication on PID System operation and usage. Employing Magnet Link enables the creation of maintenance free PIDs, which do not require target URL adjustments and thus reduce the residual efforts on data repository owner sides. With the

support of overlay network usage and NDN data access, *better* data dissemination is achievable with augmented resilience through multi-peer data hosting.

ACKNOWLEDGMENTS

We like to thank Sven Bingert from the European Persistent Identifier Consortium for his support on the Handle evaluation infrastructure used for this paper. We acknowledge research funding by Deutsche Forschungsgemeinschaft (DFG) under grant SFB 963/2 “Astrophysical Flow Instabilities and Turbulence”, projects INF.

REFERENCES

[1] N. Paskin, “Digital Object Identifier (DOI) System,” in *Encyclopedia of Library and Information Sciences*, 3rd ed. Boca Raton, FL: CRC Press, 2011, pp. 1586–1592.

[2] B. Ahlgren, C. Dannewitz, C. Imbrenda, D. Kutscher, and B. Ohlman, “A survey of information-centric networking,” *IEEE Comm. Magazine*, vol. 50, no. 7, pp. 26–36, 2012. doi: 10.1109/MCOM.2012.6231276

[3] E. Van der Sar, “The Pirate Bay Tracker Shuts Down for Good,” Nov. 2009. [Online]. Available: https://torrentfreak.com/the-pirate-bay-tracker-shuts-down-for-good-091117/

[4] K. Sollins, “Pervasive persistent identification for Information centric networking,” in *Proc. of the Second Edition of the ICN Workshop on Information-centric Networking*. Helsinki, Finland: ACM, 2012. doi: 10.1145/2342488.2342490 pp. 1–6.

[5] T. Koponen, M. Chawla, B.-G. Chun, A. Ermolinskiy, K. H. Kim, S. Shenker, and I. Stoica, “A Data-oriented (and Beyond) Network Architecture,” in *Proceedings of the 2007 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications*, ser. SIGCOMM ’07. New York, NY, USA: ACM, 2007. doi: 10.1145/1282380.1282402 pp. 181–192.

[6] C. Dannewitz, M. Herlich, and H. Karl, “OpenNetInf - prototyping an information-centric Network Architecture,” in *Proceedings of the 37th IEEE Conference on Local Computer Networks Workshops 2012*, Clearwater, USA, Oct. 2012. doi: 10.1109/LCNW.2012.6424044 pp. 1061–1069.

[7] N. Fotiou, D. Trossen, and G. C. Polyzos, “Illustrating a publish-subscribe Internet architecture,” *Telecommunication Systems*, vol. 51, no. 4, pp. 233–245, Dec. 2012. doi: 10.1007/s11235-011-9432-5

[8] C. Dannewitz, J. Golic, B. Ohlman, and B. Ahlgren, “Secure Naming for a Network of Information,” in *Proc. of IEEE Conference on Computer Communications INFOCOM*. San Diego, USA: IEEE, 2010. doi: 10.1109/INFCOMW.2010.5466661 pp. 1–6.

[9] S. Haun and A. Nürnberger, “Towards Persistent Identification of Resources in Personal Information Management,” in *Proc. of the 3rd International Workshop on Semantic Digital Archives (SDA 2013)*, vol. 1091. Valetta, Malta: CEUR Workshop Proc., Sep. 2013, pp. 73–80.

[10] H.-W. Hilse and J. Kothe, *Implementing persistent identifiers: overview of concepts, guidelines and recommendations*. London: CERL, 2006.

[11] T. Cruse, “General Assembly 2016, moving DataCite forward,” 2016. [Online]. Available: https://blog.datacite.org/general-assembly-2016/

[12] “DataCite Metadata Stats,” Apr. 2016. [Online]. Available: http://stats.datacite.org/

[13] M. Fenner, “Digging into Metadata using R,” Aug. 2015. [Online]. Available: https://blog.datacite.org/digging-into-data-using-r/

[14] B. Cohen, “The BitTorrent Protocol Specification - BEP 3,” Oct. 2013. [Online]. Available: http://www.bittorrent.org/beps/bep_0003.html

[15] A. Loewenstern and A. Norberg, “The BitTorrent Protocol Specification - BEP 5,” Mar. 2013. [Online]. Available: http://www.bittorrent.org/beps/bep_0005.html

[16] P. Maymounkov and D. Mazières, “Kademlia: A Peer-to-Peer Information System Based on the XOR Metric,” in *Peer-to-Peer Systems*. Berlin, Heidelberg: Springer, 2002, vol. 2429, pp. 53–65.

[17] V. Jacobson, D. K. Smetters, J. D. Thornton, M. F. Plass, N. H. Briggs, and R. L. Braynard, “Networking named content,” in *Proceedings of the 5th international conference on Emerging networking experiments and technologies*. Rome, Italy: ACM Press, Dec. 2009. doi: 10.1145/1658939.1658941 p. 1.

[18] G. Mohr, “Magnet URI - Draft Tech Overview/Spec,” Jun. 2002. [Online]. Available: http://magnet-uri.sourceforge.net/magnet-draft-overview.txt

[19] K. Sollins and L. Masinter, “RFC 1737 - Functional Requirements for Uniform Resource Names,” Dec. 1994. [Online]. Available: https://tools.ietf.org/html/rfc1737

[20] E. Van der Sar, “Download a Copy of The Pirate Bay, It’s Only 90 MB,” Feb. 2012. [Online]. Available: https://torrentfreak.com/download-a-copy-of-the-pirate-bay-its-only-90-mb-120209/

[21] S. X. Sun, S. Reilly, and B. Boesch, “RFC 3650 - Handle System Overview,” 2003. [Online]. Available: https://tools.ietf.org/html/rfc3650

[22] T. Berners-Lee, R. Fielding, and L. Masinter, “RFC 3986 - Uniform Resource Identifier (URI): Generic Syntax,” 2005. [Online]. Available: https://tools.ietf.org/html/rfc3986

[23] Y. Yu, A. Afanasyev, Z. Zhu, and L. Zhang, “NDN Technical Memo: Naming Conventions - NDN, Technical Report NDN-0023, Revision 1,” Jul. 2014. [Online]. Available: http://named-data.net/wp-content/uploads/2014/08/ndn-tr-22-ndn-memo-naming-conventions.pdf

[24] D. Thaler, T. Hansen, and T. Hardie, “RFC 7595 - Guidelines and Registration Procedures for URI Schemes,” Jun. 2015. [Online]. Available: https://tools.ietf.org/html/rfc7595

[25] R. Fielding and J. Reschke, “RFC 7231 - Hypertext Transfer Protocol (HTTP/1.1): Semantics and Content,” Jun. 2014. [Online]. Available: http://tools.ietf.org/html/rfc7231#section-6.4.4

[26] CNRI, “4.9 Handle Value Line Format,” in *HANDLE.NET (version 8.1) Technical Manual*, Nov. 2015, pp. 28–29. [Online]. Available: https://hdl.handle.net/20.1000/105

[27] ——, “Handle.Net software (HN_v8.1),” 2015. [Online]. Available: http://www.handle.net/download_hnr.html

[28] M. Hellkamp, “Bottle: Python Web Framework,” Feb. 2016. [Online]. Available: http://bottlepy.org/docs/0.12/

[29] European Persistent Identifier Consortium, “pidconsortium/EPIC-API-v2,” Mar. 2016. [Online]. Available: https://github.com/pidconsortium/EPIC-API-v2

[30] A. Norberg, “libtorrent python binding,” 2015. [Online]. Available: http://www.rasterbar.com/products/libtorrent/python_binding.html

[31] Transmission Project, “Transmission,” Mar. 2016. [Online]. Available: https://www.transmissionbt.com/

[32] CNRI, “Handle.Net Registry,” 2015. [Online]. Available: https://www.handle.net/proxy_servlet.html

[33] A. Afanasyev, S. Chen, W. Shang, and J. Shi, “repo-ng: Next generation of NDN repository,” Nov. 2015. [Online]. Available: https://github.com/named-data/repo-ng

[34] CNRI, “HDL® Identifier and Resolution Services,” Oct. 2015. [Online]. Available: http://www.handle.net/factsheet.html

[35] International DOI Foundation, “DOI News - September 2014,” Sep. 2014. [Online]. Available: http://www.doi.org/news/DOI_News_Sep14.pdf

[36] A. K. M. M. Hoque, S. O. Amin, A. Alyyan, B. Zhang, L. Zhang, and L. Wang, “NLSR: Named-data Link State Routing Protocol,” in *Proc. of the 3rd ACM SIGCOMM Workshop on Information-centric Networking ICN*. New York, USA: ACM, 2013. doi: 10.1145/2491224.2491231 pp. 15–20.

[37] S. Farrell, C. Dannewitz, P. Hallam-Baker, D. Kutscher, and B. Ohlman, “RFC 6920 - Naming Things with Hashes,” Apr. 2014. [Online].

5th International Conference on Wireless Sensor Networks

TOPICS

DEVELOPMENT of sensor nodes and networks
Sensor Circuits and Sensor devices – HW

- Applications and Programming of Sensor Network – SW
- Architectures, Protocols and Algorithms of Sensor Network
- Modeling and Simulation of WSN behavior
- Operating systems

Problems dealt in the process of WSN development

- Distributed data processing
- Communication/Standardization of communication protocols
- Time synchronization of sensor network components
- Distribution and auto-localization of sensor network components
- WSN life-time/energy requirements/energy harvesting
- Reliability, Services, QoS and Fault Tolerance in Sensor Networks
- Security and Monitoring of Sensor Networks
- Legal and ethical aspects related to the integration of sensor networks

Applications of WSN

- Military
- Health-care
- Environment monitoring
- Transportation & Infrastructure
- Precision agriculture
- Industry application
- Security systems and Surveillance
- Home automation
- Entertainment – integration of WSN into the social networks
- Other interesting applications

EVENT CHAIRS

- **Hodoň, Michal,** University of Žilina, Slovakia
- **Kapitulík, Ján,** University of Žilina, Slovakia
- **Micek, Juraj,** University of Žilina, Slovakia
- **Ševčík, Peter,** University of Žilina, Slovakia

PROGRAM COMMITTEE

- **Al-Anbuky, Adnan,** Auckland University of Technology, New Zealand
- **Baranov, Alexander,** Russian State University of Aviation Technology, Russia
- **Brida, Peter,** University of Zilina, Slovakia
- **Dadarlat, Vasile-Teodor,** Univiversita Tehnica Cluj-Napoca, Romania
- **Diviš, Zdenek,** VŠB-TU Ostrava, Czech Republic
- **Elmahdy, Hesham N.,** Cairo University, Egypt
- **Fortino, Giancarlo,** Università della Calabria
- **Fouchal, Hacene,** University of Reims Champagne-Ardenne, France
- **Furtak, Janusz,** Military University of Technology, Faculty of Cybernetics, Poland, Poland
- **Giusti, Alessandro,** CyRIC - Cyprus Research and Innovation Center, Cyprus
- **Grzenda, Maciej,** Orange Labs Poland and Warsaw University of Technology, Poland
- **Gu, Yu,** National Institute of Informatics, Japan
- **Hudik, Martin,** University of Zilina
- **Husár, Peter,** Technische Universität Ilmenau, Germany
- **Jin, Jiong,** Swinburne University of Technology, Australia
- **Jurecka, Matus,** University of Žilina, Slovakia
- **Kafetzoglou, Stella,** National Technical University of Athens, Greece
- **Karastoyanov, Dimitar,** Bulgarian Academy of Sciences, Bulgaria
- **Karpiš, Ondrej,** University of Žilina, Slovakia
- **Kochláň, Michal,** University of Žilina, Slovakia
- **Laqua, Daniel,** Technische Universität Ilmenau, Germany
- **Milanová, Jana,** University of Žilina, Slovakia
- **Monov, Vladimir V.,** Bulgarian Academy of Sciences, Bulgaria
- **Ohashi, Masayoshi,** Advanced Telecomunications Research Institute International / Fukuoka University, Japan
- **Papaj, Jan,** Technical university of Košice, Slovakia
- **Ramadan, Rabie,** Cairo University, Egypt
- **Scholz, Bernhard,** The University of Sydney, Australia
- **Shaaban, Eman,** Ain-Shams university, Egypt
- **Shu, Lei,** Guangdong University of Petrochemical Technology, China
- **Smirnov, Alexander,** Linux-WSN, Linux Based Wireless Sensor Networks, Russia
- **Staub, Thomas,** Data Fusion Research Center (DFRC) AG, Switzerland
- **Teslyuk, Vasyl,** Lviv Polytechnic National University, Ukraine
- **Wang, Zhonglei,** Karlsruhe Intitute of Technology, Germany
- **Xiao, Yang,** The University of Alabama, United States

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1043–1047
DOI: 10.15439/2016F511
ACSIS, Vol. 8. ISSN 2300-5963

Calculating the Speed of Vehicles Using Wireless Sensor Networks

Omar Alfandi*, Arne Bochem, Mehdi Akbari Gurabi,
Alberto Rivera Díaz, Md.Istiak Mehedi, Dieter Hogrefe
Institute of Computer Science, University of Göttingen, Germany
*College of Technological Innovations, Zayed Universtiy, United Arab Emirates
Email: {alfandi,bochem,hogrefe}@cs.uni-goettingen.de,
{mehdi.akbarigurabi, alberto.riveradiaz, mdistiak.mehedi}@stud.uni-goettingen.de

***Abstract*—Speed measurement is an important issue for some types of Wireless Sensor Networks (WSN), especially for Vehicular Ad-hoc Networks (VANETs). However, calculating this value is error-prone and costly. This report intends to demonstrate the calculation of speed of an object without the use of any additional devices or sensor boards, only using Received Signal Strength Indication (RSSI) for localization of the vehicles and time calculation using synchronization. We implemented these methods in actual IRIS motes, and tested them. The results show that, while not perfectly accurate, our method proved to be reliable and close to the real speed. In addition, the results do not have any linear correlation in divergence of real speed and calculated speed, which means the system avoids systematic errors.**

I. Introduction

NOWADAYS, Wireless Sensor Networks form a big part of our networks and their utilization has seen a rapid increase. This growth is effected due to newly developed wireless multimedia services and attributions such as data, voice and video [1]. In this way, WSN have started to play a significant role in our everyday life. They have a wide range of applications, from monitoring the environment and the human body to industrial and military applications which are slowly becoming ubiquitous [2].

Mobility is a major issue for several WSN applications. An example of a useful application is in the branch of VANETs, a subclass of Mobile Ad-hoc Networks (MANETs), which relies on vehicles to provide functionality. Their approach is of great importance to Intelligent Transportation Systems (ITS). The main differences between normal MANETs and VANETs are the mobility constraints, high mobility, and the driver's behaviour all of which are found in the latter [3]. VANETs provide a variety of applications that ranging from safety solutions, collision detection and crash avoidance to Internet access and multimedia [4]. These applications are expected to be deployed on a larger scale in the future due to their usefulness, for example in providing safety to drivers or enhancing inter-vehicle communication [1].

VANETs make use of several traffic flow parameters to perform the monitoring, such as speed of the vehicle. Several methods have been implemented to calculate this parameter, for example: Global Positioning System (GPS)localization[5], the Laser Doppler Techniques[6], or Received-Signal-Strength(RSS)[5]. Most of these methods need additional devices for their calculation such as directional antenna [7] or GPS module that can increase the cost of deployment.

In this report, a simple method using RSSI according to characteristics and attributes of VANETs mobility model was implemented. We are modelling a street by using some IRIS motes as fixed nodes for gathering data as the infrastructure of the network, and send information of cars to our base station. The base station calculates the velocity of cars by means of this data.

The paper is organized in seven sections: Section 2 discusses related works. Section 3 explains the methodology. In section 4, various advantages and disadvantages of the chosen approach are discussed. Section 5 explains further challenges. Results of experimental measurements are provided in section 6. In section 7 conclusions are drawn.

II. Related works

There exist several ways to measure the speed of a moving object. This is of particular interest to law-enforcement for speed limits of vehicles on the roads. The most popular and widespread method is through the use of radio signals.

In this section previous approaches and works on this topic are highlighted.

A. Doppler Effect

Doppler radar speed measuring unit: Radar guns or speed guns are commonly used in the enforcement of speed limits on highways. They employ the doppler effect to measure the speed of moving objects by detecting a change in frequency of the returned radar signal. From the difference in frequency, the speed of the object can be calculated.

There are two types of speed guns[6]: Stationary and moving radar. For the stationary radar, a signal with a frequency equal to this difference is created by mixing the received radio signal with a little of the transmitted signal. These two radio signals are mixed to create a "beat" signal (called a heterodyne) and an electrical circuit that measures this frequency using a digital counter and displays the number on a digital display as the object's speed. In moving radars, a gun receives reflected signals from both the target vehicle and stationary background objects such as the road surface, nearby road signs, guard rails and street light poles.

978-83-60810-90-3/$25.00

Although very useful, this method has several drawbacks. In order to function correctly, radio waves must leave the gun in a narrow beam that do not spread out much to avoid receiving a false return from nearby objects or vehicles. Another disadvantage, is that user training and certification are required so that a radar operator can use the equipment effectively. They also do not work very well in traffic and significant vehicle separation is needed to ensure an accurate measurement.

B. GPS

GPS tracking units: The Global Positioning System is another technology that can be utilized to obtain the speed of a vehicle. It provides time and location information, regardless of the weather, or the location on Earth. A GPS tracking unit is a device that uses the Global Positioning System to determine the precise location of a vehicle, person, or other objects; to which it is attached and to record the position of the asset at regular intervals. The recorded location data can be stored within the tracking unit, or it may be transmitted to a central location data base, or internet-connected computer, using a cellular (GPRS or SMS), radio, or satellite modem embedded in the unit.

One type of GPS tracking unit is the data pusher; that is used for asset tracking, personal tracking and Vehicle tracking systems. Also known as a GPS beacon, this kind of device sends the position of the device as well as other information to a determined server to be stored and analysed. By knowing the exact location of a moving object, its speed can be easily calculated, thus making GPS a tool for obtaining the said value.

C. Other methods

LiDAR speed gun: LiDAR is a remote sensing technology that measures distance by illuminating a target with a laser and analysing the reflected light. The term is a portmanteau of ”light” and ”radar.” The police uses LiDAR[6] speed guns for speed limit enforcement, which in turn use LiDAR technology to detect the speed of a vehicle. Unlike radar speed guns, which rely on doppler shifts to measure the speed of a vehicle, these devices allow a police officer to measure the speed of an individual vehicle within a stream of traffic.

RSSI: RSS is defined as the voltage measured by a receiver's received signal strength indicator (RSSI) circuit. Often, RSS is equivalently reported as measured power, i.e., the squared magnitude of the signal strength. We can consider the RSS of acoustic, RF, or other signals. Wireless sensors communicate with neighbouring sensors, so the RSS of RF signals can be measured by each receiver during normal data communication without presenting additional bandwidth or energy requirements. RSS measurements are relatively inexpensive and simple to implement in hardware. Some approaches used for localization are Time of Arrival (ToA) [5] , Angle of Arrival (AoA) [7] and Triangulation Method [8].

The Infra-Red Traffic Logger: The Infra-Red Traffic Logger, more commonly known simply by the acronym TIRTL, is a multi-purpose traffic sensor that can be used as a traffic counter, speed sensor, red light camera sensor, heavy vehicle tracker, over-height vehicle sensor, rail crossing sensor and network management system.

The system consists of a receiver unit and transmitter unit placed on opposite sides of the road perpendicular to the direction of travel. The transmitter sends two cones of infra-red light across the roadway, and the receiver records vehicles as they break and remake these cones [9].

D. Chosen approach

This project aims at demonstrating how to calculate the speed of an object, particularly a vehicle, which moves from one point to another by using only traditional wireless sensor motes without the aid of any specialized sensor boards. In order to obtain said speed, two measures are required: distance and time elapsed. For our project, the distance is a set value, which is given by the distance between two fixed nodes. The time elapsed however, is the main concern. It is calculated by using the RSSI approach to obtain the approximate point in time and space in which the vehicle passes through a fixed point. The higher the received signal, the closer the vehicle is to the fixed point. This allows calculating the difference in times between the nodes, and thus, obtain also the total time elapsed while the vehicle stayed between these fixed points. Implementing other procedures such as GPS to obtain the location of the moving nodes results in high costs and is not energy efficient. Therefore it is intended to demonstrate that a simpler and cheaper approach can be applied with approximately equal effectiveness, rather than a costly and complex one such as GPS or radar based systems.

III. Methodology

In this section, the methodology and experimental setup will be described in details.

A. Lab setup

Sending nodes: These are the main target of the project. A wireless mote is attached to the vehicles or moving objects. It will be constantly sending dummy packets, which also contain the nodes ID to be processed by the fixed nodes. The

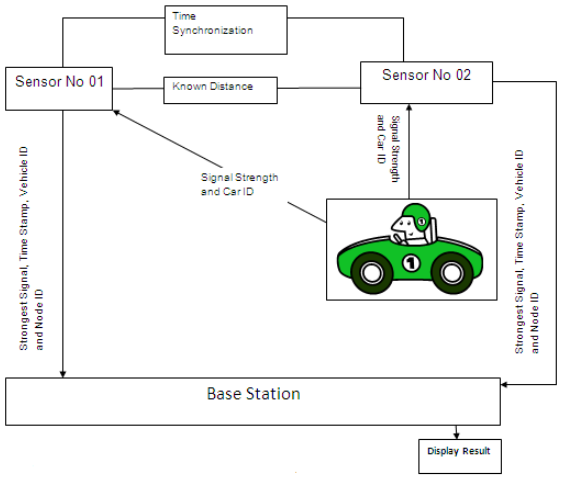


Fig. 1: Overview diagram.

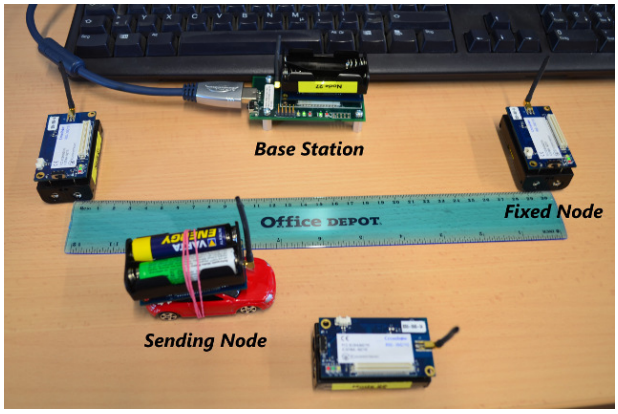


Fig. 2: Lab setup.

ID distinguishes all the different moving nodes. The speed measurement obtained at the end of the process is calculated based on the linear speed of this node assuming it has a constant speed and moves in a linear path. It then sends the packets at a high speed in order to achieve a higher accuracy.

Fixed node: Its function is to receive data, i.e. sense the moving motes and receive their ID and send it to the Base Station. In these motes, the RSSI for the packets sent by vehicular nodes is obtained, as well as the time-stamp of the moment in which the packets are received from the moving mote. Only the strongest signals are chosen to be sent to the Base Station together with the ID and time of receipt. These motes are scattered across the lane and have a fixed distance between them. This allows for the calculation of the speed of the vehicle in a portion of the road, between two fixed nodes.

Base station: The base station gathers the data from the fixed points and makes decisions based on the first packet having the signal strength of the moving mote. Using this information, a graphical interface is deployed to calculate the speed of the vehicle by filtering the data received and finally displaying the results.

B. Process

1. Sending packets: The mobile sending motes are the representation of the vehicles in a highway. These nodes are flashed with a program that executes the following functions: First, when the node is booted, the radio capability is enabled. This will allow communication between the node and the fixed points. Afterwards, packets are assembled by obtaining the nodes ID which will be used to identify the vehicle or object among all the different nodes. These packets are then continuously broadcast.

2. Receiving packets: Fixed motes are scattered across the lane or path which the mobile nodes pass by. They represent the second stage of the process and are flashed with a program called RssiBase that performs the following actions:

When a fixed node receives a packet from a sending node, the fixed node starts preparing a new packet with the information obtained from the sending node, that is the nodes ID, and two new values: the time-stamp and the Received Signal Strength Indication (RSSI) from said mote, which are calculated at the time of reciept. This time is synchronized by applying the Flooding Time Synchronization Protocol(FTSP).

The RSSI value of the sending nodes is used to determine when a mobile node gets to the closest point to the fixed node. When this happens, a time-stamp will allow for the calculation of the total time elapsed between fixed nodes, which is one of the values required in the formula of speed. Therefore, only the values with maximum RSSI received must be used. The fixed node calculates this from a set number of packets, which are received and compared to the previous RSSI value to determine if it is higher or not. If the new value received is higher, then it is stored along with its node ID. After a maximum amount of packets received is reached, the values are encapsulated into a packet and broadcast in order to be received by the base station.

3. Gathering data: The base station receives the packets which contain the highest RSSI received at each fixed node, along with the associated ID and timestamp. The base station mote additionally participates in time synchronization by means of the FTSP protocol.

4. Calculating and displaying the results: After gathering the information needed from all the fixed notes, the results are displayed. This is achieved by a Java application running on a PC, to which the base station is attached. It obtains the data collected from the base station and displays it in a graphical user interface. The actions executed by this program are described in the following steps:

A packet is received at the base station. First, any packets that are useless to the application are filtered out. This avoids cases like getting a second reading from the same car and same fixed node, to preserve the order of events, which could confuse calculation.

There is also another filter that checks if all the packets from the participant fixed nodes have been sent and received by the base station. Calculations can only be done when the condition is met.

After passing both filters, the elapsed time is calculated. This value is obtained by subtracting the time-stamp of the last fixed node's packet from the previous fixed node's packet, and so forth. This will result in the total time between the first and last fixed node.

The elapsed time, along with the vehicle's ID are then passed down to another function within the program, which finally calculates the vehicle's speed.

The common formula to calculate velocity is applied inside that function:

$$v = \frac{d}{t}$$

Here, "d" is the distance travelled, and "t" elapsed time. For the distance value, since the locations of fixed nodes are known, the travelled distance is simply found as shown in figure 4. A fixed distance is set before running and compiling the Java program, and is dependent on the distance between fixed nodes.

After obtaining the speed value, the speed and node ID are displayed in the GUI. (Results are given in m/s.) This process

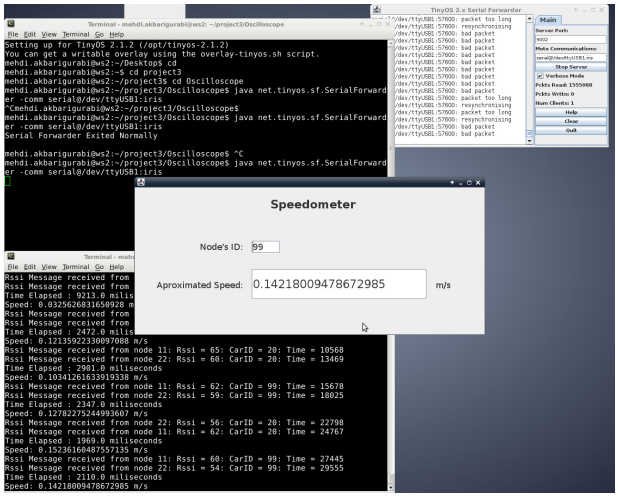


Fig. 3: Results screen.

will repeat indefinitely every time a vehicle passes by the fixed nodes from start to end.

IV. Evaluation of the approach and challenges

In the following, advantages and disadvantages of the described approach are explained.

A. Advantages

Cheaper and easier deployment compared to other approaches that require expensive or complicated tools. By using only the RSSI of the moving nodes, the time elapsed from a fixed point to another can be calculated without the need of special sensor boards or complex schemes to determine the exact moment when the node passes by a point.

By avoiding the use of additional devices, battery consumption is reduced, as opposed to cases where more power-demanding sensors are required. This also translates to lower costs spent in energy and maintenance.

According to our measurements, this approach has reasonably good accuracy.

B. Disadvantages

The approach also has some disadvantages and drawbacks, as explained in the following:

The vehicle or moving object requires a wireless sensor mote. This is the biggest drawback of the project since it will only calculate the speed of an object that has a wireless sensor mote attached to it. If it were to be implemented in a real scenario, vehicle manufacturers would have to be required to implement a similar device which can transmit similar packets. Alternatively, wireless devices of this type could be distributed to vehicle owners. Another possibility that could be investigated would be the use of ubiquitous smart phones as a packet source.

In free space, signal power decays proportional to d^2, where d is the distance between the transmitter and receiver. In real-world channels, multipath signals and shadowing are two major sources of environmental influence on the measured RSS.

Multiple signals with different amplitudes and phases arrive at the receiver, and these signals add constructively or destructively as a function of the frequency, causing frequency-selective fading [5].

These factors may lead to lower accuracy in real world scenarios.

C. Challenges

As of now, this project is very limited due to the magnitude of implementing a real life system where actual cars and fixed points are used. It can be considered a proof of concept to be built upon in future works. As explained in the previous section, several physical constraints appear.

One of the problems with this approach is that the speed of cars could surpass that which our current motes can send at. Packets can get lost easily, not arrive at the correct time or not arrive at all at the fixed nodes, which would lead to incorrect calculations. A solution to this would be deploying more powerful devices with a better transmitting power range.

Another challenge to be tackled is the scalability of the project. For simplicity reasons, we have only implemented a two-component system where one or several moving nodes pass by only two fixed nodes. Real world implementation would require several fixed nodes to have a practical use. This is because the distance between motes has to be within their transmitting range, which limits the distance for a given measurement. Therefore several fixed motes should be employed to be within the range of the network and measure a much greater distance in a highway or path for the purpose of increasing accuracy.

V. Results and discussion

In the following, the testbed is described and experimental measurements are given.

A. Test setup

To prove the effectiveness and accuracy of our project, several tests were performed. A ruler was used to set the exact distance between the fixed nodes. The ground truth is measured using a stop watch. Two IRIS nodes are used as fixed nodes, one IRIS node is attached to a toy car and one acts as a base station.

B. Steps

The system was tested by flashing all the motes with their respective programs, so they could perform their respective duties. The sending mote was attached to a toy car, and proceeded to place both of the fixed nodes at a fixed distance. We tested with four different distances. First, a distance of 30cm, then 60cm, 100cm, and finally 200cm were tested. We tried the system with different random paces: slow (around 0.2 m/s), medium (around 0.5 m/s) and fast (around 1 m/s). We test the system 10 times for each distance and speed. The measurement of time was done with a stopwatch, starting and

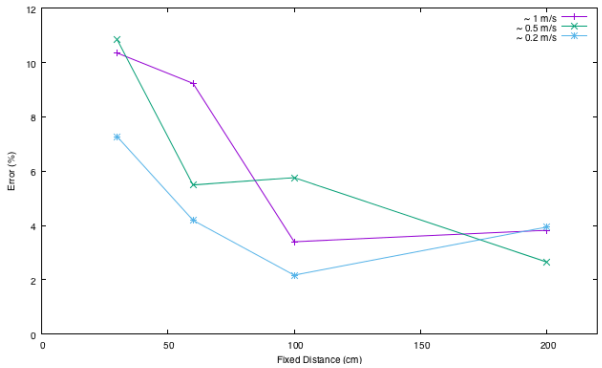


Fig. 4: Accuracy of speed calculations, Error percentage $= 100 * \left| \frac{v_{\text{real}} - v_{\text{calc}}}{v_{\text{real}}} \right|$

ending the timer whenever the sending mote passed next to one fixed node. This allowed us to obtain the total time elapsed. At the end of each test, the manual results obtained from our calculations (The real, accurate speed) with the results displayed on screen given as output from our program.

C. Results

Our results are shown in figure 4. From these results, it is observed that the error percentage between the manually measured values and the values obtained with our application is very small. The most accurate result obtained had an error of 0.94%, whereas the least accurate result had an error of 14.86%. This is out of 120 tests performed which is averaged from 10 tests per each pair of distance and speed.

Another observation is that the set of tests with the shortest distance between fixed nodes (0.30m) has the least accurate results, whereas the set of tests with the longer distances (1m or 2m) has the most accuracy. This can indicate that the closer the fixed nodes are, the less accurate calculations are obtained, possibly due to the packet's arrival times at the Base Station. Shorter paths mean that packets may arrive at close or same time. Additionally, over longer distances, small errors in measurement time will have less influence than equal errors over short distances, because the total amount of time is higher and, overall, outweighs it.

VI. Conclusion

The approach used in this project to calculate the speed of a vehicle using wireless sensor networks proved to be effective and reasonably reliable as shown in the results of the testing. Calculating the time elapsed without the aid of any extra devices or sensor boards is a big advantage, compared to the higher costs of deploying a network of several more power-consuming and expensive devices, as the RSSI of a wireless mote can be obtained without specialized sensor boards. The results show that our method proved to be reliable and close to the real velocity, and no linear correlation in difference of real and measured velocity shows that the speed measurement is free of systematic errors.

In the future, the approach can be extended in various ways, such as including more fixed nodes, using commonly available devices such as smartphones for the mobile node role and scaling everything up to a real world scenarios outside of a controlled testbed.

References

[1] E. Spaho, L. Barolli, G. Mino, F. Xhafa, and V. Kolici, "Vanet simulators: A survey on mobility and routing protocols." *Broadband and Wireless Computing, Communication and Applications (BWCCA), 2011 International Conference on (pp. 1-10). IEEE.*, 2011.

[2] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks." *Communications magazine, IEEE, 40(8), 102-114.*, 2002.

[3] S. Yousefi, M. S. Mousavi, and M. Fathy, "Vehicular ad hoc networks (vanets): challenges and perspectives." *ITS Telecommunications Proceedings, 2006 6th International Conference on (pp. 761-766). IEEE.*, 2006.

[4] G. M. Abdalla, M. A. Abu-Rgheff, and S. M. Senouci, "Current trends in vehicular ad hoc networks." *Ubiquitous Computing and Communication Journal, 1-9.*, 2007.

[5] N. Patwari, J. N. Ash, S. Kyperountas, A. O. Hero, R. L. Moses, and N. S. Correal, "Locating the nodes: cooperative localization in wireless sensor networks," *Signal Processing Magazine, IEEE, 22(4), 54-69.*, 2005.

[6] M. Adnan, N. Zainuddin, N. Sulaiman, and T. Besar, "Vehicle speed measurement technique using various speed detection instrumentation," in *Business Engineering and Industrial Applications Colloquium (BEIAC), 2013 IEEE*, April 2013, pp. 668–672.

[7] J. R. Jiang, C. M. Lin, F. Y. Lin, and S. T. Huang, "Alrd: Aoa localization with rssi differences of directional antennas for wireless sensor networks," *Information Society (i-Society), 2012 International Conference on(pp. 304-309). IEEE.*, 2012.

[8] Z. Li, W. Trappe, Y. Zhang, and B. Nath, "Robust statistical methods for securing wireless localization in sensor networks," *Information Processing in Sensor Networks, 2005. IPSN 2005. Fourth International Symposium on*, 2005.

[9] TIRTL, *http://www.ceos.com.au/index.php/products/tirtl*, 2014, [Accessed October 5, 2014].

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1049–1056
DOI: 10.15439/2016F89
ACSIS, Vol. 8. ISSN 2300-5963

Modelling and evaluation of a multi-tag LED-ID platform

Grzegorz Blinowski
Institute of Computer Science,
Warsaw University of Technology,
Nowowiejska 15/19, 00-665
Warszawa; Poland
Email: g.blinowski@ii.pw.edu.pl

Adrianna Kmieciak
Institute of Computer Science,
Warsaw University of Technology,
Nowowiejska 15/19, 00-665
Warszawa; Poland
Email: akmiecia@elka.pw.edu.pl

Abstract—**An LED-ID system works like an electronic "tag" transmitting a short digital broadcasted message. Low complexity LED-ID installations, being a subset of an emerging class of visible light communication (VLC) systems, may be considered as a replacement of popular RFID tags, Bluetooth tags and Wi-Fi beacons. In this work, we focus on multi LED-ID environments with "dense" tag placement. The problems that we focus on are estimating the level of cross-tag interference and the issue of tag proximity: how closely can we place the tags without making the system unusable? We present a theoretical model with a numerical simulation of sample arrangements. We also describe the results of experiments we conducted in a real-world test environment under different external lighting conditions.**

I. Introduction

Visible light communication (VLC) is wireless optical communication technology through which baseband signals are modulated on the light emitted by an LED [1] – [3]. The decreasing cost and hence rapid adaptation of LED-based light make VLC a promising communication technique and an excellent alternative to radio-based wireless communication. A unique feature of a VLC system is that it performs two functions simultaneously: illumination and communication. This results in a reduction of costs because a separate system for data transmission is not needed any more – existing illumination infrastructure is used instead.

VLC systems have been proposed and implemented both for indoor and outdoor applications (see [2] and [4]). Indoor applications include a range of communication facilities provided today by WLAN and personal area networks (PAN) such as office communication [5], multimedia conferencing [6], peer-to-peer data exchange, data broadcasting (especially multimedia such as home-audio and video streams – see [7] – [10]). A relatively simple VLC system is able to achieve data rates of up to 100 Mbit/s over a distance of 1 – 3 m with a single light source and a simple equalized receiver [11]. Data rates of over 1 Gbit/s have been recently obtained for more complex transmitter-receiver configurations.

One application of VLC are LED-ID platforms, which can be used in numerous environments including shops and supermarkets, museums, plenum spaces, etc. An LED-ID system works as an electronic "tag" transmitting a short digital broadcasted message. LED-ID systems, with their low complexity, may be considered as replacements of popular RFID and Bluetooth tags. An example of LED-ID systems in use are "smart" supermarket carts, which via illumination infrastructure record a shoppers' path for subsequent analysis. LED-ID systems may also be used to "tag" particular shop shelves and areas to enable fast product localization. Digital signage systems used in museums, exhibitions, etc. are another example of LED-ID technology. These signage systems may be used with specialized applications for mobile platforms to provide information about objects in proximity. Yet another LED-ID field of application arises in environments where the usage of radio-based technology, such as Bluetooth, ZigBee or RFID, is hazardous or limited by regulations, for example in mines, petrochemical plants, aeronautics and hospitals.

In comparison to more complex VLC systems, LED-ID tags are simple: their functionality is limited to broadcasting digital information. LED-ID tags typically do not provide duplex communication; tag "programming" is done via wired or wireless connections and in some cases the ID is simply hardcoded into the tag's microcontroller unit. In many cases, the tag is simple enough that it does not support cooperation in a multi-transmitter environment – it simply broadcasts its information with no regard for other tags competing for the same medium. As was explained in [12] and [13], an optical communication link can be modelled as a Poisson channel. In the general case of multiple transmitters, it was shown that the maximum total throughput of the Poisson MAC monotonically increases with the number of transmitters and is bounded from above. Therefore, adding more inputs to a Poisson

This work was supported by the Statutory Grant of the Polish Ministry of Science and Higher Education to the Institute of Computer Science, Warsaw University of Technology.

MAC eventually saturates the entropy rate (and hence the information content) of the output. Given the channel capacity limitation, a signal source with sufficient transmitting power will be able to saturate the channel, obscuring the data source. The same result may also be obtained by a larger number of low-power transmitters.

In this work, we will focus on multi LED-ID environments with "dense" VLC tag placement. Examples of such environments include article tagging on shop shelves, the tagging of individual items in museum exhibitions, and other cases where light-tagged items are placed closely together. In such environments with dense arrangements of tags, the cones of light emitted by different luminaires overlap. The problems that we focus on in this work are as follows: what measures may we use to evaluate such an environment? What is the level of cross-tag interference? How closely can we place the tags without making the system unusable?

The structure of this paper is as follows: in section II we present the architecture of LED-ID systems, which leads us to the theoretical system model then described in section III. We use the model for the numerical simulation presented in section IV. In section V, we show the results of an experiment that we conducted on a sample installation built from commercially available LED-ID components. Our work is summarized in section VI.

II. Architecture of an LED-ID system

An LED-ID system consists of a transmitter ("tag") and a receiver ("reader"). The transmitter must be able to modulate the emitted light to transmit the digital tag. It consists of a luminaire which may use one or more LEDs (typically a high power white-light LED in blue-LED / yellow phosphorous technology), an LED-driver IC and a microcontroller unit driving the amplifier.

The critical difference between VLC and radio-based communication is that in VLC, data can not be encoded in the phaseof the light signal. The information has to be encoded in the varying intensity of the emitted light. The demodulation depends on direct detection at the receiver - hence IM/DD (Intensity Modulated/Direct Detection) modulation techniques are used in VLC. Modulation in VLC must also take into account the requirements of dimming and flicker mitigation. Various modulation schemes have been proposed for VLC systems, including:

- On-Off Keying (OOK) - the data bits 1 and 0 are transmitted by turning the LED on and off respectively. In the "0" state, the LED is not completely turned off but rather the light intensity is reduced. The advantages of OOK include its simplicity and ease of implementation.
- Pulse Width Modulation (PWM) - the widths of the pulses are adjusted based on the desired level of light dimming while the pulses themselves carry the modulated signal in the form of a square wave.
- Pulse Position Modulation (PPM) - the position of the pulse in a series of pre-defined time-slots identifies the transmitted symbol.
- Orthogonal Frequency Division Multiplexing (OFDM) - the channel is divided into multiple orthogonal subcarriers and data is sent in parallel sub-streams modulated over the subcarriers. Standard "radio-based" OFDM techniques need to be adapted for application in IM/DD techniques because OFDM generates complex-valued bipolar signals which need to be converted to real values.
- Frequency Shift Keying (FSK) – the instantaneous frequency of a constant-amplitude carrier signal is changed between two (for BFSK) or more (for MFSK) values by the baseband digital message signal.

The modulation methods described above have numerous pros and cons [14]. OFDM is very effective in high speed transmission, when inter-symbol interference and multipath fading start to dominate the channel capacity. However, it is difficult to implement OFDM with the LED-driving analogue hardware that is currently used. PWM, PPM and their numerous variants provide light dimming and a simple way to eliminate flicker while maintaining good channel bandwidth. In some cases, the dominant factor in choosing a modulation method is the hardware available and its limitations. For example, with customer mobile devices, a plug-in photodetector is the simplest and the cheapest choice (see the receiver section below), and a compatible modulation method therefore must be used – FSK in this case. In this study, we assume that FSK modulation is used, as it is currently the dominant modulation method for mobile platforms.

In general, VLC systems may use two types of receivers: (1) a photodetector – typically a photodiode (a non-imaging receiver); (2) an imaging sensor (a camera). In LED-ID systems, where low cost is an important factor, simple photodetector receivers are used. Even with no or with very simple analog equalization they provide bandwidth that is more than adequate for LED-ID applications. In customer-grade VLC, a smart-phone or a similar device is used as a reader. In this case, the phone's built-in camera could be considered as the receiving device.

However, this type of imaging sensor is very slow and inadequate for data transmission applications[1], hence plug-in photodetector modules compatible with a standard audio-in/out port are used instead.

III. System model

The components of an LED-ID system include an LED transmitter consisting of one light source and a photodiode receiver. The received signal depends on the physical characteristics of the transmitting LED, the receiver, and channel characteristics. We use ray optics theory to calculate signal and noise levels and derive adequate metrics. We assume the Multiple Input Single Output (MISO) model, with multiple transmitting LEDs and one photodiode detector. A single transmitting LED is characterized by a half-power semi-angle and central luminous intensity (measured in candelas). The receiver is a simple non-imaging photodetector with an optical filter, optical concentrator and a single photodiode element with a field of view (FOV) angle, gain, a photodetector area and conversion efficiency (measured in A/W).

The metric that we use to measure the impact of the interference is bit error rate (BER), which depends on the signal to noise ratio (SNR) and modulation scheme. The relationship between BER and SNR depends on the modulation type and modulation parameters. For binary frequency shift keying (BFSK) with non-coherent detection [15]:

$$BER_{BFSK}(SNR) = \frac{1}{2} exp\left(-\frac{SNR}{2}\right) \quad (1)$$

we calculate SNR as follows:

$$SNR_s = \frac{\overline{s_{data}^2}}{\left(N + \overline{s_{interf}^2}\right)} \quad (2)$$

where s_{data} is the data signal, s_{interf} is the signal transmitted by other luminaires, and N is noise.

The problem of noise in VLC environments has been studied in detail [16]. In general, the following noise sources should be considered:background and transmitter LED shot noise, thermal noise in the detector and the influence of inter symbol interference (ISI). The background or ambient noise comes from the sun and artificial light sources:

$$N = \sigma_{shot}^2 + \sigma_{thermal}^2 + \sigma_{ISI}^2 \quad (3)$$

where N is the total noise variance and $\sigma_{shot}, \sigma_{thermal}, \sigma_{ISI}$ is the standard variance of shot noise, thermal noise and ISI respectively. The proper estimation of noise in VLC environments is crucial in studying the maximum attainable transfer rates under various conditions and modulation schemes. The input referred noise variance depends on the signal data rate. For low data rates in the range of 10^2 – 10^4 bits/s, the major noise factor is shot noise:

$$\sigma_{shot}^2 = 2qRPB + 2qI_{bg}I_2B \quad (4)$$

Where q is the electronic charge, R is the responsivity of the photodiode, B is the equivalent noise bandwidth, P is the received power, I_{bg} is the background current, and for a p-i-n/FET receiver we assume $I_2 = 0.56$. In the multi-luminaire study that we conduct in this paper, the dominant noise factor is the interfering signal from neighboring luminaires and not physical noise itself.

Now we will present the analytical model of the optical wireless channel which will let us derive SNR and BER measures for different physical scenarios. Our analysis is based on the fundamental paper by Komine and Nakagawa [17].

A single LED is a Lambertian emitter – its radiation intensity is a cosine function of the viewing angle and is given by

$$I(\theta) = P_t \frac{(m+1)}{2\pi} cos^m(\theta) \quad (5)$$

where θ is the irradiance angle, P_t is the transmitted power and m is the order of Lambertian emission given by irradiance semi-angle $\theta_{1/2}$ (half power angle)

$$m = -\frac{\ln 2}{\ln(\cos(\theta_{1/2}))} \quad (5)$$

Light propagates from the LED to the receiver via a channel which is modeled by direct channel transfer function h_d:

[1]It is possible to use a more complex multi-light source transmitter which takes advantage of the "imaging" properties of the sensor, however this is much more expensive than a simple single luminaire solution.

$$h_d = \begin{cases} \frac{(m+1)A\cos^m(\theta)}{2\pi d^2}\cos(\psi)\,R(\psi) & 0 \le \theta \le \theta_{FOV} \\ 0 & \theta > \theta_{FOV} \end{cases} \quad (7)$$

whereθ is the irradiance angle, ψ is the angle of incidence, A is the receiver area, $R(\psi)$ is receiver gain, d is the distance from the LED to the receiver and $\theta_{FOV}$is the receiver's FOV semi-angle. The geometric model of this simple line of sight (LOS) case is shown in Fig. 1.

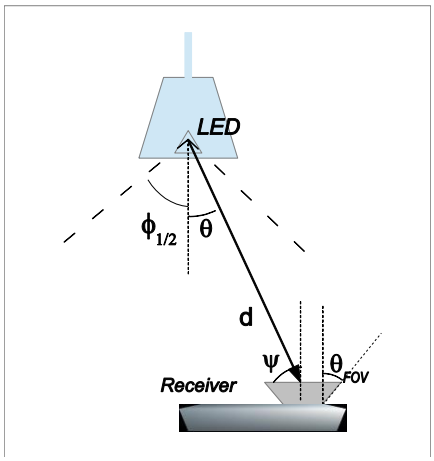


Fig. 1. Geometric model of LOS communication

For a single source, the output signal of the LED transmitter is given by the following general formula:

$$p_0(t) = P_t[1 + \mu\, x(t)] \quad (8)$$

where$P_t$is the power transmittedfrom a single LED, μ is the modulation index and *x(t)* is the modulating signal. Assuming that the receiver is DC blocked, we get the following general formula for the received signal:

$$s(t) = h_d P_t\, \mu\, x(t) \quad (9)$$

Considering the "legitimate" and "interfering" sets of transmitters, we obtain the following:

$$s_{data}(t) = \sum_{data\_LEDs} \{P_{LED}\, \mu\, x(t) h_d\} \quad (10)$$

$$s_{interf}(t) = \sum_{interf\_LEDs} \{P_{LED}\, \mu\, x(t) h_d\} \quad (11)$$

We use (10) and (11) in a numerical model to calculate BER as given in (1) for our study.

IV. Simulation results

For our numerical simulations, we designed sample scenarios with 3 and 9 luminaires. The scenarios' dimensions are 2m x 2 m x 2m. We assume that the detector's photodiode is parallel to the luminaire plane. We simulated two luminaire placement scenarios: L1 - with 3 luminaires arranged in a line as shown in Fig. 2,and scenario G1 - with 9 luminaires arranged in a 3x3 square gridas shown in Fig. 3. The first scenario relates to a "shop shelf" arrangement and the second to an exhibition cabinet or stand. The physical parameters are summarized in Table I.

TABLE I
PHYSICAL PARAMETERS OF THE SIMULATED SCENARIOS

Photodetector parameters	
FOV (field of view)	60^o
Detector area	1 cm^2
Detector gain	1.3
Scenario parameters	
Dimensions	2m x 2 m x 2 m
Luminaire spacingΔx, Δy	L1: 16 cm G1: 16 cm, 16 cm
# of luminaires, scenario L1, G2	3, 9
Luminaire parameters	
Optical power	1 W
Radiation semi-angle	20^o

In both scenarios we show the logarithmic plots of the computed BER for data transmission. We assume that the BER level of maximum 10^{-2} is required for effective transmission of the LED-ID tag.

In scenario L1 we calculated BER for outer lamps, while the inner lamp is the interfering transmitter. BER is calculated on a plane at a distance of 30, 40 and 50 cm from the luminaire plane – Fig. 4. BER decreases as we move the receiver away from the luminaires and achieves values in the range of 10^{-6}, 10^{-2} and 10^{-1} respectively. We can conclude that BER becomes intolerably high when the light cones (as limited by the radiation semi-angle) start to fully overlap each other, i.e. when the radius of the luminaire light cones is equal to the distance of their centers.

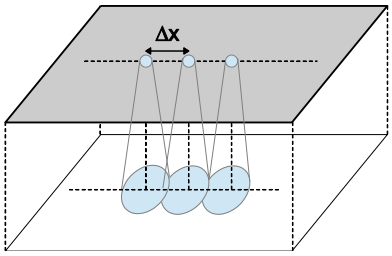


Fig. 2. Simulated scenario arrangement L1.

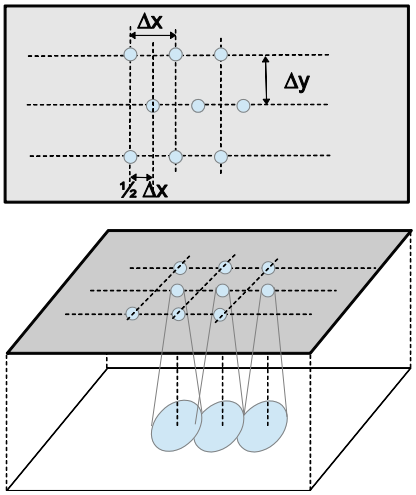


Fig. 3. Simulated scenario arrangement G1.

In scenario G1 we also calculated BER on a plane at a distance of 30, 40 and 50 cm from the luminaire plane – Fig. 5. BER decreases as we move the receiver away from the luminaires and achieves values in the range of 10^{-6} 10^{-2} and 10^{-1} respectively. The LED-ID tag under respective luminaires can be properly resolved, as was in the case of a single luminaire line.

The scenarios prove that the resolution of LED-ID tagging is quite satisfactory – even with dense luminaire placement, we are still able to obtain a reliable tag readout.

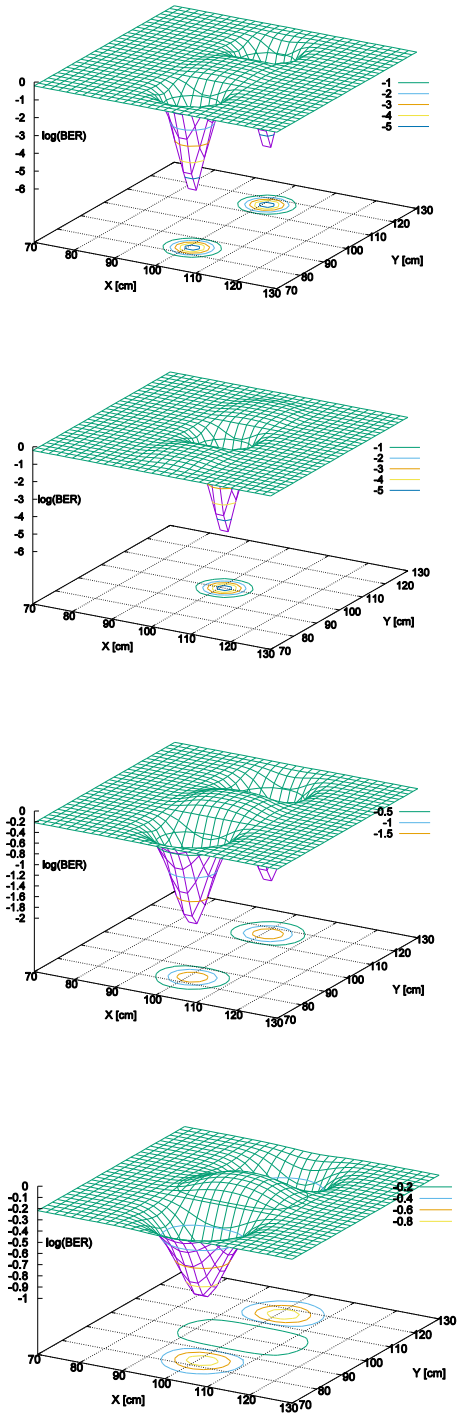


Fig.4.BER - simulation results for scenario L1. From top: (1) outer luminaires, distance 30 cm; (2) inner luminaire, distance 30 cm; (3) outer luminaires, distance 40cm; (4) outer luminaires, distance 50cm.

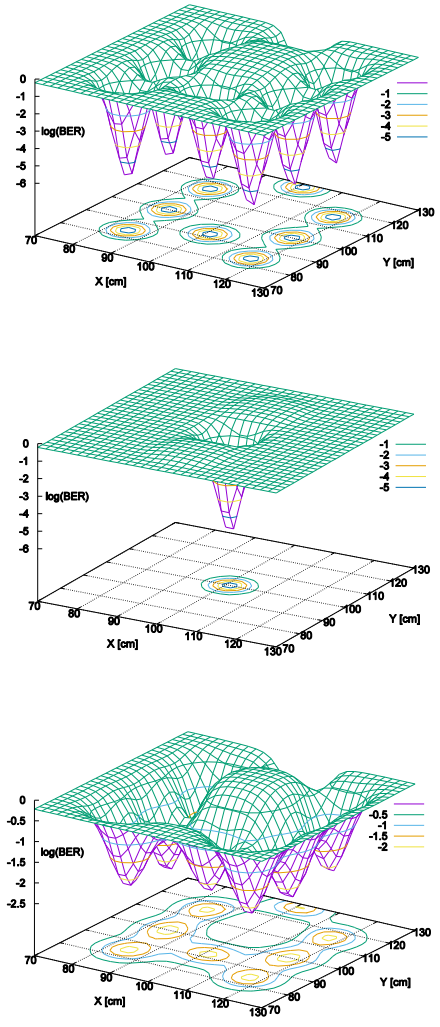


Fig. 5.BER - simulation results for scenario G1. From top: (1) outer luminaires, distance 30 cm; (2) inner luminaire, distance 30 cm; (3) outer luminaires, distance 40cm.

V. SAMPLE SYSTEM EVALUATION RESULTS

For our experiments, we used LED-ID devices manufactured by OLEDCOMM as shown in Fig.6. These luminaires came in the form of a desktop lamp with a 1W single LED light source, with a ~15^{o} radiation semi-angle (as declared by the manufacturer, this parameter varies from unit to unit, and in most cases is a few degrees larger than declared). The luminous flux when measured 50 cm from the light source is ~ 900 lx (it varies by 5% between different luminaires). The OLEDCOMM kit also contained an audio-port plugin receiver compatible with most Android devices and an SDK library. The receiver uses a simple PIN photodiode with no optical concentrator or filter.

Fig.6. LED-ID equipment used in experiments

For the tests, we implemented a client-server test suite consisting of an Android client program written in Java which gathers information regarding light intensity and lamp-ID numeric tags as reported by the library and sends it to the data-collecting server. The client has provisions for recording semi-automatic LED-sensor distance and is also able to buffer the data if the server is not accessible. Collected data may be manually tagged in the application to record various field conditions such as test series name, external illumination conditions, etc. The server stores data received for the client for further analysis. The server was implemented with the Django Rest Framework [18].

A. Testing under various field conditions

To establish the baselines, we tested three sets of communication kits under the following external light conditions: (1) minimal external light source (< 10 lx); (2) ambient dispersed light (50-200lx); (3) unmodulated direct light from an external LED source (up to 3000 lx); (4) direct sunlight (3000 – 5000 lx). The ambient light intensity levels were measured with a certified lux meter.

In each case we measured the maximum distance that guaranteed reliable ID transmission (5 tags correctly received in sequence). Measurements were collected with 1 - 5 cm intervals for d and x values – see Fig. 7, for 3 different luminaires and repeated 2-3 times.The results were averaged. As expected, we can conclude that as interfering conditions vary, so does the maximum reliable distance and to the lesser extent the maximum reliable angle. Table II summarizes the obtained data.

TABLE II
MAXIMUM DISTANCE AND MAXIMUM ANGLE FOR RELIABLE TRANSMISSION UNDER DIFFERENT CONDITIONS

Condition	Maximum reliable distance [cm]	Maximum reliable angle [deg]
Declared	370	30
Measured w/o and with interfering light sources		
No external light	250	38
Ambientlight	230-240	34
Unmodulated LED	100-220	36
Direct sunlight	60-180	26

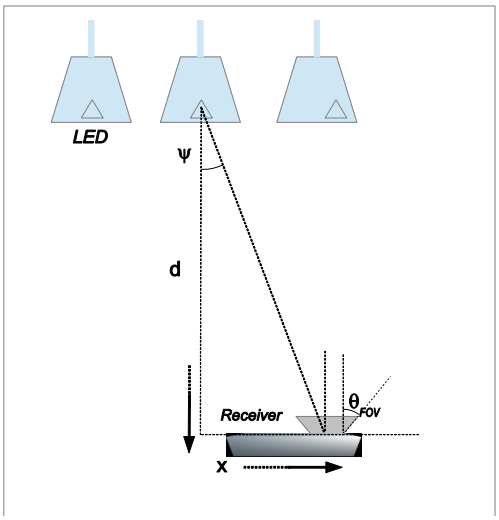


Fig.7. Single and multiple transmitter experiment setup.

B. Testing with multiple luminaires

The experiment was set up to verify simulation results. We used three lamps, placed at a distance of 16 cm from each other. We collected tag read-outs with the receiver moving directly under the lamps on a parallel plane distanced 30 cm from the luminaires(*d*). The horizontal distance corresponds to *x* from Fig. 7. Fig.8 shows the obtained results – the resolution of tag readouts is compatible with the results of the simulation, and the error rate (number of bad or inconclusive tag readouts) was ~ 5%, with errors occurring in the transition area.

VI. CONCLUSION

We tested a multi-tag LED-ID system both via numeric simulations and by means of an experiment. We have concluded that in a dense transmitter setup, i.e. with overlapping light cones, it is still possible to resolve transmitted digitaltags, up to the point where light cones start to totally overlap. The methodology that we have presented should be useful for planning more complex LED-ID scenarios.It should also be helpful to the vendors of LED-ID hardware.

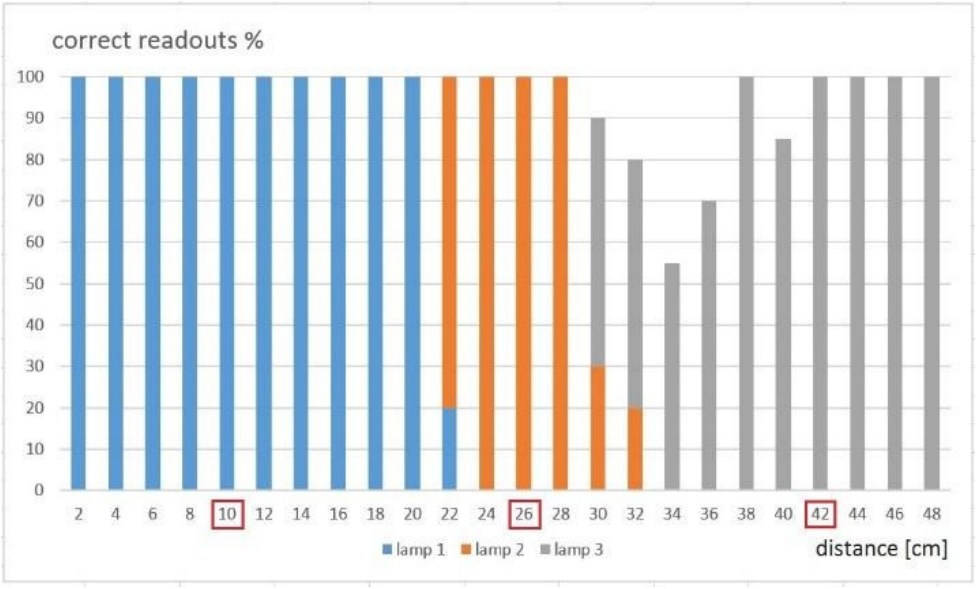


Fig.8. Summary of tag readouts from experiment. Transmitters were placed at positions: 10, 26, 42 cm (marked as squares on the axis).

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their helpful comments in the preparation of this article.

REFERENCES

[1] M. Nakagawa, "Visible Light Communications," In Proc. Conference on Lasers and Electro-Optics/Quantum Electronics and Laser Science Conference and Photonic Applications Systems Technologies, Baltimore, 2007, DOI: 10.1109/CCNC.2012.6181092

[2] H. Elgala, R. Mesleh and H. Haas, "Indoor Optical Wireless Communication: Potential and State-of-the-Art," IEEE Communications Magazine, Volume: 49, Issue: 9, 2011, pp. 56-62.

[3] A. Tsiatmas, C. P. A. Baggen, F. M. Willems, J. P. Linnartz and J. W. Bergmans, "An illumination perspective on visible lightcommunications," In Communications Magazine, IEEE, 52.7, 2014, pp. 64-71.

[4] Samsung Electronics, ETRI, VLCC, University of Oxford, "Visible Light Communication: Tutorial," 2008, http://www.ieee802.org/802_tutorials/2008-03/15-08-0114-02-0000-VLC_Tutorial_MCO_Samsung-VLCC-Oxford_2008-03-17.pdf

[5] M. B. Rahaim, A. M. Vegni and T. D. Little, "A hybrid radio frequency and broadcast visible light communication system," in Poc. IEEE Global Communications Conference (GLOBECOM) Workshops, 2011, pp. 792–796.

[6] L. B. Chen et al. "Development of a dual-mode visible light communications wireless digital conference system," In Consumer Electronics (ISCE 2014), The 18th IEEE International Symposium on, 2014, pp. 1-2.

[7] J. P. Javaudin, M. Bellec, D. Varoutas and V. Suraci, "OMEGA ICT Project: Towards Convergent Gigabit Home Networks," in Proc. International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), Cannes, France, 2008

[8] K. D. Langer et al., "Optical Wireless Communications for Broadband Access in Home Area Networks," In Proc. International Conference on Transparent Optical Networks, ICTON, 2008, pp. 149 - 154, DOI: 10.1109/ICTON.2008.4598756

[9] D. C. O'Brien et al, "Home access networks using optical wireless transmission," In Proc. Personal, Indoor and Mobile Radio Communications, IEEE 19th International Symposium on, 2008, pp. 1-5

[10] D. C. O'Brien et al, "Gigabit Optical Wireless for a Home Access Network," in Proc. IEEE 20th International Symposium on Personal, Indoor and Mobile Radio Communications, 2009, pp. 1-5.

[11] H. Le Minh, et al. "100-Mb/s NRZ visible light communications using a postequalized white LED." Photonics Technology Letters, IEEE 21.15 (2009): pp. 1063-1065.

[12] G. Blinowski, "Security issues in visible light communication systems," IFAC-PapersOnLine 48.4 (2015): 234-239.; http://www.sciencedirect.com/science/article/pii/S2405896315008149

[13] A. Lapidoth and S. Shamai, "The Poisson multiple-access channel," Information Theory, IEEE Transactions on, 44(2), 1998, pp. 488-501.

[14] P. H. Pathak et al. "Visible Light Communication, Networking, and Sensing: A Survey, Potential and Challenges," Communications Surveys & Tutorials, IEEE 17.4 (2015): 2047-2077.

[15] F. Xiong. Digital Modulation Techniques, (Artech House Telecommunications Library). Artech House, Inc., 2006.

[16] A. Agarwal and S. Garima, "SNR Analysis for Visible Light Communication Systems," International Journal of Engineering Research and Technology. Vol. 3. No. 10 (October-2014). ESRSA Publications, 2014.

[17] T. Komine and M. Nakagawa, "Fundamental analysis for visible-light communication system using LED lights," Consumer Electronics, IEEE Transactions on , vo1.50, no. I , 2004, pp. l00 -107,

[18] Django REST framework version 3, http://www.django-rest-framework.org/

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1057–1066
DOI: 10.15439/2016F202
ACSIS, Vol. 8. ISSN 2300-5963

Accurate Event Detection and Velocity Estimation in Wireless Environments

Falk Brockmann, Sascha Jungen, Chia-Yen Shih, Marcus Handte and Pedro José Marrón
University of Duisburg-Essen,
Schützenbahn 70, 45127 Essen, Germany
Email: see http://www.nes.uni-due.de/staff/

***Abstract*—Radio signals can be used to detect the presence of a person (target) in an environment by analysing the fluctuations in the Received Signal Strength Indicator (RSSI). The velocity of the target can be estimated by examining the sequence of disturbances in consecutive radio links over a period of time. This requires knowledge of the deployment of the radio transceivers and the time when the target crosses the Line of Sight (LoS) of each radio link. However, it is not trivial to precisely estimate the exact time of the link crossing due to the broad range of RSSI fluctuations generated as the target approaches the link. In this paper, we evaluate and compare 15 techniques for estimating the velocity of the target and propose enhancements to some of the techniques. In our experiments the techniques perform with an average accuracy in the range between 13.02% and 96.18%, which corresponds to an average error of 0.05m/s for a moving target.**

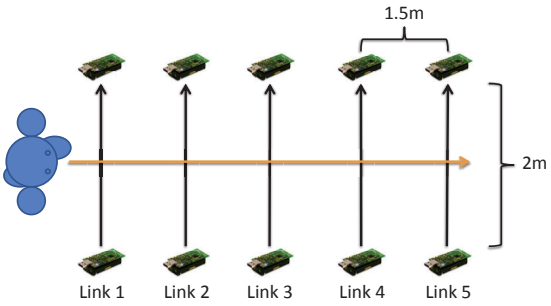


Fig. 1. Concept diagram of experiment set-up

I. Introduction

WIRELESS sensor networks (WSN) have been the interest of the research community for many years. They consist of a number of small, affordable devices (*motes*), typically equipped with a microprocessor, some sensors, for example a light or humidity sensor, and a transceiver chip for radio transmissions. Traditionally, WSNs are used to monitor environments for long-term changes in attributes like temperature, but recent studies have shown, that they can also be used to detect the presence of persons by analysing disturbances in the radio links between the motes [1], [2], [3], [4], [5].

Generally, if a person (the target) enters a radio link, the human body causes multipath fading of the radio signal [2]. This is detected by analysing for example the *Received Signal Strength Indicator* (RSSI) of the transmission. The resulting characteristic of the RSSI values will show the presence of a target, but can take different shapes. A highly sensitive link will react early to the presence of a human. The collected RSSI samples will have strongly varying values in a relatively large time interval, even when the target is still some distance away. On the other hand human presence might only create a few higher or lower spikes than the average RSSI signal on an insensitive link. Furthermore, if the target is not just standing still, but moving through the link, this will cause additional fluctuations in the RSSI values [6].

In addition to simply detecting a target it is also possible to estimate the targets' position. One approach would be employing a grid of motes, creating a mesh of multiple radio links [3]. If events are simultaneously detected on several intersecting links, the targets' position can be estimated at this intersection. To track a target through such a system, the detection can be repeated periodically, each time updating the targets' position. Using multiple position estimates it is also possible to derive further information about the targets' behaviour, for example the targets' trajectory or velocity.

There are various reasons why determining the velocity of a target can be useful. For example in a system tracking not one but multiple targets, like in [7], [8], [9], it is beneficial to not only know the position of every target, but also to predict future positions. This information can be used to differentiate between two targets crossing their paths in close vicinity [9]. Knowing the direction and speed is also interesting when monitoring an area where only a limited amount of sensors is available. Since the coverage might be too sparse to allow continuous tracking, a general idea about the movement could be helpful, especially when monitoring an area with movement restrictions. For example, if monitored targets are moving towards a dangerous location, like a broken elevator shaft or the site of a fire, then an alarm sound could be triggered to prevent possible harm.

In any case, the challenge of deducing additional information from the pure detection of a target lies in pinpointing a precise moment within the detection event. This moment should be identifiable, even when the radio link behaviour is different. For our experiments we use the point of the detection event when the target has the greatest influence on the RSSI, which we define as the peak. Since humans are three

978-83-60810-90-3/$25.00

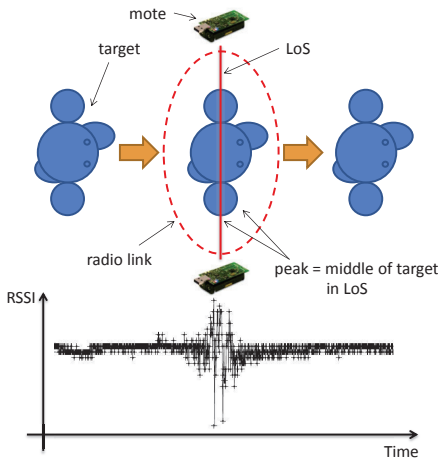


Fig. 2. Target passing the Line of Sight of a radio link

dimensional, it takes a certain time for a target to pass a radio link, as seen in Fig. 2. The peak can be seen as the point in time, when the middle of the target is in the Line of Sight of the link [4]. The goal of this paper is to evaluate 15 different techniques on how to best detect the velocity of a target with the highest possible accuracy. For this we estimate the peak of the radio link crossing event and use the generated results in the velocity estimation of a target moving at different speeds.

Consider the velocity estimation following the simple formula $v = \frac{s}{t}$ as an example. To compute the speed of a target, the locations of at least two radio links are required, and the RSSI values and their respective timestamps are taken from the links. If a target consecutively moves through the radio links, as for example seen in Fig. 1, the actual crossing of a link is recorded as an interval of collected RSSI samples while the link was disturbed. The detection events on both links need to be identified to estimate the time it took the target to cross the distance between the links. Because of the unsteady nature of the RSSI it is not possible to simply use the earliest distorted sample, since its occurrence can be different for dissimilar links. For accurate velocity estimation we have to identify the targets position in the link, respectively for each link. However, the best samples to indicate the actual crossing do not necessarily need to be the samples with the highest RSSI attenuation or the middle sample of the collected interval. The different link states and general fluctuation in the RSSI distort the collected data and make accurate predictions difficult.

The remainder of this paper is structured as follows: Section II gives an overview of research similar to our approach. Section III describes the different techniques used, followed by an evaluation which details the results of our experiments in Section IV. The paper is concluded in Section V with a summary and a short outlook.

II. Related Work

Traditionally, instead of using the RSSI to record an event, a system detecting the presence of a target can be based on using a laser range finder. A range finder emits a laser beam that is reflected off of any object it hits. The reflection is caught by the light sensor of the device, the Time of Flight is measured, and the distance to the object computed. If the distance suddenly decreases, the laser beam is reflected at a shorter distance. This is the case, if a target is in the LoS between the laser emitter and the light sensor. However, using a range finder is unfavourable in scenarios, where a larger region needs to be covered, since the laser beam only stretches over a relatively narrow area. Furthermore, the sensor measuring the light intensity needs to be shielded from direct sunlight, to avoid false negatives. Also, these systems can easily be fooled, when a target is avoiding the beam by stepping over it, without interrupting it. Depending on the scenario, a RSSI-based device-free localisation system might be more suitable. Since the RSSI is already disturbed when a target moves near the radio link, avoiding it is not easily possible.

Multiple systems employing the RSSI to detect the presence of a target have been developed [2], [10], [11]. In these systems sensor nodes are deployed at all edges of the monitored area and a grid of radio links is created. A target standing or walking through the grid will cause a characteristic disturbance which can be detected [2].

However, the link characteristic is not specific to a target. In [7] multiple targets are tracked by dividing the monitored area into voxels. A target entering the area will cause disturbances in a set of voxel simultaneously, which are then clustered. Each cluster in the area symbolises a target and is tracked while in the system. But intersecting trajectories create the difficulty of continuously keeping track which target is matched to a specific cluster.

To cope with RSSI irregularities, alternatives to the use of the absolute RSSI or its average have been explored. In [8] the time of flight of radio messages is used in an antenna set-up with five transmit antennas and five receive antennas. The signals reflecting off of the human body are analysed, which allows for the tracking of up to five people. While no longer based on the absolute RSSI, like in [7] still no clear identification of the tracked target out of a group is possible.

The behaviour of the radio links is further analysed in [12]. In general, RSSI values are unstable and always slightly fluctuating because of minimal changes in the environment. Slight variations in the area, for example moving a chair or opening a window, can have a huge impact on the radio propagation. Despite that, [12] defines two classes of observed link behaviour when a target enters the link. Both are based on the different qualities of a link. An *anti-fade* link has very high RSSI values and will experience a strong attenuation and a significant drop in the RSSI values when a target enters the area of the link. A *deep-fade* link on the other hand already has a very weak signal quality. For such links it is possible that the quality increases when a target enters the link, due

to refraction and scattering of the signal. Thus, the resulting RSSI characteristic is generally an inverted version of the anti-fade one. For this reason and other environmental influences, a detection event can have a very dissimilar RSSI characteristic compared to the characteristic of an event on a different link.

The authors of [4] suggest focusing on the changes in the variance of the RSSI, instead of looking at the absolute value. Regardless whether the absolute RSSI average in- or decreases after a target enters the area of the link, the RSSI variance will always be higher, because the link gets disturbed. Fundamentally, a long-term and a short term variance are computed, compared, and incorporated to the *alert magnitude*. The larger the alert magnitude is, the more likely a target is present in the detection area. This method simultaneously reduces the influence of different link qualities but also of changes in the environment on the RSSI in long-term measurements. By adapting the long-term variance in phases of tranquillity, in which the variance of the RSSI is low, small changes in the radio propagation model caused by the environment are handled.

Using the RSSI to detect the speed of driving vehicles, [13] collects samples of the signal strength of four links covering a road section. The resulting measurements are analysed using a statistical and a curve fitting technique. In the statistical approach, the time between the entering and exiting event of the monitored area is used to estimate the speed. The curve fitting technique exploits the relation of the height of the variance and the speed of a car. The faster the car passes the monitored area, the higher is the influence on the recorded RSSI variance. However, this technique is applied to fast moving cars made of metal [14], not to humans.

III. Analysed Techniques

In this paper we analyse 15 different techniques for estimating the velocity of a target. The techniques are divided into three classes based on their input data. The techniques of the first class are working with the RSSI. They either use all available samples or search for the samples with the minimum value, as introduced in [15]. The techniques of the second class use the *alert magnitude*, the result of the RSSI variance based algorithm introduced in [4]. The techniques either select the maximum alert magnitude values or again use all available samples. We propose enhancements to the second class to form the third class of techniques. Instead of only using samples with the maximum value, more samples with high values are included into the input data. We evaluate the different methods for estimating the targets' speed in an experiment.

To accurately compute the targets' speed two things are needed: The distance travelled and the time it takes the target to do so. This can be measured for example in a system, where the crossing of the *Line of Sight* (LoS) between two sensor nodes creates a disturbance in the RSSI values of periodically sent messages. Two measures of the targets' position at two different times can be subtracted from each other to determine the distance. Since the motes in our experiments are set at fixed positions, in the evaluation we consider the distance between them as known. To determine the time, the presence of the target at each position needs to be detected and the detection event analysed. The event itself stretches over an interval consisting of several recorded RSSI samples. Each sample includes a RSSI value and a timestamp. During the recording interval, the target approaches the link, enters it, passes the Line of Sight, and leaves it again. Disturbances in the stream of RSSI values are caused during the time the target stays in the area of the radio link.

Would at this point a sensor producing a boolean result be deployed to record the disturbances, the outcome would be binary: Either a body part of the target is perceived by the sensor and the target is detected, or the sensor is not triggered and no detection is registered. However, the human movement also incorporates the swinging of arms and legs. Therefore, an event can be triggered prematurely, when for example an arm is detected before the rest of the body enters the monitored area, artificially prolonging the event. The targets' distance to the LoS of the radio link will be shifted during the whole event, when compared to the event of a different link. Also, it cannot be assumed that the LoS crossing of the target is in-line with the middle of the detection event. This reduces the accuracy of the velocity estimation.

Fortunately, in the case of detecting an event using RSSI values, the attenuation following the disturbance of the radio link is the highest when the target is closest to the LoS [16]. However, this does not necessarily make the lowest RSSI value the best point describing the event. There might be more than one recorded RSSI sample with the lowest value. Additionally, the RSSI values of a disturbed link are fluctuating. Another sample could time-wise better describe the link crossing, but may not match the lowest RSSI recorded.

The following sections illustrate the techniques to cope with these restrictions and estimate the velocity of a target moving through a group of radio links. For a better overview, all introduced techniques are listed in Table I. Afterwards, the methods are experimentally evaluated.

A. RSSI-based Techniques

The RSSI-based techniques use the unaltered data stream of RSSI values as input to estimate the events' peak and the targets' velocity. They can be found in the first column of Table I.

1) minRSSI: The first approach to identify the best RSSI sample indicating the peak is the naïve approach, abbreviated minRSSI. This appraoch was also used in [15]. The RSSI data stream of the monitored link is searched for the sample with the minimum value. If multiple samples with the same lowest value are found, all of them are collected. After the time interval of the event has been analysed, the timestamp of the first occurring sample with the lowest RSSI value is selected for the velocity estimation. This approach is easy to implement and low on computational and space complexity, since the samples can be discarded once analysed.

2) medianRSSI: The next approach, medianRSSI, does not simply collect all samples with the minimum value to select

TABLE I
APPLIED TECHNIQUES ORDERED BY INPUT DATA

Technique	Based on Value		
	Minimum RSSI	**Maximum Alert Magnitude**	**Maximum Alert Magnitude Set**
Raw	minRSSI [15]	maxAlertM [4], [13]	topMaxAlertM
Median	medianRSSI	medianAlertM	topMedianAlertM
Average	avgRSSI	avgAlertM	topAvgAlertM
Linear Regression	linRegRSSI	linRegAlertM	topLinRegAlertM
Curve Fitting	CurveFittingRSSI	CurveFitting [13]	-
Cross Correlation	CrossCorrelationRSSI	CrossCorrelation	-

the first one, but uses the timestamp of the middle sample. This change is introduced as a precaution against outliers and incorporates additional information about the order in which the samples arrive, the data stream. However, this approach ignores the timestamps of the other collected samples, which might have a beneficial influence on the peak estimation, especially when only two samples with the minimum value exist.

3) avgRSSI: Simply selecting the first sample or selecting the median sample is not always the best choice, given the fact, that an event can have a non-symmetrical characteristic. The avgRSSI approach is addressing this issue by removing the dependency on the precise timestamp of the RSSI sample. Again, all samples with the minimum RSSI value are collected, but then the average of all available timestamps is computed. The new value does not necessarily need to match the timestamp of any sample, but can take an arbitrary point in the event interval. Since the peak does not need to match the sample with the minimum value, or more precisely any single sample, this technique avoids the very limiting dependency of selecting a timestamp from a predefined set.

4) LinRegRSSI: An alternative method, which is also not bound to the exact timestamp of a sample, is the LinRegRSSI approach. LinRegRSSI models the relationship between the detected events on different links by computing a linear regression, based on the RSSI values. The input data set for this approach is the distance between the links and the timestamps of the selected samples. To obtain the necessary data, first all samples with the minimum RSSI value as well as their timestamps are collected for each link. With the distance between the links known in our scenario, the parameters of the linear regression are then estimated from the input.

The simple equation $y = m \cdot x + b$ is the equation of a linear function, where m describes the slope, b the y-intercept.The result of the linear regression is the linear function with the minimum squared error towards all input data points.

In order to estimate the velocity, data from at least two links is needed, but can be extended onto more. In our case, values from all available links are used. The slope of the resulting linear regression line equals the speed of the moving target. To get the time of the peak, the position of a link can be set as y and the equation then solved for x.

Ignoring the restrictions of having to choose a precise timestamp, this approach additionally uses information from multiple links.

5) CurveFittingRSSI: A technique introduced in [13] is the fitting of a curve to the data stream of the RSSI variance. The highest point of the curve would indicate the time of the peak. To analyse how well this performs based on the unaltered RSSI stream, we use the CurveFittingRSSI approach. A Gaussian function is fit as continuous curve to the stream of all RSSI values per link. The maximum of the resulting curve is defined by the characteristic of the detection event and will be shifted towards the highest disturbance of the link. Unfortunately this approach is rather computationally expensive, since all samples of the event interval have to be collected and incorporated into the Gaussian function.

6) CrossCorrelationRSSI: Another approach using all available RSSI samples per link is the Sliding Dot Product, also known as cross correlation. The cross correlation originates in the field of signal processing. Two signals are shifted towards each other along the x-axis, while in each step their data points are multiplied with each other. The resulting product is called the correlation coefficient coef_{cc}. The higher the correlation coefficient is, the more similar are the data streams. This method is often used to measure how far the signals need to be shifted in order for them to reach their point of highest similarity.

In this paper we use the cross correlation to sample-wise compare the data streams of two links by computing their coef_{cc}. Given the characteristic of the input data, the highest similarity is obtained when the events of both links are overlapping.

The approach works in detail as follows: The first sample of the first link is multiplied with the last sample of the second link. The resulting cross correlation coefficient is stored. Next the links are shifted towards each other by increasing the number of samples that are compared. The first two samples of the first link are compared with the last two of the other and again the correlation coefficient is stored. This is repeated until the last sample of the first link has been compared with the first sample of the last link. Afterwards, the highest correlation coefficient between the links is selected and the number of shifts to reach this coefficient is counted.

Since the frequency with which the samples are sent, is known, the number of shifts indicates how many samples have

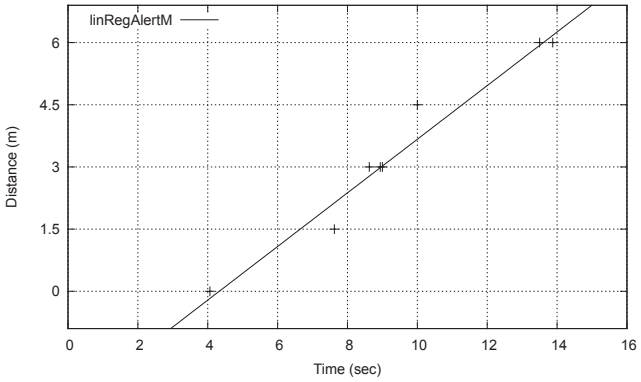


Fig. 3. Linear regression using alert magnitude values of five links

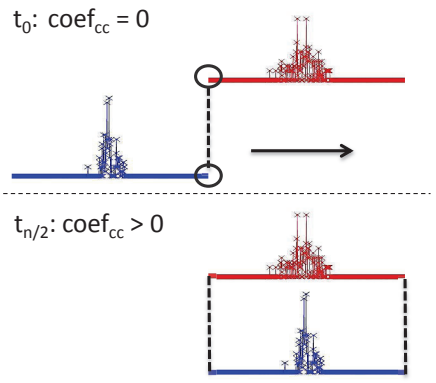


Fig. 4. General mechanism behind the cross correlation

been sent in the time it took the target to cross the space between first and second link. Analysing the sample rate we deduce the time between the LoS crossings. Strictly speaking, the cross correlation is not searching for the best sample to describe the event, but computes the velocity directly.

B. Alert Magnitude based Techniques

All unaltered RSSI-based techniques work best with anti-fading links, since they use the minimum RSSI value and on these links the maximum drop in the RSSI level is characterizing the crossing event. In the case of deep-fading links, where the signal improves after a target enters the link, a different method is needed. On those links the RSSI values near the peak are higher, even though fluctuations cause some RSSI samples to have a very low value.

To handle both cases, a RSSI variance based technique is used. Contrary to the absolute value, the variation of the RSSI always increases when a target is in the link [16]. In this paper the algorithm from [4] is applied. While the better handling of the issue of anti- and deep-fading links is addressed by changing the input data, the general issue of selecting a sample still remains the same. This being the case, the aforementioned methods of finding the peak of the LoS crossing event can also be applied on the alert magnitude.

All techniques using the alert magnitude are summarized in the second column of Table I.

1) maxAlertM: Following the given nomenclature in [4], the technique searching for the maximum alert magnitude is called maxAlertM. After all samples with the highest alert magnitude are found, the first occurring sample, which is the sample with the minimum timestamp, is selected.

2) medianAlertM: Like the medianRSSI approach, the medianAlertM approach selects the middle one of all samples with the highest alert magnitude. If an even number of samples with the same highest alert magnitude is encountered, the first occurring of the two middle samples is selected as a tie-breaker.

3) avgAlertM: The avgAlertM approach computes the average time from the timestamps of the collected samples with the highest alert magnitude. Again, this is analogous to the RSSI-based approach, in this case avgRSSI.

4) linRegAlertM: The approach to compute the linear regression, based on the alert magnitude, is LinRegAlertM. It is computed from all samples with the maximum alert magnitude of each link, using data from all available links. The resulting regression line is visualized as an example in Fig. 3. The displayed data is taken from the first run of the first measurement of our experiment. The y-axis value represents the distance of a radio link to the first link of the experiment set-up, the x-axis value represents the arrival time of the samples. For this example the distance between the links was set to 1.5 m and samples from five different links are used.

5) CurveFitting: While RSSI values are always unsteady and slightly fluctuating, the alert magnitude takes the form of a steady line, except during an event. All values are zero, as long as the link is not disturbed. However, they become greater than zero when indicating a target in the area of the radio link. This behaviour is beneficial to the CurveFitting approach, which fits a Gaussion function to all available alert magnitude values. Again, the precise time of the peak is not restricted by the timestamp of the samples.

Since the alert magnitude is based on the RSSI variance, this more closely resembles the curve fitting used in [13], then CurveFittingRSSI does.

6) CrossCorrelation: Computing the cross correlation from the alert magnitude values is the CrossCorrelation approach. The mechanism behind this approach is visualized in Fig. 4. In the beginning of the correlation process $coef_{cc}$ equals 0, but is gradually increased, when the data streams of the alert magnitude are shifted towards each other.

C. Alert Magnitude Set based Techniques

To enhance the performance of the methods using the alert magnitude and to take precautions against outliers, we introduce a new group of techniques. Instead of only searching for the samples with the maximum alert magnitude, the set

Fig. 5. Gird of motes in an lab environment

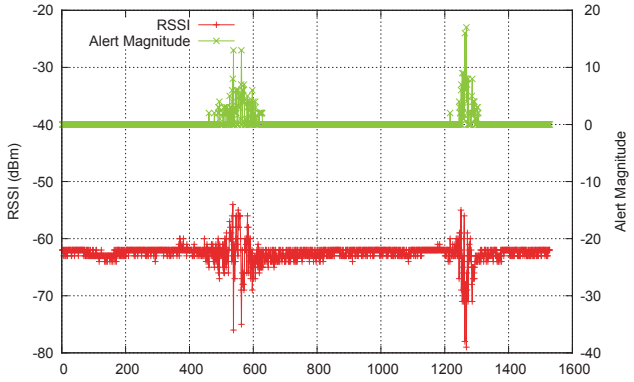


Fig. 6. Example of recorded RSSI and alert magnitude

of n samples with the highest values is collected. The value of n is analysed further in section IV. The set includes a guaranteed minimum number of n sample, but can include an even greater number, when more samples with the same high alert magnitude exist. In this case all samples are added to the set. Using a set of samples reduces the effect of the sample with the maximum alert magnitude being an outlier and thus unfavourable influencing the result.

Since the input data has been reduced to a selective subset of the complete data stream, the curve fitting and cross correlation techniques cannot be used. However, most methods estimating the velocity from the highest alert magnitude value can also be applied to this slightly enlarged set of samples.

1) topMaxAlertM: The topMaxAlertM approach uses the first occurring sample of the set with the n highest values. This could be realized by either selecting the first occurring sample with the highest alert magnitude, or the first occurring sample of the complete sample set. However, the first case is identical to the maxAlertM approach, in the second case a timestamp from the border of the event is selected, not representative for the peak.

Because of the above mentioned issues, the topMaxAlertM approach is not further considered in the evaluation.

2) topMedianAlertM: The topMedianAlertM approach selects the middle sample of collected set. This sample does not necessarily have to be the one with the highest value.

3) topAvgAlertM: The topAvgAlertM technique uses the timestamps of all samples in the collected set to estimate the time of the peak by computing the average time, again loosening the restriction of the precise timestamp.

4) topLinRegAlertM: topLinRegAlertM models the linear regression using the slightly larger data set, again with input from multiple links.

IV. Evaluation

For the evaluation of the different techniques we perform an experiment with a test set-up, as seen in Fig. 5. The set-up consists of ten Crossbow telosB sensor nodes, five on each side of a 2m wide and 6m long detection area. The motes are set in an interval of 1.5m on the same side. Radio links are formed between the motes directly opposite of each other in a way, that five consecutive links are generated across the space. During the experiment the links exhibited a mean RSSI of -62.5, -57.9, -58.3, -64.5 and -60.26 respectively from link one to link five. The motes use the TI MSP430 microcontroller and the CC2420 transceiver chip. They are sending on a frequency of 2.4 GHz and use a TDMA scheme with a 62.5 millisecond cycle time to avoid collisions. This value is chosen as a trade-off between energy consumption and detection accuracy. For our analysis all sent messages are recorded as samples using the IRIS tool introduced in [17]. IRIS is an experiment management tool, that allows for simultaneous data collection and visualisation. Samples collected can be analysed using provided or self-written functions and can be saved for further processing.

To evaluate the accuracy of the velocity estimation of a moving target, we compute the average error respective to the actual velocity. Two measurements with 12 runs each are performed in an lab environment. The error values are calculated as an average of the 12 runs for each measurement. During one experiment run, the target is entering the lab, walking crossing the links through the detection area once and then leaves the room. For the purpose of this evaluation, we assume that the target is walking with constant speed, does not suddenly change direction and walks directly in the middle of the detection area in a straight line. In the first measurement the target is walking with a speed of 0.6m/s, in the second measurement with 1.35m/s. The behaviour of the RSSI and the corresponding alert magnitude can exemplary be seen in Fig. 6. The figure shows a target crossing a link first with the slower, then with the faster speed. Markings on the floor indicating the step interval and a metronome to time the steps are used to help the target maintain constant speed, while still walking normally. Also, to not be reliant on the human perception of time, the precise moment of the LoS crossing is recorded as ground truth, using a laser based system.

The results of the measurements are summarized in Table II and Table III. Table II contains the average error in % and its Sample Standard Deviation. Column one and two list the values for the measurements with the slower speed of

TABLE II
AVERAGE ERROR AND SAMPLE STANDARD DEVIATION OF THE VELOCITY ESTIMATION

	Technique	Measurement Velocity 0.6m/s		Measurement Velocity 1.35m/s	
		Average Error (%)	Sample Standard Deviation	Average Error (%)	Sample Standard Deviation
1	minRSSI	16.77	13.40	11.15	6.82
2	medianRSSI	17.25	13.97	11.50	6.91
3	avgRSSI	16.18	13.94	11.33	6.84
4	linRegRSSI	12.79	10.95	11.03	7.77
5	CurveFittingRSSI	17.10	23.15	27.00	20.40
6	CrossCorrelationRSSI	83.99	1.64	86.98	1.59
7	maxAlertM	12.03	12.15	11.59	8.34
8	medianAlertM	11.41	12.57	12.16	8.16
9	avgAlertM	11.72	12.34	11.88	8.22
10	linRegAlertM	7.00	6.21	13.21	15.91
11	CruveFitting	12.03	12.15	11.59	8.34
12	CrossCorrelation	5.04	3.40	7.17	4.04
13	topMedianAlertM$_5$	4.12	3.83	6.16	5.04
14	topAvgAlertM$_5$	4.12	4.30	3.82	4.16
15	topLinRegAlertM$_5$	21.73	13.32	22.76	23.45

0.6m/s, column three and four the values for the faster speed of 1.35m/s. Each row shows the results for a different technique. Table III contains the results of the measurement runs with the maximum error for each approach and the Squared Mean Error, also for both measurements and velocities.

A. RSSI

The results of the measurements using the minimum RSSI data stream, minRSSI, can be found in the first row of Table II, for both slower and faster speed. Estimating the velocity from the first found sample selected by the minRSSI has an average error of 16.8% with a sample standard deviation of 13.4 for the slow velocity measurement and 11.2% error with a deviation of 6.8 for the faster velocity. This corresponds to an absolute error in the estimation of 0.1m/s for a speed of 0.6m/s and an absolute error of 0.15m/s for a speed of 1.35m/s.

Achieving a high accuracy using this technique turns out to be easier when detecting a target walking in a faster pace. The duration of the LoS crossing event is shorter, involving fewer samples and a narrower time interval, as seen in Fig. 6.

Using the median of all samples with the minimum RSSI (medianRSSI) or the average of those (avgRSSI) does not improve the performance significantly. The deep-fade link behaviour is ignored, which leads to a less accurate event detection and higher error values.

The most accurate approach according to our results, when working with pure RSSI values, is computing a linear regression. linRegRSSI shows a slightly better performance in the measurement with slower speed, and average performance, when the target is moving faster. Since multiple links are used in this approach, irregularities of one link can be evened out by the others.

Modelling the stream of RSSI values with a Gaussian function in the CurveFittingRSSI approach is unfavourable compared to the other techniques. Since the RSSI varies around a baseline, creating both positive and negative spikes in the case of an event, CurveFittingRSSI has a very high sample standard deviation. This causes very inaccurate behaviour.

Comparing two RSSI data streams using a cross correlation is possible, but since the similarity of two links is compared, the original input data needs to be similar. The unsteady and fluctuating nature of the RSSI with a high variance during the crossing event prevents this. Also, the deep-fading and anti-fading behaviour of the RSSI creates completely different characteristics, that cannot directly be compared. The resulting error values are constantly above 80% in both our measurements and have a very small Sample Standard Deviation, achieving in the worst performance of all tested RSSI-based methods.

B. Alert Magnitude

1) maxAlertM, medianAlertM and averageAlertLevel: Using the alert magnitude as a basis for the detection achieves better results for the slow velocity measurement and similar results to the RSSI-based techniques for faster speeds. Exploitation of the RSSI variance, and by doing so coping with the deep-fade qualities of the links, is the cause of this improvement. The average estimation error of maxAlertM can be seen in row seven of Table II, the results for medianAlertM and averageAlertLevel are in row eight and nine. All three techniques perform similar to each other, with error values around 11.5% $\pm$ 1%, which is an improvement of about 4% towards the RSSI-based techniques in the 0.6m/s measurement.

2) linRegAlertM: Using the linear regression has an improvement of 5% when applied to the data of the slow velocity measurement. However, linRegAlertM shows worse performance with the data of the faster velocity measurement. The average error is 13.2%, which is 2% worse than when based on the RSSI.

TABLE III
MAXIMUM AND SQUARED MEAN ERROR OF THE VELOCITY ESTIMATION

	Technique	Measurement Velocity 0.6m/s		Measurement Velocity 1.35m/s	
		Maximum Error (%)	Squared Mean Error	Maximum Error (%)	Squared Mean Error
1	minRSSI	51.89	460.85	28.07	170.74
2	medianRSSI	51.89	492.89	28.07	179.94
3	avgRSSI	51.89	455.93	28.07	175.09
4	linRegRSSI	39.45	283.59	26.85	181.98
5	CurveFittingRSSI	85.23	828.11	69.77	1144.93
6	CrossCorrelationRSSI	87.15	7057.48	91.27	7567.14
7	maxAlertM	46.85	292.43	32.73	203.99
8	medianAlertM	46.856	288.19	32.73	214.34
9	avgAlertM	46.85	289.73	32.73	208.78
10	linRegAlertM	22.71	87.58	59.34	427.44
11	CruveFitting	46.85	292.43	32.73	203.99
12	CrossCorrelation	11.36	36.88	14.71	67.68
13	topMedianAlertM$_5$	12.77	31.61	17.74	63.30
14	topAvgAlertM$_5$	18.12	35.53	15.87	31.94
15	topLinRegAlertM$_5$	49.03	649.73	71.50	1067.68

Since the target is moving with a higher speed, the time between the LoS crossing of the links is shorter. This has an impact on the slope of the linear regression line. It is more steep and outliers have a higher influence since fewer samples are sent, explaining the behaviour in the faster velocity measurement.

3) Curve Fitting: Curve fitting is performed using all values of the respective data streams, without previous filtering for minimum RSSI or maximum alert magnitude values. Fitting a curve through the alert magnitude values achieves similar results to maxAlertM, medianAlertM and avgAlertM, as seen in row 11 of Table II. The average error is at 12% for the first measurement and at 11,6% for the second.

The results of the alert magnitude based CurveFitting achieve a higher accuracy than fitting a curve to the RSSI data stream. The steady nature of the alert magnitude, which is only interrupted in the case of a detection event, positions the maximum of the Gaussian function close to the peak of the LoS crossing. Still, the results are on the same accuracy level as maxAlertM, medianAlertM and averageAlertLevel.

4) Cross Correlation: The cross correlation computes the correlation coefficient coef$_{cc}$ between two links. To utilize the approach and estimate the velocity of a target, all possible link combinations for both measurements are analysed first. The results of this test are shown in Fig. 7. There, the average error is plotted over the distance of the links towards each other. A clear trend is visible, showing that the estimation becomes more accurate, the further the links are apart. The overall error is being reducing and the minimum and maximum error values are closer to the average. The messages for the experiment are being sent with a frequency of 62.5 milliseconds and it takes the target a certain time to cross the distance between two links. The larger the distance, the longer is the time needed to cross it and the more samples can be sent in the duration,

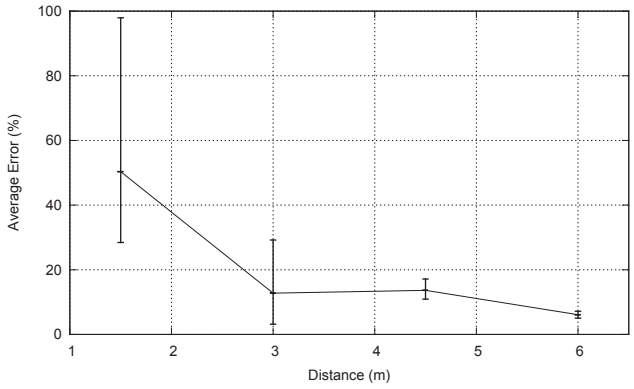


Fig. 7. Average of the velocity estimation error over the link distance

explaining the increase in the accuracy. For this reason the cross correlation between the two links farthest apart is used for the comparison with the other techniques.

As mentioned in Section IV-A, estimating the velocity using the cross correlation of the RSSI values has a very low accuracy, because of the RSSI fluctuations. However, applying it on the alert magnitude is very promising. The average error is shown in row 12 of Table II. It is at 5.0% with a sample standard deviation of 3.4 for the slow velocity measurement. For the fast velocity measurement these values are at 7.2% average error with 4.0 sample standard deviation.

These results surpass the previous techniques using the maximum alert magnitude. Since the alert magnitude values are zero except during an event, as seen in Fig. 6, the correlation coefficient reaches its maximum when the events on two different links are directly overlapping. This achieves a precise approximation of the time between the LoS crossing events on two compared links used in the velocity estimation.

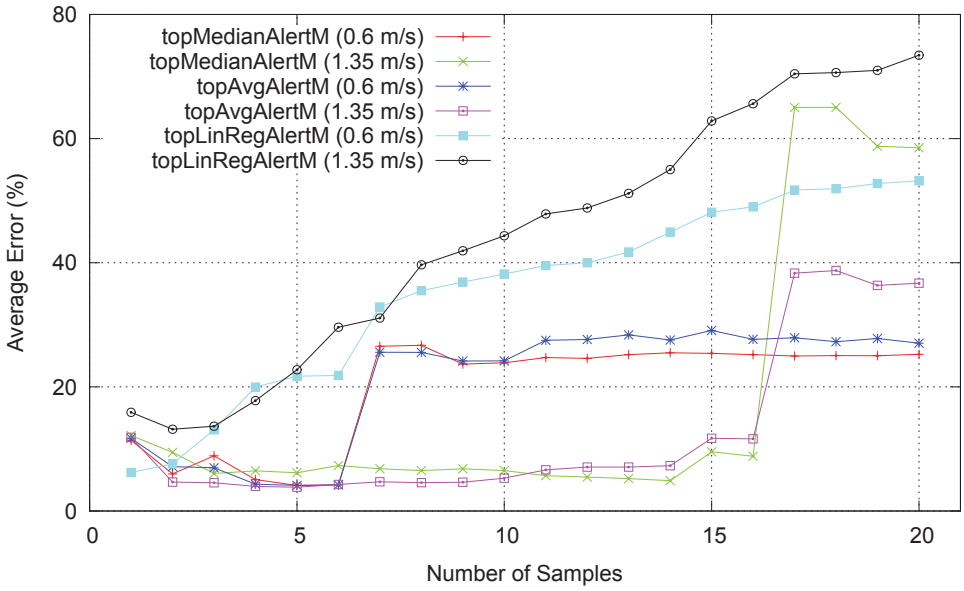


Fig. 8. Average velocity estimation errors with different sizes of alert magnitude sets

C. Alert Magnitude Set

Instead of using the sample(s) with the maximum alert magnitude, using the set of the n highest alert magnitude values can achieve a very different result. The set provides a more balanced picture of the event, since it includes all of the most interesting samples and not just the highest, reducing the influence of outliers. Unfortunately, this is not beneficial to all methods. The CurveFitting and CrossCorrelation techniques use all the samples from the complete measurement and cannot be applied on a subset. Furthermore, since the sample with the maximum alert magnitude is still in the set of the highest n samples, using the topMaxAlertM approach will not have a different outcome than maxAlertM, as explained in section III-C1. However, for the techniques of topMedianAlertM, topAvgAlertM and topLinRegAlertM the improved sample set is applicable.

The influence of the size of the sample set is analysed in Fig. 8, showing the average error over the number of samples. For each of the remaining three techniques all sample set sizes from 1 to 20 were examined. The set size of 1 equals the respective approach based on the maximum alert magnitude value. The figure displays six graphs, two for each technique, split by the velocity of the measurement. To further differentiate between the techniques an index i is introduced, indicating the size of the sample set. topMedianAlertM$_3$, for example, would describe the topMedianAlertM approach with a sample set size of 3 samples.

The analysis shows a high improvement of the topMedianAlertM and the topAvgAlertM approach for the measurement with a velocity of 0.6m/s, when increasing the sample size to up to 6 samples in the set. The lowest error is achieved by topMedianAlertM$_5$ with an average error of 4.12%, directly followed by topAvgAlertM$_5$ with an average error of 4.124% for the slower velocity. This is an accuracy of around 95.8%, estimating the velocity of the target. These improvements are due to the fact that both techniques benefit from the larger selection of possible samples describing the peak of the crossing event.

For the fast velocity measurement of 1.35m/s the topAvgAlertM$_5$ approach achieves the lowest average error of 3.82%, giving it an estimation accuracy of 96.1%. For the topMedianAlertM techniques topMedianAlertM$_1$4 has the lowest error with 4.88%.

Additionally, Fig. 8 indicates a trade-off between too few or too many samples in the sample set. While the performance of the topMedianAlertM and topAvgAlertM techniques initially improves when adding samples to the set, the average error abruptly increases after a certain threshold. For the slow velocity measurement this happens in our experiment at sample set size 7, for the fast velocity measurement at sample set size 17.

The reason for this is the inclusion of samples in the set which are not representative for the peak. These sample cause a reduction in precision. The effect occurs, when the number of samples with unique values is not very large. If only a few samples with high alert magnitude values exist, all of them will be included into the sample set. However, when raising the set size, more samples will need to be added. Eventually, some samples with only moderate alert magnitude values will be in the sample set. Those are less likely to describe the detection event. As soon as that happens, the accuracy of the alert magnitude set based techniques diminishes.

Table II summarizes the results for topMedianAlertM$_5$ in row 13, and topAvgAlertM$_5$ in row 14.

Since more samples are guaranteed to be in the sample set, the linear regression loses much of its accuracy. Fig. 8

shows: The more samples are added to the set, the worse the topLinRegAlertM technique performs. This deterioration occurs, because the values of all samples are included in the computation. The average error drops from 7% (linRegAlertM) to 21,7% (linRegAlertM$_5$) in the slow velocity measurement and from 13,2% to 22,8% in the fast velocity measurement. The results for linRegAlertM$_5$ are listed in Table II, row 15.

The evaluation shows that the topMedianAlertM$_5$ and the topAvgAlertM$_5$ techniques are the best choices for accurately detecting the peak of an event and estimating the velocity of a target. Furthermore, they are also less computational expensive then a linear regression, the curve fitting technique or the cross correlation.

V. Conclusion & Future Work

In this paper we experimentally evaluated 15 techniques in order to estimate the velocity of a target moving through a set of radio links. The precise moment of the Line of Sight crossing is determined based on the collection of RSSI samples. The techniques are separated into groups, based on their input values. Analysing the average of the minimum RSSI attains less accurate results than focusing on the maximum alert magnitude or the set of the *n*-highest alert magnitude samples. This is caused by the RSSIs unstable character.

Three techniques were found delivering unsuitable results below 80% accuracy, nine techniques performed moderately, three techniques achieved a performance above 90% estimation accuracy. Performing a cross correlation of the links with the highest distance from each other achieves good results in estimating the velocity. The accuracy for slow moving targets is around 95.0%, for fast moving targets it is around 92.9%. Furthermore, not focusing on the samples with the highest absolute value, but on the set of highest values offers a layer of protection against outliers and improves simple median or average computing techniques. Selecting the median timestamp of the set of samples with the highest detection value achieves 95.8% accuracy for slow moving targets and 95.1% for fast moving ones. Computing the average of those samples also achieves 95.8% accuracy for slow moving targets and 96.1% accuracy for fast ones.

In the future we want to extend our research by using the velocity estimation to differentiate between multiple targets in the same radio environment.

References

[1] K. Woyach, D. Puccinelli, and M. Haenggi, "Sensorless sensing in wireless networks: Implementation and measurements," *2006 4th International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks, WiOpt 2006*, 2006. doi: 10.1109/WIOPT.2006.1666495. Available: http://dx.doi.org/10.1109/WIOPT.2006.1666495

[2] S. Hussain, R. Peters, and D. Silver, "Using received signal strength variation for surveillance in residential areas," *SPIE Defense ...*, 2008. Available: http://proceedings.spiedigitallibrary.org/proceeding.aspx?articleid=837114

[3] J. Wilson and N. Patwari, "Radio tomographic imaging with wireless networks," *Mobile Computing, IEEE Transactions on*, vol. 9, no. 5, pp. 621–632, 2010. Available: http://ieeexplore.ieee.org/xpls/abs{_}all.jsp?arnumber=5374407

[4] O. Kaltiokallio and M. Bocca, "Real-Time Intrusion Detection and Tracking in Indoor Environment through Distributed RSSI Processing," *2011 IEEE 17th International Conference on Embedded and Real-Time Computing Systems and Applications*, pp. 61–70, aug 2011. doi: 10.1109/RTCSA.2011.38. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6029830

[5] M. Kochláň, J. Miček, and P. Ševčík, "2.4ghz ism band radio frequency signal indoor propagation," in *Proceedings of the 2014 Federated Conference on Computer Science and Information Systems*, ser. Annals of Computer Science and Information Systems, M. P. M. Ganzha, L. Maciaszek, Ed., vol. 2. IEEE, 2014. doi: 10.15439/2014F299 pp. pages 1027–1034. Available: http://dx.doi.org/10.15439/2014F299

[6] J. Yang, Y. Ge, and H. Xiong, "Performing Joint Learning for Passive Intrusion Detection in Pervasive Wireless Environments," ... , *2010 Proceedings IEEE*, 2010. Available: http://ieeexplore.ieee.org/xpls/abs{_}all.jsp?arnumber=5462148

[7] M. Bocca, O. Kaltiokallio, N. Patwari, and S. Venkatasubramanian, "Multiple target tracking with rf sensor networks," *IEEE Transactions on Mobile Computing*, vol. 13, no. 8, pp. 1787–1800, 2014. doi: 10.1109/TMC.2013.92. Available: http://dx.doi.org/10.1109/TMC.2013.92

[8] F. Adib, Z. Kabelac, D. Katabi, R. C. Miller, I. Nsdi, and Z. Kabelac, "Multi-Person Localization via RF Body Reflection," *Usenix Nsdi*, 2014. Available: https://www.usenix.org/conference/nsdi15/technical-sessions/presentation/adib

[9] T. Li, Y. Wang, L. Song, and H. Tan, *Wireless Sensor Networks: 12th European Conference, EWSN 2015, Porto, Portugal, February 9-11, 2015. Proceedings*. Cham: Springer International Publishing, 2015, ch. On Target Counting by Sequential Snapshots of Binary Proximity Sensors, pp. 19–34. ISBN 978-3-319-15582-1. Available: http://dx.doi.org/10.1007/978-3-319-15582-1_2

[10] J. Wilson and N. Patwari, "See-through walls: Motion tracking using variance-based radio tomography networks," *Mobile Computing, IEEE Transactions on*, vol. 10, no. 5, pp. 612–621, 2011. Available: http://ieeexplore.ieee.org/xpls/abs{_}all.jsp?arnumber=5582100

[11] O. Kaltiokallio, M. Bocca, and N. Patwari, "Follow @grandma: Long-term device-free localization for residential monitoring," *Proceedings - Conference on Local Computer Networks, LCN*, pp. 991–998, 2012. doi: 10.1109/LCNW.2012.6424092. Available: http://dx.doi.org/10.1109/LCNW.2012.6424092

[12] J. Wilson and N. Patwari, "A Fade-Level Skew-Laplace Signal Strength Model for Device-Free Localization with Wireless Networks," *IEEE Transactions on Mobile Computing*, vol. 11, no. 6, pp. 947–958, jun 2012. doi: 10.1109/TMC.2011.102. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6188339

[13] N. Kassem, A. E. Kosba, and M. Youssef, "RF-Based Vehicle Detection and Speed Estimation," *2012 IEEE 75th Vehicular Technology Conference (VTC Spring)*, pp. 1–5, may 2012. doi: 10.1109/VETECS.2012.6240184. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6240184

[14] O. Karpis, "Sensor for vehicles classification." in *FedCSIS*, 2012, pp. 785–789. Available: https://fedcsis.org/proceedings/2012/pliks/215.pdf

[15] O. Kaltiokallio, "Intrusion Detection Based on Embedded Processing of Received Signal Strength Indicator," Master's thesis, Aalto University, 2011.

[16] D. Zhang and L. M. Ni, "Dynamic clustering for tracking multiple transceiver-free objects," *2009 IEEE International Conference on Pervasive Computing and Communications*, pp. 1–8, mar 2009. doi: 10.1109/PERCOM.2009.4912777. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4912777

[17] R. Figura, M. Ceriotti, C.-Y. Shih, M. Mulero-Pâzmâny, S. Fu, R. Daidone, S. Jungen, J. J. Negro, and P. J. Marrôn, "Iris: Efficient visualization, data analysis and experiment management for wireless sensor networks," *EAI Endorsed Transactions on Ubiquitous Environments*, vol. 14, no. 3, 11 2014. doi: 10.4108/ue.1.3.e4. Available: http://dx.doi.org/10.4108/ue.1.3.e4

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1067–1074
DOI: 10.15439/2016F154
ACSIS, Vol. 8. ISSN 2300-5963

RF-Tania protocol and system architecture for location based sensor measurements

Soultana Ellinidou
Dept. of Informatics & Telecommunication Engineering
University of Western Macedonia
Kozani, Greece
t.ellinidou@gmail.com

Sotirios Kontogiannis
Dept. of Mathematics
University of Ioannina
Ioannina, Greece
skontog@gmail.com

George Kokkonis
Dept. of Applied Informatics,
University of Macedonia
Thessaloniki, Greece
gkokkonis@uom.gr

***Abstract*—In this paper, we propose a new protocol for sensors Frequency Shift Keying data transmission named RF-Tania protocol framework. This protocol focuses on Energy Efficiency and it is based on an existent open source protocol stack. Tania protocol framework includes a set of three protocols: The Sensors Data Transmission Protocol (Tania-SDTP), the Ad-Hoc On Demand Protocol (Tania-AHOD) and the Ad-hoc Alerts Protocol (Tania-AHA), each one servicing different user functionalities.**

Tania protocol is used on a specific system architecture of a crowded sensors network. Such network is composed of two different types of equipment: The measuring sensor transponders and the Ad-hoc receivers. Sensor transponders are battery operated systems with installed sensors. Ad-hoc receivers are mobile phones equipped with transponders that operate as both sensor output devices as well as sensor data re-transmitters.

In the case study, a test bed system architecture for monitoring CO_2 and temperature levels has been used. In this study, tests of energy endurance, RF coverage and energy performance have been performed. From the evaluation results, it is proven that RF-Tania protocol outperforms in terms of energy existing RF protocols that use periodic broadcast data transmissions.

Keywords—Wireless Sensor Networks, RF communication, Energy Efficient Operation

I. Introduction

ENERGY efficiency is a fundamental issue concerning the design of communication protocols for wireless sensor networks (WSNs). Hence, the proposed protocol focuses on that key objective.

Another important attribute of an RF sensor network is the relaxed mobility of sensors' monitoring clients. While the term mobility is a well-defined term that signifies the communication means and interoperability between clients and wireless sensors, we define the term "*relaxed*" as a new term. The term "relaxed" determines a wireless communication state where sensors exchange information with one another; intermediate gateways traverse sensor information to allotted servers and clients act both as sensors' data repeaters and as actual monitoring devices.

Based on these concepts a new RF protocols suite for data transmissions has been implemented. This set of protocols tries to maintain both Energy efficiency by reducing utilization of the sensors' RF transponders and MCU, while pertaining the characteristics of a relaxed mobility network architecture.

RF-Tania protocol is divided into three sub-protocols: The WSN sensor protocol for the broadcasting of sensor data, the WSN client protocol, for the reception and acknowledgement of sensor protocol messages and the WSN alerts protocol for the transmission of sensor critical information. The major function of RF-Tania is to save energy by putting the sensor transmitter into sleep-mode, when the sensors measurements remain constant between two or more consecutive periods. As far as the transmitted measurements are concerned, the RF-Tania protocol forces the receivers to handle frame re-transmission, whenever the sensor measurement remains at a constant level (Sensors Data Transmission Protocol, Tania-SDTP).

The mobile user requests a sensor's measurement through a mobile phone client application. If the sensor is in sleep mode, the nearby clients will inform the client that issued the request about the last sensor measurement (Ad-Hoc On Demand Protocol, Tania-AHOD).

Ultimately, Tania-RF protocol is capable of informing all WSN nodes whenever sensor measurement or measurements correlation function is higher or equal to a threshold value. In such alert cases the transmitter will immediately send an alert-message to all connected receivers in the network (Ad-Hoc Alerts Protocol, Tania-AHA).

RF-Tania protocol is comprised of 3-layer according to the OSI model. The physical layer includes a Frequency-Shift Keying (FSK) data transmission. The Data link layer is consisted of a MAC layer of node broadcast frames called Jeelab frames [1]. The network layer includes the RF-Tania protocols.

This paper structure is organized as follows: In section II we discuss the related work, in section III we present the proposed protocols suite, in section IV we provide the implementation of protocol and in section V we conclude to

our contribution and present a future system and protocol modifications.

II. Related Work

There is a wide number of WSN energy efficient protocols seen in the literature. A comprehensive classification and survey on this topic presented in [2]. These protocols work on the assumption that energy is limited and exhaustible. Consequently, the effort is primarily shifted towards prolonging the network lifetime. However, with energy harvesting capability [28], there is a need of fresh perspective on protocols and network design. Specifically, in the scenario of RF energy transfer, the protocol proposed in [6], and its subsequent analytical model in [7], adopts a duty-cycle based on the proportion of harvested energy.

The authors in [3] proposed a Medium Access Control (MAC) protocol, called RF-MAC, which ensures optimal energy delivery to the requesting node. In RF-MAC, a node broadcasts its request for energy harvesting (RFE) frames containing its ID, and then waits to hear for the Energy feeding Transmitters (ETs) in the neighborhood.

In [4], the authors proposed a network architecture which consists of two types of RF sensors. One class of sensors harvests RF energy on the DTV band (614MHz), while another uses the 915MHz ISM band. The energy transfer stage begins when the ET sends out the Request to Charge (RTC) packet, at 915MHz (operating frequency of Mica2 mote), offering to transfer wireless energy to the sensors in the network. Consequently, both types of devices can hear the RTC packet sent by the ET. The sensors that received the RTC then acknowledge this packet by sending back a response, called energy pulses. Once the ET receives the energy pulses from the responding sensors, it estimates the average power that sensors will receive during the charging process.

In [7] the authors implemented the WirelessHart (Highway Addressable Remote Transducer), which is a digital protocol for two-way communication between a host application and smart field instruments, providing access to diagnostics, configuration and process data. It specified a physical layer which used FSK to superimpose digital communication signals at a low level of 4-20mA. It supports two types of networks: Point to point and multi-hop network. WirelessHart is an RF protocol used by industrial vendors such as ABB and OMRON (PLC). It prevents message collisions using a TDMA use of the radio channel on predefined time slots instead of sensing the medium to transmit (CSMA/CA-CSMA/CD). However, its adoption to industrial application limits the feasibility for use in commercial and residential application due to its increased deployment cost [10].

Synkro RF network specification was developed by Freescale [11]. Synkro is mainly implemented for ad-hoc and low cost sensors. It offers a maximum transmission rate of 250kbit/s, three independent channels in the 2.4GHz band and two network node types: a controller and controlled nodes. The main weakness over its PAN alternatives is the maximum number of 32 controlled nodes per controller (similar to Jeelab implementation). Moreover, the SMAC protocol [6] is used for developing proprietary RF transceiver applications using a Freescale 802.15.4 transceiver. It supports point to point communications by having a very-low power, proprietary, bi-directional RF communication link between nodes. Freescale is porting its protocol and network specification into home appliances such as DVDs and TVs offering a two-way communication alternative, in order to replace existing infrared technologies.

Bluetooth Low Energy is a 2MHz BW protocol at 2.4 GHz and uses TDMA as a medium access mechanism (IEEE 802.15). It uses a low to medium power transmission of -20 up to 10 dBm in comparison with Bluetooth Class 1/2/3 accordingly [12, 13, 16]. ZigBee [15] is a Bluetooth alternative protocol created for industrial applications. It uses a variety of IEEE 802.15.4 [14] standard and the same frequency band with Bluetooth and BLE. ZigBee is more of a mesh wireless protocol rather than a P2P protocol such as BLE [16].

The RF4CE [5, 8] is a convenient, low-cost, low-power wireless transmission protocol used both by the ZigBee company and the RF4CE association. It is built from the standard specs of the networking layer and the application layer is implemented by the IEEE 802.15.4. It differs from ZigBee, as it does not have a complicated Internet routing protocol or multiple transmission communication mechanisms. It operates in the 2.4GHz frequency band utilizing three RF channels with a total BW of 2MHz [8].

WiMedia Ultra-WideBand (UWB) radio platform uses frequencies from 3.1 GHz - 4.85GHz, or at 6.2GHz - 9.7GHz [18]. WiMedia frames may be sent in either unicast or broadcast fashion. Unicast frames are directed to their destination based on a 16-bit device address; certain addresses are also reserved for broadcast groups. WiMedia Ultra Wide Band (UWB) is a short range high throughput technology using Tx Power of (12-20dBm) covering a maximum range of 10-20m with speeds up to 500Mbit/s. It is commonly used for multimedia applications.

Aside from the WiMedia implementation, an extension of IEEE 802.15.4, IEEE 802.15.4a-2007 uses the unlicensed Gigahertz bands. The standard provides two physical protocols the UWB PHY and Crisp Spread Spectrum – CSS PHY. UWB provides high data rate channels in three distinct unlicensed GHz bands, providing speeds of 110Kb/s up to 26Mb/s and maximum range of 100m, with the optional feature of precision ranging(used for indoor positioning) [19]. CSS PHY on the other hand, is a low power consumption, low range and low data rate alternative used at the 2.4GHz band for sensor data transmissions up to 1Mb/s [19, 20, 22].

III. PROTOCOLS SUITE FOR SENSORS DATA TRANSMISSION

The main purpose of the RF-Tania protocols suite is to save energy by reducing the utilization of the sensors' RF transponders and MCU, while maintaining relaxed mobility network architecture.

Our protocols suite has been implemented on the RF12 driver, used for the RFM12B JeeNode module [1]. The RF12 driver operates as follows:

- Nodes can only communicate with each other if they are in the same "net group" (1-250)
- Nodes in the same net-group have a unique ID(0..31)
- Frames (Jeelab frames) are 0-66 bytes long
- Frames utilize extra 9 bytes of overhead, including the preamble
- Data is sent at a maximum rate of 45-50Kbit/s

The RF-Tania protocol stack consists of 3-layers, according to the OSI model: Physical layer, Data link layer, Network layer. Each one of these layers is explained below.

A. Network Architecture

Network architecture consists of two entities, the transmitter and the receiver, as we can see in Fig 1. Specifically, the transmitter or sensor entity contains RF-transponder and microcontroller equipped with sensors. The receiver or mobile client entity contains mobile phones equipped with RF-transponders.

The transmitter, called sensor entity, is placed on a stable position transmitting the sensors measurements to the mobile receivers. The mobile receivers, called client receivers, are moving around. The sensor entity transmits data every period T_p to all mobile receivers, located in the coverage area.

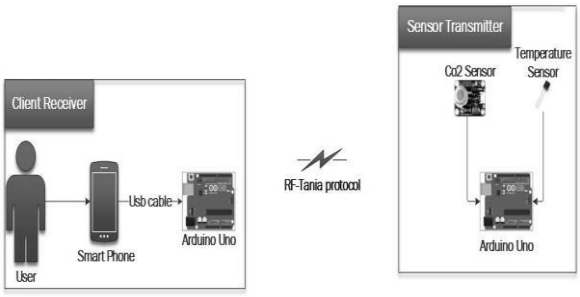


Fig.1: WSN receiver and sensor entities

In the WSN network we use a series of RFM12B transponders. Technical characteristics of RFM12B chip are outlined at [23] and [24] for the long-range transceivers.

B. Physical Layer

The RFM12B initialization includes the communication transponder used for our WSN network. The physical layer includes the following configuration fields: Configuration Settings, Power Management, Frequency Settings, Data Rate and the Receiver Control.

As far as the configurations settings are concerned we used the band of 433MHz, in order to minimize interference with other transmissions in the area (GSM frequencies, Wi-Fi, Bluetooth and GPRS data transmissions). Furthermore, we can activate the RX Register and TX FIFO Buffer. These chip configuration settings parameters are set from the Arduino microcontroller (sender or receiver entity) via the SPI (Serial Peripheral Interface), sending 16bit commands to the RFM12B chip.

The parameters control the power to the RFM12 submodules and allow the selection of circuits, until the RFM12 is turned on or turned off accordingly. So by disabling these circuits, when it is not required, we can control the amount power consumption of the device.

The RFM12B chip has the ability of frequency hopping-shifting to nearby frequencies. The default carrier frequency shifting is 90 KHz. If there is an interference on one frequency, another frequency can be selected manually and interference problems can be avoided.

The RF12B incorporates a fully integrated Power Amplifier (PA) with antenna tuning and a Low Noise Amplifier (LNA) with switchable gain. The Power Amplifier (PA) has an open-collector differential output and can directly drive a loop antenna with a programmable output power level of maximum of 0dBm transmission power.

An automatic antenna tuning circuit is built in to avoid costly trimming procedures and the so-called "hand effect". The Low Noise Amplifier (LNA) has approximately 250Ohm input impedance, which functions well with the proposed $\lambda/4$ monopole antennas (of 17.3cm length). If the RF input of the chip is connected to 50Ohm devices, an external matching circuit will be required to provide the correct matching and to minimize the noise figure of the receiver. The LNA gain can be selected in four steps (between 0 and -20dBm) according to RF signal strength required.

The Data Rate command sets the bitrate of the transmitted data or the expected bitrate of the received data. The actual bit rate in transmit mode and the expected bit rate of the received data stream in receive mode are determined by the 7-bit parameter R (bits r6 to r0) and bit cs. The highest data rate is set at 57.4Kbit/s. This gives 17usec per bit transmission and 139usec per byte (7192Kbyte/s or 57,553Kbit/s).

The Receiver Control command contains a series of various bits, according to [25]:

Bit 10 (P20): sets the function of INT/VDI pin on the RFM12 module. It configures the module as input (Interrupt from MCU) or output (VDI Valid Data Indicator).

Bits 9-8 (d1 to d0): VDI (valid data indicator) signal response time setting

Bits 7-5 (i2 to i0): Receiver baseband bandwidth (BW). The receiver bandwidth is selectable by programming the bandwidth (BW) of the baseband filters. The bandwidth settings are linked to both data rate, and Tx modulation

commands. When data rate is fast a higher receiver bandwidth is required. The default baseband BW of the RF12B is set to 134 KHz. The highest receiver BW is set to 450 KHz [24].

Bits 4-3 (g1 to g0): LNA (Low Noise Amplifier) gain (dBm). Typically, 0dBm is the LNA output of the RFM12B chip. Consequently, the maximum transmission power of the transponder is 0dBm.

Bits 2-0 (r2 to r0): bits that set the RSSI detector threshold. The RSSI threshold is based on the 433MHz beacon carrier and expresses how strong or weak the transmitter beacon signal is (as spotted at the receiver end). If it is below a certain threshold (set by bits) the receiver will ignore the incoming transmitted frame.

C. Data Link Layer

The Data link layer includes the MAC layer, which is responsible for controlling how and when network devices gain access to medium and permission to transmit data. The nodes broadcast frames, called Jeelab frames constructed in this layer.

The Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) is a network multiple access method in which carrier sensing is performed. The nodes attempt to avoid collisions by transmitting data only when the channel is sensed to be idle. Moreover, a rule of 1% channel utilization applies at the data link layer. This means that each transmitting node in our network should send up to 1% of the time, on average. The 1% rule is a very simple collision avoidance mechanism. There is also a random back off collision avoidance mechanism included in Jeelabs library whenever a collision occurs in cases of frames retransmission. Both of these mechanisms are implemented on the rf12_easy transmit Jeelab's library function [1].

The original frame header, implemented by Jeelabs [1] is shown in Fig. 2 (Preamble, SYN, Head and CRC).The proposed protocol suite header field is encapsulated in a Jeelab frame, shown in Fig. 3 (Packet_ID, Node_Id Measurements and Send_time).

The Preamble and SYN fields are used for data transmission synchronization and group selection (Each group can contain up to 32 nodes with unique node ids). The Head field includes Jeelab protocol options and source or destination node id. The frame Payload is 0-56bytes. The CRC (Cyclic Redundancy Check) is the error-detecting code field.

In order to transmit frames, RF-Tania protocol uses the Head field of the original Jeelabs frame. There are three bits: C = CTL, D = DST, A = ACK and a 5-bit node ID. (Fig. 2 – Head field). Node id values can be 0-31. The A bit (ACK) indicates whether this frame wants to get an ACK back. The C bit needs to be zero in this case. The D bit (DST) indicates whether the node ID specifies the destination node or the source node. For frames sent to a specific node, DST = 1 (The destination node id is included in the Head field).

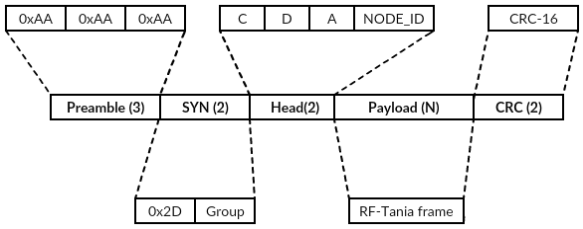


Fig 2: The Jeelab frame header [1]

For broadcasts, DST = 0, in which case the node ID field refers to the originating node. The C bit (CTL) is used to signify ACK request and should be combined with the A bit set to zero. The node receiving an ACK can check the originator of the ACK reply.

As for the Payload field of the Jeelab frame, we propose the format shown at Fig. 3. The first field of RF-Tania payload contains an auto incremented number of frame ID. The second field carries the ID of the sending node and it is 5 bits (3 bits left for future use possibly QoS or Collision experience provisions). The Measurements field includes 4 bytes sensor measurements and the Send_time field includes a 4 bytes timestamp of the frame. In case of sensor entity measurement, Send_time is represented by the MCU millis time.

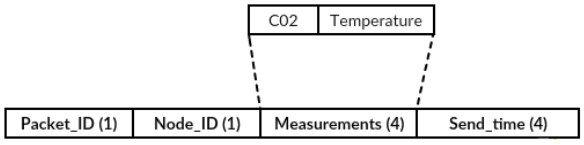


Fig. 3: RF-Tania frame

RF-Tania frame can include simultaneously 24 measurements of 2bytes each one (48 bytes overall) or 12 measurements of 4 bytes each one. In our case study we use 2 measurements of 4 bytes from a CO_2 and Temp sensor accordingly.

D. Network Layer

The network layer includes the RF-Tania protocol, which is categorized into 3 sub-protocols:

Tania-SDTP: According to SDTP protocol the receiver entity periodically collects sensor measurements from the sensor entities (Fig. 4). Each sensor entity collects measurements from a number of sensors connected to the entity. The sensor entity device has two periods: the sensor device data collection period (T_{sc}≈2s) and the sensor device network data transmission period T_p (initially T_p=30s). For a period T_p, all sensors measurements retrieved from a sensor entity, are stored in the sensor entity SRAM. At the end of each period T_p, the sensor entity calculates the mean sensor values using equation (1) and checks whether equation (2) is satisfied:

$$x_p = \frac{1}{n}\sum_{i=1}^{n} x_{Ci} \quad (1)$$

$$\left(\left|\frac{xc_i - xc_{i-1}}{dt}\right|\right)_{Tc} \geq 4\min\left(\left|\frac{dxp}{dt}\right|\right)_{Tp} \quad (2)$$

In the above functions, xc_i is one sensor measurement at time T_{sc} (i=1..n sensor measurements collected in a time interval T_p), x_{pi} (x_p)is the mean calculated sensor measurement ready for transmission, $xp_{i-1}$is the previously transmitted sensor measurement. If equation (2) is satisfied, the current sensor mean measurement will be considered of high variance and an alert will be set by the sensor entity using Tania AHA protocol. A sensor entity alert forces the sensor entity to perform a multiplicative decrease of its transmission period ($T_p=T_p/2$, $T_p>2s$).Transmission period value multiplicative decrease is performed in each transmission interval until equation (2) is satisfied. If equation (2) does not apply, the sensor entity will perform an increase of the transmission period T_p using the following equation: $T_p=T_p+1\times f_p$ (sec), where f_p is the sensor data transmission frequency coefficient.

In a previous data transmission interval T_p, a sensor entity collected n number of measurements (M) from a CO_2 sensor in period $T_{sc}<<T_p$. For the T_p period interval the maximum measured CO_2 sensor value is M_{max} and the minimum measured value is M_{min} (Equation (3)). For each transmission period, a frequency coefficient parameter f_p is calculated based on equations (3) and (4) and used for the determination of the next period T_p value.

$$cp = \frac{M_{max}^{Tp} - M_{min}^{Tp}}{\frac{1}{n}\sum_{i=1}^{n-1}[M_{Tsc}^{i} - M_{Tsc}^{i-1}]} \quad (3)$$

$$f_p = \frac{1}{cp} \quad (4)$$

M_{TSC} values are the sensor measurements per time intervals called ticks (T_{sc}).The denominator of equation (3) expresses the average M_{TSC} value for a period T_p. If $f_p<1$, f_p will be set to 1. Coefficient (c_p) is called energy efficient parameter. When its value decreased, the sensor will become less sensitive to measurement variations or sensor variations are minimum. Moreover, when the energy efficient parameter value decreased, the sensor values variations will be significant and at the same time the sensor entity will spend more energy for data transmission.

If equation (2) is satisfied, the sensor entity will send alerts to the receiver entities using Tania AHA alerts protocol. As soon as a frame is sent to the receiver, it should send back an ACK frame to the sensor entity (see Fig. 6).

Tania-AHOD: This is the on demand measurements request protocol. Specifically the client-receiver, which is connected to the users' mobile-phone, requests measurements from a sensor entity. In that case, either the sensor entity (if it is in listen mode-Fig. 4, Tania on demand request listen period) might respond to the request or another nearby receiver entity might respond. In case of reception of multiple responses by the receiver entity, the receiver will choose the sensor entity responses first. If nobody response, the receiver entity will response with the most current timestamp field in its Tania header frame.

Once the receiver makes a decision about the best frame to keep, then it will forward it to the mobile phone application layer. The requested measurement will appear on user mobile display (see Fig. 5).

Tania-AHA: This protocol is responsible for informing all receiver nodes for critical measurement values. Sensor alerts occur whenever sensor measurements satisfy the equation (2) or the sensor measurements are higher or equal to a threshold value.

For example, in a sensor network that monitors environmental CO_2 values, the sensor entity collects CO_2 sensor data. Furthermore, it is checking once for every period T_p whether equation (2) is satisfied or the mean sensor value for that period is above threshold value of 1000ppm for CO_2 (see Fig.6). If equation (2) is satisfied or mean sensor value is above threshold value then the sensor entity will immediately send an alert-message to all connected receiver entities in the WSN network using unicast transmissions (Alert frames will be sent to all sensor nodes that acknowledged last periodic sensor entity transmission using Tania-SDTP protocol). As soon as the alert-messages are sent, the receiver entities will open a serial connection with mobile phone, in order to inform the user (Fig. 6).

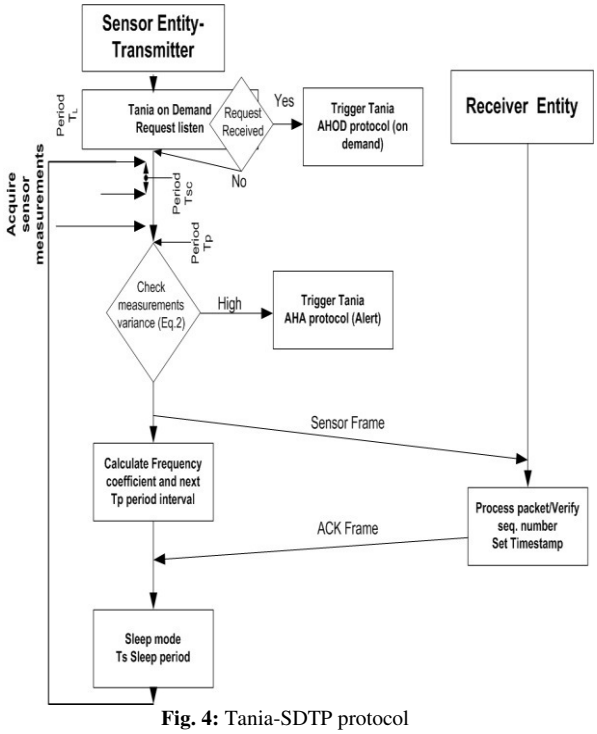


Fig. 4: Tania-SDTP protocol

The Tania AHA alerts transmission phase is persistent and energy consuming. If a receiver entity does not acknowledge the reception of the alert frame, the alert message transmission will be broadcasted and the alert broadcast will be repeated every 1sec for a period T_p equal to the period of 30sec (30 alerts back to back). This step is per-

formed until at least one receiver entity acknowledges the alert reception to the sensor entity.

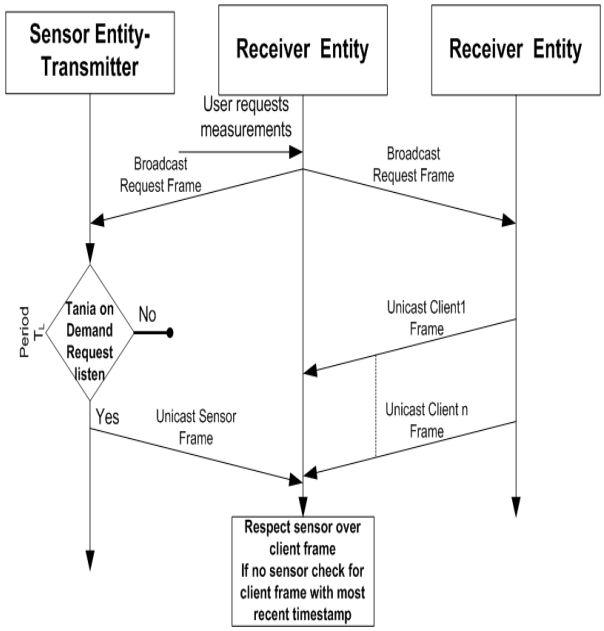


Fig. 5: Tania-AHOD protocol

The sensor entity timeout value for all unicast acknowledgements reception is set to 2sec (RTO=2s). If RTO is reached, a frame alert retransmission for non-acknowledged nodes will be performed at least three times for each node. As soon as the alert frame transmission is completed, a multiplicative decrease in transmission period will be performed.

IV. CASE STUDY AND PERFORMANCE MEASUREMENTS

The case study includes one sensor entity (see Fig. 1 – MCU transmitter with temperature and CO_2 sensors). The sensor entity equipment uses a 3.3V Arduino pro mini microcontroller connected to an MG-811 C02 analog sensor and a DS18B20 digital temperature one-wire sensor. On the Arduino SPI pins the RF12B transponder is connected in order to transmit periodically measurements to the WSN network. The WSN network consists of two receiver entities nodes. These nodes include an Arduino UNO 5V microcontroller with an SPI RFM12B transceiver. The Arduino is back to back connected via OTG cable to Android mobile phones where the measurements monitoring application resides. The application measurements real time is set by the mobile phones as soon a measurement is received.

The Arduino board used at sensor entities operates at 3.3V DC, since there is a CO_2 sensor requirement to operate at 5V DC the sensor is powered directly from the battery unit. The typical power consumption of an Arduino UNO 5V board in idle mode is 282-300mW. However, the same board Arduino pro mini, operated at 3.3V provides power requirements at idle mode of 27-35mW, including the voltage regulator 5V-3.3V chip power consumption (10 times less than an Arduino5V UNO board used at primer experimentation).

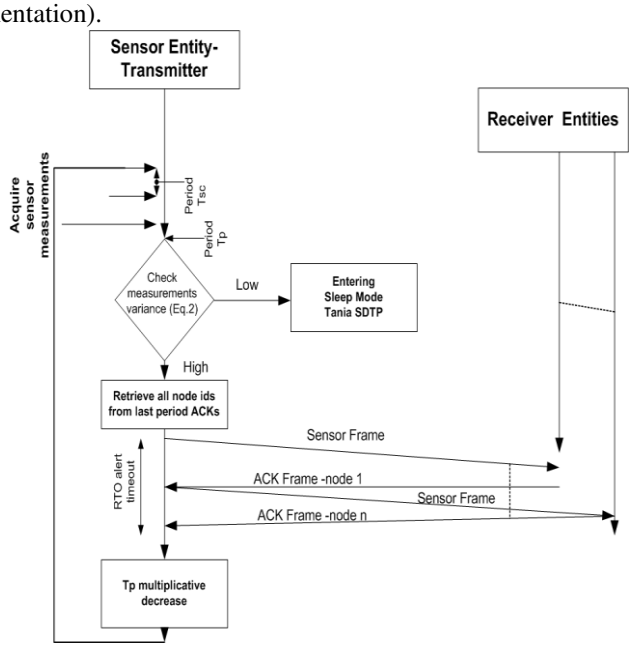


Fig. 6: Tania-AHA protocol

We are using an MG-811 CO_2 sensor [25], which includes a signal conditioning circuit along with a LNA amplifier. Its power consumption is rated from 15-18mA, thus giving a power consumption outcome for 5V sensor required voltage of max 90mW. This is still 3 to 4 times less than Arduino UNO idle power consumption. On the other hand, both digital thermometer DS18B20 and RF12B transponder operate at 3.3V, with transmission RF12B power consumption of 76mW and thermometer consumption of 6mW.Moreover, more power-enhanced improvements have been performed by providing a low power consumption algorithm for the sensor entity. The algorithm's description follows.

Arduino board has six different sleep mode states: idle, ADC noise reduction, power-down, power-save, standby and extended standby [26]. Power save and power down modes power off all Arduino peripheral chips such as SPI, TWI and ADC, contrary to idle mode that still consumes power. Both power save and power down can be triggered by external Interrupts (INT0 and INT1). However, power save mode still feeds power to Arduino timers (0/1/2), while power down mode disables their functionality completely. This of course leads us to Arduino timing problems (millis function connected to timer0) [26].

Arduino includes a watchdog timer (WDT). This timer has a separate on chip RC clock operating at 128 KHz and has 9 prescaler modes of operation starting from 30ms up to 8s. From these modes authors selected the following sleep interval modes to be configured at their algorithm's initialization step: 1s, 2s, 4s and 8s. WDT timer is functional on all Arduino power modes (including power down).

The Brown out detector (BOD) [26] is another functionality of the Arduino board that consumes power and is not required for the sensor entity. The BOD detector checks for voltage anomalies or dips in the voltage of the board and resets the Arduino chip by calling the watchdog timer. This BOD functionality is set by the Arduino fuse bits (extended fuse bit 0x05 to 0x07) and is programmatically pushed to the sensors using in circuit programmers. This BOD functionality reduces the board energy consumption from 0.1 to 0.3mW of continuous power. Moreover, removing the Arduino power led also contributes to less power consumption of 10-12mW.

We propose the following low power reduction steps for our sensor entity measurements:

Step 1a– Setup step: Arduino disables the watchdog timer and initializes timer0 and RFM12B transmission parameters. The CO_2 board is powered directly from 6V battery pack as well as the Arduino Pro mini. RFM12B is powered from Arduino Pro mini Digital pin 5 set to HIGH (3.3V), DS18B20 is powered from Digital pin 6 set to HIGH (3.3V). T_{sc} period counter is set to zero, millis() value is at 5000ms (5s).

Step 1b – Sensor main loop: This is the initiation of the MCU sensors measurement process of sensor probing and it is repeated every T_{sc} time interval set at 2sec.

Step 2a – Check for Ad-hoc requests: Arduino opens the RFM12B transponder for 100ms and checks for any Ad-hoc on demand requests. If a request is sensed it transmits data to the request node.

Step 2b – T_p period reached: Arduino checks whether the periodic counter value of sensor measurements initially set at T_p=30sec has elapsed. If it has elapsed, it will initialize the SPI interface and transmits sensor data via the RFM12B transponder. Afterwards, it opens an I2C communication with the EEPROM chip, an Atmel AT24C128 I2C EEPROM chip [27], reading past period measurements and performs calculation (period increase or decrease) of the next period interval based on the algorithm described at section III.D.

Step 2c – Sensors probing Loop: If T_p<30 or $T_{pcalculated}$, Arduino enters a probing Loop, sets its ADC frequency to 125KHz and performs sequentially 96 measurements of CO_2 and 96 measurements of temperature (around 2ms of time). The time completion and averaging result is produced and saved to an external EEPROM memory I2C communications chip [27].

Step 3 – Sensor MCU Sleep mode: Watchdog timer is set for time of 1s with interrupt1 enabled. The Arduino goes to power down mode and the digital pins 5,6 are set to low. The WDT timer is set as the main millis timer (128 times slower counting – for time period of 1s WDT increases millis value by a value of 8 instead of 1000).

Step 4 – Sensor MCU wake up: Interrupt service routine of Interrupt 1 is entered disabling the watchdog timer and setting the Arduino chip out of the power down mode.

Step 5 – MCU clock calibration: End of Microcontroller loop, set millis counter to millis()+1000-8. T_{sc} period counter calculation equal to (millis()-5000)/1000 and loop re-start from step 2a.

TABLE 1:

PROPOSED SENSOR ENTITY ARCHITECTURE COMPONENTS AND POWER CONSUMPTION

Equipment	Power Consumption-idle mode
Arduino UNO 5V	282mW
Arduino Pro mini 3.3V –idle state	33mW
MG-811 sensor 5V	90mW
DS18B20	6mW
RFM12B transmission	76mW
RFM12B data reception, ad-hoc sensing	39.6mW
AT24C128 EEPROM	17mW (100,000 writes)

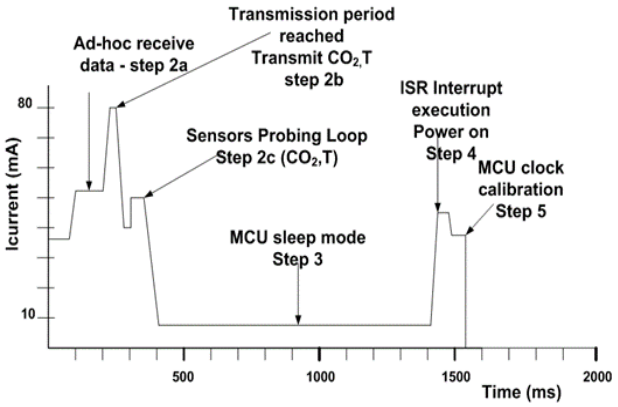


Fig. 7: Sensor entity energy footprint for a Tp=2sec.

TABLE 2:

WSN COVERAGE AREA FRAME LOSS AND MAXIMUM CHANNEL CAPACITY

Coverage Distance (meters)	Frame loss	Mean delay (sec)	Throughput (bits/sec)
10	0%	0.08	2373.5
20	0%	0.1	2362.3
30	3%	0,15	2286.2
40	7%	0,19	2187,6
50	15%	0,22	1996,5

From the sensor entity and WSN experimentation the maximum energy footprint for a minimum sensor entity period of T_p=2s is presented at Figure 7. In addition, the WSN sensor pin powered RFM12B transponder-receiver RFM12B transponder maximum coverage distance and

maximum channel throughput capacity are shown at Table 1.

The WSN sensor pin powered RFM12B transponder-receiver RFM12B transponder maximum coverage distance and maximum channel throughput capacity are shown at Table 2.

V. CONCLUSIONS

This paper presents a new WSN architecture comprised of sensor entities and mobile receiver entities, for the transmission of sensory data to mobile roaming clients with medium to small transmission ranges coverage. Our approach is based on existing cheap and open source hardware and software solutions. On the proposed WSN architecture, a set of three network layer protocols is presented, called RF-Tania protocol framework, for the transmission of relaxed sensory data, ad-hoc on demand transmission of sensory data and alert transmissions in cases of sensors' threshold events.

It is set as a future work, a further detailed experimentation of proposed architecture and protocols, as well as comparison results of the sensor entity energy consumption and coverage with similar transponder devices energy consumption and coverage results.

REFERENCES

[1] Jeelabs, RF12B MAC protocol, *http://Jeelabs.org*, 2013.

[2] Z. A. Eu, H-P. Tan, and W. K. G. Seah, "Design and performance analysis of MAC schemes for wireless sensor networks powered by ambient energy harvesting". *Ad Hoc Networks*, vol. 9, no. 3, ISSN 1570-8705, pp. 300-323, 2011.

[3] M. Y. Naderi, P. Nintanavongsa,K. R. Chowdhury, "RF-MAC: A Medium Access Control Protocol for Re-Chargeable Sensor Networks Powered by Wireless Energy Harvesting".*IEEE Transaction on Wireless Communications*, vol. 3, issue 7, ISSN 1536-1276, pp. 3926 -3937, July 2014.

[4] P. Nintanavongsa, M.Y. Naderi, and K. R. Chowdhury, "A Dual-band Wireless Energy Transfer Protocol for Heterogeneous Sensor Networks Powered by RF Energy Harvesting".*International Conference on Computer Science and Engineering Conference (ICSEC)*,IEEE,pp.400-405, Sept. 2013.

[5] Dong-feng, Su, C. Xiang-jian, L. Di, X. Zhi-jun and C. Zhi-feng, "Research of New Wireless Sensor Network Protocol: ZigBee RF4CE."*International Conference on Electrical and Control Engineering (ICECE)*, IEEE, pp. 2921-2924, 2010.

[6] Song, Wen-Miao, Yan-Ming Liu, and S-E. Zhang. "Research on SMAC protocol for WSN." *International Conference on Wireless Communications, Networking and Mobile Computing*, IEEE, pp. 1-4, 2008.

[7] Song, Jianping, S. Han, A. Mok, D. Chen, M. Lucas M. Nixon and W. Pratt, "WirelessHART: Applying wireless technology in real-time industrial process control." *Real-Time and Embedded Technology and Applications Symposium, IEEE*, pp. 377-386, 2008.

[8] RadioPulse LM2470/2475 RF4CE Technical Specification Datasheet, http://www.radiopulse.co.kr/, 2015.

[9] Microchip Application Notes, "AN1283: Microchip Wireless Media Access Controller MiMAC", "AN1284: Microchip Wireless Application Programming Interface MiApp" and "AN1066: Microchip Wireless Networking Protocol Stack MiMAC", http://www.microchip.com/miwi, 2011.

[10] T. Lennvall, S. Svensson and F. Hekland, "A comparison of WirelessHART and ZigBee for industrial applications", *IEEE International workshop on Wireless Factory Communication Systems*, pp. 85-88, 2008.

[11] Freescale Product Brief and development kits MC132X,http://www.element14.com/community/servlet/JiveServlet/download/46712-2-98858/MC13242_RFFS.pdf,2010.

[12] C. Gomez, J. Oller and J. Paradells, "Overview and Evaluation of Bluetooth Low Energy: An Emerging Low-Power wireless technology.", *Sensors Journal*, vol. 12, no. 9, ISSN 1424-8220, pp.11734-11739, 2012.

[13] Nordic Semiconductor, "Multiprotocol Bluetooth low energy /2.4GHz RF System on Chip - nRF51822 product specification", https://www.nordicsemi.com/.../nRF51822_PS v3.1.pdf

[14] 802.15.4, Part 15.4, "Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for Low-Rate Wireless Personal Area Networks (LRWPANs)".

[15] ZigBee Alliance, "ZigBeeRFCE specification: ZRC profile. Version 2.0 ZigBee", http://www.zigbee.org/zigbee-for-developers/network-specifications/zigbeerf4ce/, 2014.

[16] P. Rohitha, P. R. Kumar, N. Adinarayana and V.N. Rao, "Wireless Networking Through ZigBee Technology", *International Journal of Advanced Research in Computer Science and Soft. Eng*. IJARCSSE, ISSN 2277-128X, vol. 2, no. 7, pp. 49-54, 2012.

[17] ZigBee Alliance, "ZigBee RFCE specification: ZigBee v.1.0", http://www.zigbee.org/zigbee-for-developers/network-specifications/zigbeerf4ce/, 094945r00ZB 2010.

[18] WiMedia Specifications, "The WiMedia common Radio Platform", http://www.wimedia.org/en/docs/10001r01WM_MPI-WiMedia_MAC-PHY_Interface_Specification_1.5.pdf, Rel 1.5, 2009.

[19] E. Karapistoli, F.N. Pavlidou, I. Gragopoulos and I. Tsetsinas, "An overview of the IEEE 802.154a standard", *IEEE Communications Magazine*, ISSN 0163-6804, pp.47-52, Jan 2010.

[20] WG802.15 - Wireless Personal Area Network (WPAN) Working Group, "802.15.4-2011 - IEEE Standard for Local and metropolitan area networks-Part 15.4: Low-Rate Wireless Personal Area Networks (LR-WPANs)", IEEE standard, 2011.

[21] WG802.15- Wireless Personal Area Network (WPAN) Working Group, "802.15.4-2015 - IEEE Standard for Low-Rate Wireless Personal Area Networks (WPANs)", IEEE standard 2015.

[22] I. Howitt, J. A. Gutierrez, "IEEE 802.15.4 low rate - wireless personal area network coexistence issues", *Wireless Communications and Networking*, ISSN 1525-3511, vol. 3, pp. 1481-1486, 2003.

[23] HopeRF transceiver 433MHz RF12B datasheet, http://www.hoperf.com/upload/rf/RFM12B.pdf, 2012.

[24] HopeRF RF69W/RF69HW high range 433MHz transceiver datasheet, http://www.hoperf.com/upload/rf/RFM69CW-V1.1.pdf, 2015.

[25] MG-811 CO_2 sensor datasheet, Sandbox Electronics, http://sandboxelectronics.com/?P=147,2011.

[26] Atmel ATmega328 8 bit microcontroller datasheet and user's manual, http://www.atmel.com/.../atmel-8271-8-bit-avr-microcontroller-atmega48a-48pa-88a-88pa-168a-168pa-328-328p_datasheet_complete.pdf, 2015

[27] Atmel 128K/16KBytelow voltage operation EEPROM TWI chip, www.smartrobots.pl/download/AT24C256.pdf, 2012.

[28] AKBARI, Saba. Energy harvesting for wireless sensor networks review. In: Computer Science and Information Systems (FedCSIS), 2014 Federated Conference on. IEEE, 2014. p. 987-992.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1075–1084
DOI: 10.15439/2016F240
ACSIS, Vol. 8. ISSN 2300-5963

Comparison of MANET self-organization methods for boundary detection/tracking of heavy gas cloud

Mateusz Krzysztoń
Institute of Control and Computation Engineering, Warsaw University of Technology
Nowowiejska 15/19, 00-665 Warszawa, Poland
Email: mateusz.krzyszton@gmail.com

***Abstract*—Mobile wireless ad hoc network (MANET) becomes increasingly popular in responding to emergency situation. In this paper a possibility to support rescue team in monitoring heavy gas cloud with MANET comprised of mobile sensing devices is investigated. In the view of the current state of research, two methods for controlling mobile sensing devices during MANET self-organization are presented. The first one is based on a greedy approach whereas the second on a repulsion from the estimated centroid of a cloud and other nodes. Various variants of both methods are considered and their efficiency in terms of detection quality and energy saving is evaluated with MobASim simulation software. The results are discussed and one variant is chosen as the basis for the future research.**

I. Introduction

ALL over the world great amount of toxic substances is transported and stored. Some of these substances after release form clouds of gas heavier than air [1]. Despite high safety standards severe accidents, in which dangerous substance is released to the atmosphere, occur. Examples of accidents involving heavy gas release include those with chlorine [2], nitrogen dioxide [3] and sulfur dioxide [4]. Heavy gas cloud can be created by natural reasons as well — in 1986 a massive, sudden release of carbon dioxide occured from Lake Nyos, a volcanic crater lake, and as a result around 1700 people were killed [5]. Another source of the toxic gas cloud can be military or terrorist attack [6], [7].

As heavy gas cloud is formed a rescue team has two goals: evacuate endangered area and neutralize the cloud. In both knowledge of position, boundary and direction of the cloud is substantial as it supports rescue team in surrounding and then suppressing the cloud, making decision which area to evacuate first and identifying safe evacuation routes. Usually there is no need for modeling exact concentration level of gas inside the cloud.

Because of the negative buoyancy behavior of the heavy gas cloud is different than the one showed by positively or neutrally buoyant clouds. The main difference is gravitational velocity field (gravitational slumping) which influences the way cloud moves and changes its shape with time. Hence, special group of mathematical models was developed to describe the dispersion of heavy gas in the atmospheric air [8]. However, usage of these models to define boundary and position of the cloud demands specifying values of many parameters as direction and speed of wind, quantity of released gas, type of release (instantaneous or continuous), etc., which are often unknown and/or variable. Most of the described models assume obstacle-free, flat terrain; Even if obstacles or topography of terrain is considered only simple scenarios can be modeled. These issues create need for more universal, environment independent methods for the cloud boundary tracking in an unknown terrain.

In this work implementation of Mobile Ad Hoc Networks (MANET) for supporting rescue team in emergency action is proposed. MANET is comprised of wireless mobile devices, which can dynamically and autonomously self-organize into temporal networks to provide a discovering and tracking boundaries of dynamic heavy gas cloud. The network adapts to current conditions in deployment area by forming adequate topology. Nodes communicate wirelessly to exchange their knowledge about the environment.

The article is organized as follows. First, state of the current research on applying MANET in emergency situations and methods for boundary detection/tracking is presented. Then in the section III problem is formulated. Next two simple distributed methods for boundary of heavy gas cloud discovering and tracking are proposed. In the section V different variants of these methods are evaluated and compared in terms of both quality and energy efficiency. Finally, results are discussed and future directions of work are briefly outlined.

II. Related work

Contemporary lots of research focus on both using sensor networks to support detection and response to emergency situations and on detecting boundary of phenomena. Below the most interesting works in these two fields are briefly presented.

MANET can be used in emergency situation to address various issues. Usually as the result of natural disaster communication systems are down [9]. Thus MANET can be used to establish new communication layer for rescue team [9], [10]. In [11] system for firefighters that provides possibility to conduct audio and video conference during action is described. Another communication network architecture is presented and assessed for existing Telemedicine Service in [12]. In [13] MANET is created to establish communication with single robot that searches building in emergency scenario. The comparison of the MANET based solution with the static network in terms of throughput shows improvement

978-83-60810-90-3/$25.00

if network's topology does not change often, but decrease otherwise. The influence of relay nodes placement on network reliability, stability and availability is discussed in [14].

Another purpose of MANET can be supporting decision making process of emergency team by providing important data. In [15], network is created to serve as a communication layer and collect medical data from remote sources about mass casualty incidents, thus supports decision making. Similar scenario of MANET application is presented in [16] — locations and current state of victims are monitored and provided to rescue team. Another application of MANET is to support evacuation of people from endangered area by providing an appropriate evacuation route in real time [17], [18].

MANET can be also used in disaster detecting — methods based on analyzing people behavior are proposed in [17], [19], whereas distributed control system for maximizing the joint detection probability of events in a given area is presented in [20].

In the same time much effort has been spent on research on detecting and tracking *phenomena clouds* — events characterized by nondeterministic, dynamic temporal variations of cloud shape, size, speed, and direction of motion along multiple axes [21]. The examples of phenomena cloud are oil spills, movement of group of people and gas clouds. Most interesting approaches are presented below, divided according to the type of used devices — stationary and mobile (can move autonomously).

A number of works on phenomena cloud detection with stationary nodes exists. In [22] authors focus on meeting time and accuracy constraints in a task of tracking phenomena. In [23], scenario with nodes sparsely deployed in the area is considered. In such situation collected data can be ambiguous in terms of number and shape of detected phenomena clouds. To reconstruct contours method based on gradients of concentration is designed.

Important issue in WSN networks is energy saving. In [24], method for limiting power usage by selecting subset of WSN nodes that are localized on the boundary of phenomena (representative nodes) and reducing size of messages is proposed. Distributed boundary estimation strategy with mechanisms for decreasing number of messages sent between nodes, and adaptive turning off sensors, is described in [25]. In [21] authors propose several distributed methods and focus on minimizing resource utilization by both reducing number of active sensors and optimization of centralized query processor, which gather information about cloud in real time. Energy efficient algorithm for phenomena tracking for the case of void area (area not covered by nodes) existence is presented in [26]. Interesting scenario is considered in [27] — sensors are carried by people, thus nodes have no influence on their location, but are mobile. Authors address their method for scenarios with dense network, where lowering traffic is crucial aspect in terms of energy efficiency.

As concluded in work on target tracking [28] use of mobile devices can increase tracking performance significantly in comparison with stationary network of the same number of devices. Hence, much research on applying mobile devices has been conducted. Comprehensive survey on boundary estimation, covering and tracking with collaborative sensors can be found in [29]. In the next few paragraphs the most important approaches (including the newest research, not included in the survey) are discussed. In some works instead of detecting boundary more general task is considered – detection of perimeter defined as curve of the given constant concentration. Boundary can be seen as perimeter given by concentration close to zero.

Tracking and exploring phenomena cloud with one device is important issue in the first phase of detection, when phenomena cloud is sensed by one node only and other nodes did not arrived yet to the area of interest. The decision about the direction of movement can be based on trigonometric reasoning [30] or gradient of concentration [31]. Additionally, in [31] use of artificial neural networks for the nonholonomic mobile robot to move in the designed direction is described.

In [32] multiple autonomous vehicles are used to estimate boundary, however, with no cooperation between them. Authors focus on tracking algorithm, that enables each vehicle to go along boundary based on it's concentration measurements. Measurement noise is addressed. Data gathered from all vehicles is used to estimate boundary. The tracking algorithm was validated in static environment using testbed in [33]. The same algorithm was used in [34] to estimate phenomena cloud boundary with single mobile node. Additional support by WSN was introduced to correct mobile node's movement by predicting future center of the cloud.

Cooperation between robots can increase quality of cloud boundary detection. In [35], algorithm for even robots placement on the boundary was proposed. The algorithm was proved to converge in case of static boundaries and to be efficient for slowly-moving boundaries. However, assumptions that initial estimation of boundary is known to agents and each agent is able to locally estimate the tangent and the curvature of the boundary are needed. These assumptions are not valid in the case of heavy gas cloud, because robot can measure gas existence in one point only at the same time.

Another cooperation based method for placing evenly on the boundary was introduced in [36]. Additionally, group of mobile robots track perimeter of certain substance counterclockwise using camera. However, to uniformly distribute around perimeter the desired separation distance need to be set a priori.

An interesting method for the stationary environment, inspired by the active contour model used in the image segmentation field, was proposed in [37]. The approach is strictly dependent on the concentration of substance — it is assumed that lowest concentration is on the boundary and a gradient of concentration points to that boundary. Algorithm for scattering devices evenly, based on pairwise repulsion force, was additionally proposed. However, node needs to choose between tracking boundary or scattering in every step. Similar idea is presented in [38], where non-stationary environment is considered. Again, assumption on concentration is needed.

Another method based on concentration value is described in [39]. Every sensor knows concentration in its location and in its close neighbourhood. To reach the location with the given concentration (concentration on the boundary) the sensor decides in every step which neighbouring location move to. Various strategies of taking decision were proposed and examined in scenarios with linear and non-linear variations of concentration.

In [40], group of unmanned underwater vehicles is used to find and patrol underwater perimeter. Two methods for controlling movement of vehicles were proposed and compared. The first one is based on aforementioned snake algorithm (concentration based). The second models vehicles as a gas of particles, which affects each other speed, similarly to the one of the approaches presented in this paper. The authors conclude that free-concentration methods are better for tracking underwater perimeter in real world scenarios.

Some research was conducted on using UAVs (unmanned aerial vehicles) to detect and track the shape of an environmental boundary. In [41], authors propose the method for calculating path of UAV that allows recognition of the contaminant cloud's shape. The approach was extended with methods for predicting contaminant cloud's shape [42] and avoiding obstacles [43]. UAVs can also be used to track forest fire boundary [44].

Interesting issue of the boundary detection is discussed in [45]. Authors propose some algorithms for deciding when to start spreading sensors (how far in advance before reaching the boundary) to speed up converge. The algorithm is based on a rate of concentration change, thus is not proper for situation in which boundary is defined with small concentration value (whole area where any gas exists should be classified as cloud interior). However, the idea of early sensor spreading should be addressed in our work in the future. Other drawbacks of proposed spreading algorithm is its centralized nature and assumption on obstacles free environment.

In the view of this brief review, MANET can be successfully used in emergency situation. At the same time there is lack of proper mobility control method for nodes for case of detecting boundary of dense gas cloud. Most of above methods base on concentration distribution, which is inadequate for scenarios with dense gas dispersion, because of various reasons. Firstly, heavy gas cloud has very dynamic internal changes. Secondly, gas sensors sense gas in single point only at a time and the sensory readings can be inaccurate. Additionally, even slightly pleated area influences distribution of gas in the area - lower concentration can be caused by existence of both cloud boundary and hill. Hence, we propose two new methods for real time estimation of area covered by heavy gas cloud with MANET, based on binary information from sensors about gas existence.

III. Problem Description

The aim of the paper is to create a sensing network for a heavy gas cloud detection in a two-dimensional workspace W, and boundary tracking to estimate a size of this. In our investigation, a network is a physical system modeled as a set of n unmanned vehicles or mobile robots D_i, $i = 1, \ldots, n$. All these vehicles are equipped with radio transceivers and gas sensors. All network nodes are solid bodies with an arbitrary shape. In order to simplify the description of the system we model each network node by a polygon with its reference point $\mathbf{c}^i = [x^i, y^i]$, which is the location of the device (exactly its antena). All vehicles are forced to move in advisable direction with the speed $v \in [v_{min}, v_{max}]$.

In this work it is assumed that each pair of nodes D_i and D_j can communicate independently on the Euclidean distance between these nodes (d_{ij}), with use of external communication system. In the future, local multi-hop communication between each pair of nodes should be considered to enable deploying the network in area with no external connectivity system.

The goal of both methods is to control direction and speed of each node in such way, that would increase quality of estimation of area covered by gas and decrease total distance traveled by all nodes. The proposed methods are based on PFM (*Potential Function Mobility*) model described in details in [46]. In this model each node has single goal g defined with target point in the workspace $\mathbf{c}_g^i = [x_g^i, y_g^i]$ and its movement is influenced by positions of other nodes and obstacles. The PFM model combines concepts of an artificial potential fields and particle-based mobility schemes. Each node is treated as a self-driven particle moving from a high-value state to low-value state of artificial potential field. The artificial potential function is constructed as a sum of repulsive and attractive potentials. Its value depends on Euclidean distance between a given node and all other nodes in the network, and the distance to target position and obstacles in W. As in this work obstacles-free environment is assumed each device D_i has to calculate its new position solving problem of minimizing an artificial potential function U^i (influence of obstacles is omitted):

$$\min_{\mathbf{c}^i} \left[U^i(\mathbf{c}^i) = U_g^i(\mathbf{c}^i) + \sum_{j=1, j\neq i}^{n} U_j^i(\mathbf{c}^i) = \epsilon_g^i \left(\frac{\overline{d}_g^i}{d_g^i} - 1 \right)^2 + \sum_{j=1, j\neq i}^{n} \epsilon_j^i \left(\frac{\overline{d}_j^i}{d_j^i} - 1 \right)^2 \right], \tag{1}$$

where U_g^i and U_j^i are potentials derived from g and D_j, respectively. $\epsilon_g^i \geq 0$ and $\epsilon_j^i \geq 0$ are weighting factors determining the importance of, respectively, the goal g and the device D_j. d_g^i and d_i^i are real Euclidean distances between $\mathbf{c}^i$ and respectively, $\mathbf{c}_g^i$ and $\mathbf{c}_j^i$ after a network transformation, and $\overline{d}_g^i$ and $\overline{d}_j^i$ the reference distances between $\mathbf{c}^i$ and respectively, $\mathbf{c}_g^i$ and $\mathbf{c}_j^i$. The reference distance is understood as the expected distance between node and target or other node, accordingly.

The final network topology depends on choice of target location ($\mathbf{c}_g^i$), the values of the reference distances ($\overline{d}_g^i$, $\overline{d}_j^i$) and weighting factors (ϵ_g^i, ϵ_j^i). Hence, in the next section two distinct methods for calculating these values in task of estimating boundary of area covered by gas are presented.

IV. METHODS

The detection of area covered by gas is divided into two phases. In the first one, when only one node detects heavy gas with sensor, the node has to track the cloud and broadcast information about its position to others. All nodes that do not sense gas are attracted by the point defined by the position of the node that senses gas. Tracking phase last to the moment in which at least one more node arrives to the area covered by gas. Then, the boundary detection phase starts.

The algorithm for controlling node that senses gas in the tracking phase is presented in Algorithm 1. The algorithm is executed as only the node starts to sense gas. The idea of the algorithm is to move inside the cloud as long as gas is sensed (k steps), then when the cloud is lost return to the cloud by taking opposite direction ($k/2$ steps) and choose new direction randomly. In this phase $\epsilon^i_j = 0$ (thus $U^i = U^i_g$). Moving in the given direction is achieved by adding attracting goal g in an appropriate location $\mathbf{c}^i_g$ with $d^i_g = 1$. The advantage of this algorithm is simplicity and power efficiency thanks to little computation needed. More advanced approach, that does not base on gas concentration as well, was presented in [30].

Algorithm 1 Algorithm for controlling node in tracking phase

```
if moving then
    continue moving in that direction
else
    start moving in random direction
end if
while ( nodesSensingGas.length == 1 ) do
    k := 0
    while ( sensingGas ) do
        continue moving in that direction
        k = k + 1
    end while
    start moving in the opposite direction
    while ( not sensingGas ) do
        continue moving in that direction
    end while
    k = k / 2
    while ( k >0 and sensingGas ) do
        continue moving in that direction
        k = k - 1
    end while
    choose new direction randomly
end while
```

As at least two nodes sense gas the estimation area covered by gas phase begins. Below two different approaches for implementation of this phase are proposed. Before the approaches are described in details few remarks have to be made:

Remark 1. Binary sensory reading causes that all sensors that do not sense gas in a given moment are treated the same way, independently on how many steps have passed after last positive sensory reading. Such simplification causes that much of important information can be lost from the perspective of rescue team. Hence, three different states of node are introduced:

(a) sensing gas at time t
(b) not sensing gas at time t, but did sense gas at least once in last h steps
(c) not sensing gas at time t, and did not sense gas in any of last h steps

The boundary of the covered area is estimated as convex hull of the set C^+ defined as the set of locations of all nodes in state (a) or (b) with $h = 2$.

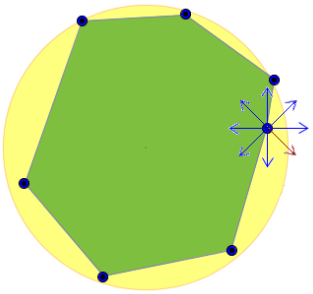


Fig. 1: Directions of movement considered by a node in the greedy approach ($k_d = 8$).

Remark 2. The term *centroid* will refer to the center of the cloud $\mathbf{c}_c$, which is calculated as an arithmetic average of positions of all nodes that sense gas:

$$\mathbf{c}_c = \frac{\sum_{\mathbf{c}^i \in C^+} \mathbf{c}^i}{|C^+|}. \tag{2}$$

Remark 3. In both approaches the nodes that do not sense gas are attracted by single goal located in centroid ($\mathbf{c}^i_g = \mathbf{c}_c$) with reference distance $d^i_g = 1$ and influence of other nodes is omitted ($\epsilon^i_j = 0$).

A. Greedy approach

In greedy approach every node in every time step chooses one of k_d directions to move. The decision is based on localization of all other nodes that sense gas (C^+). The node calculates the area of current cloud's shape estimation (convex hull of C^+), simulates the move in each of k_d directions and calculates the change of area after each move. In accordance to greedy approach the direction that increases the area most is chosen as a new direction of the node's movement. However, as decision is independent on the simultaneous decisions of other nodes, it may occur that the chosen direction is not optimal in the given situation. As in the exploration phase moving in the chosen direction is achieved by adding attracting goal g in an appropriate location $\mathbf{c}^i_g$ with $d^i_g = 1$. The example of possible directions for $k_d = 8$ is presented in Figure 1.

The above decision making process should be applied only to the node D_i which location c^i is one of the vertices of convex hull created by C^+ (lies on the estimated boundary). The node that is inside the estimated boundary should move towards one of the edges of approximated cloud's shape. To

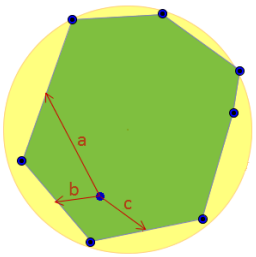


Fig. 2: The target edges chosen by the node inside estimated boundary according to the strategy: (a) — the longest, (b) — the closest, (c) — the greatest relation of the length of the edge to the distance to the edge.

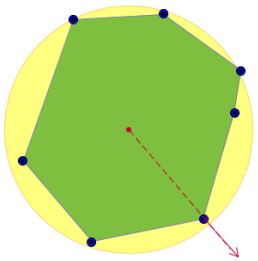


Fig. 3: The force acting on the node in the centroid approach. Red dot is the estimated centroid of the cloud, red arrow represents the force.

improve uniform distribution of nodes on the boundary the target g is located in the middle of the chosen edge. Three strategies for choosing an edge were proposed (Fig. 2):

(a) the longest (to distribute nodes uniformly on the boundary);
(b) the closest (to limit energy consumption);
(c) with the greatest value of ratio *length of the edge* to *distance of the node to the middle of the edge* (trade off between energy consumption and uniform distribution).

When the node is on the boundary of the real cloud its greedy decision to increase area of convex hull causes that it escapes from the area covered by gas. Then it is attracted by centroid and returns to the position on the boundary. Such repeated behaviour causes that node pass considerable distance within small area but do not improve quality of cloud detection significantly. Hence, simple stabilization mechanism to limit energy consumption is proposed. If the node senses gas in the given time and was outside the cloud (did not sense gas) in one of the last k_s steps then it freezes ($\epsilon_g^i = 0$[1]). The higher value of k_s is the greater network stability is expected.

B. Centroid repulsion

The second, as well decentralized, method is based on idea to repeal nodes that sense gas from the center of cloud to keep track of changing cloud's shape. To repel the node to the boundary of cloud the goal g is localized in $\mathbf{c}_c$. The reference distance of that goal $\hat{d}_g^i$ is slightly greater than the greatest distance between any of points from C^+ and $\mathbf{c}_c$:

$$\hat{d}_g^i = \max_{\mathbf{c}^i \in C^+} d(\mathbf{c}^i, \mathbf{c}_c) + \epsilon \tag{3}$$

where ϵ is slightly greater than zero and $d(\mathbf{c}^i, c_c)$ is Euclidean distance between two points $\mathbf{c}^i$ and $\mathbf{c}_c$. The result of the repulsion from the centroid is the node movement in direction shown on Fig. 3.

Similarly to greedy approach, when the node is on the boundary it is frequently being pushed out and drag into the cloud. Hence, the aforementioned stabilization mechanism is added also in this case.

[1]It is assumed that if $U^i = 0$ than the node D_i freezes.

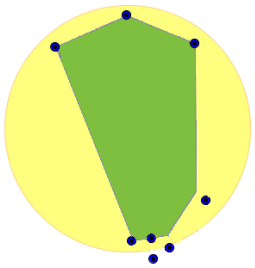


Fig. 4: Distribution of nodes after 40 steps of exemplary simulation if only goal localized in centroid is considered and other nodes' locations are ignored (i.e. $\epsilon_j^i = 0$).

If the node location is not influenced by locations of other nodes (i.e. $\epsilon_j^i = 0$) then estimated area covered by gas is strictly dependant on initial location of nodes. If the location of nodes is similar, which is a frequent situation in real life scenarios, the nodes are not uniformly distributed on the boundary, thus much of dangerous area is not discovered (Fig. 4). To uniformly distribute nodes on the boundary in decentralized manner each node should repel from each node that senses gas. Hence, ϵ_j^i should be greater than zero if node D_j senses gas and the expected distance between node D_i and D_j (d_i^j) should be greater than current distance between these nodes ($\hat{d}_j^i$). As long $\hat{d}_j^i - d_i^j = \Delta d = const$ ($\Delta d > 0$) for each node D_j it can be observed that according to (1) the closer the nodes are the bigger influence on each other they have. Fig. 5 depicts forces acting on exemplary node from all other nodes and the goal located in centroid, as all are treated as potential sources.

Due to the introduction of stabilization mechanism the question when $\epsilon_j^i > 0$ arises. Four variants exist:

(i) only when repelling from centroid is considered ($\epsilon_g^i > 0$);
(ii) only in stabilization phase — when repelling from centroid is deactivated ($\epsilon_g^i = 0$);
(iii) always (independently on value of ϵ_g^i);
(iv) never.

It can be expected that the more often interaction between nodes is taken into account the better shape of cloud will be estimated. However, the consequence of often interaction

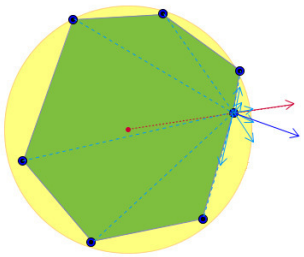


Fig. 5: Forces acting on the node if repelling from other nodes is considered. Red dot is the centroid of estimated cloud area, red arrow is force resulting from repelling from the centroid and blue arrows are forces resulting from repelling from other nodes. Dark blue arrow is resultant force acting on the node.

between nodes will be increased energy consumption due to longer distance traveled by nodes and more communication. All variants, as well as other ideas presented in this section for both greedy and centroid approach, were examined and the results are presented below.

V. Experiments

The experiments were conducted with MobASim platform [47], which contains both PFM model and simple model of heavy gas dispersion. The aim of the experiments was to verify the quality of the proposed approaches and to adjust their parameters according to the following criteria:

1) percent C_{cov} of the gas cloud covered by a MANET

$$C_{cov} = \frac{cov_{\mathcal{G}}}{cov_{GC}} \cdot 100\%, \tag{4}$$

where C_{cov} is the percent of a surface of a gas cloud detected by MANET, cov_{GC} denotes a surface of the whole gas cloud, $cov_{\mathcal{G}}$ an area of a polygon with boundary discovered by all sensing devices.

2) total distance traveled by all nodes, which is the major factor of the energy consumption.

Each experiment was composed of nine different (in terms of wind velocity and cloud initial position) simulations in which the MANET was used to detect area covered with gas (Fig. 6). The variety of scenarios allowed for gathering results independent on initial position of network relative to the cloud. In each simulation the same configuration of environment and gas (chlorine) are taken (Table I). In every experiment each of the nodes used the same configuration of the method in the process of network self-organization. The network is composed of $n = 10$ nodes.

The result of each experiment are graphs of the above criteria versus time (in each simulation $t = 0$ is the moment of detecting the gas by at least one node).

A. Greedy approach

Firstly, the strategies for choosing target edge by node inside the detected area were examined. The other parameters of this method were set as follows: k_d = 8 and k_s = 4. The results are

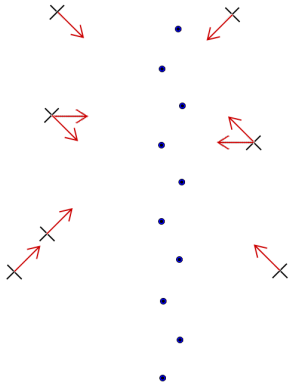


Fig. 6: Initial topology of MANET and configurations of clouds — initial position of cloud is marked with a black cross and the velocity of wind with a red arrow. If one cross is associated with two different arrows then it is two different configurations (with the same initial position of the cloud).

TABLE I: Values of parameters of the chlorine gas cloud simulation model [47].

Symbol	Value	Units
$\mathbf{c}_c$	[200, 200]	$[m,m]$
$\|v_c\|$	1	$\frac{m}{s}$
r	10	m
m_0	2000	kg
m_a	0	kg
g	9.81	$\frac{m}{s^2}$
ρ_{air}	1.20	$\frac{kg}{m^3}$
M	71	$\frac{g}{mol}$
M_{air}	29	$\frac{g}{mol}$
c_p	0.48	$\frac{kJ}{kg*K}$
c_p^{air}	1.01	$\frac{kJ}{kg*K}$
u_*	0.15	$\frac{m}{s}$
ΔH_0	661	$\frac{kJ}{kg}$
T_{air}	293.15	K
α	0.9	-

presented in Figure 7. As expected the quality of gas cloud detection is directly proportional to the distance traveled by nodes. Hence the choice of strategy depends on priorities in concrete application — if the priority is quality of detection than (a) strategy should be chosen and this strategy is chosen for further considerations. Otherwise, if the energy efficiency is critical aspect the (b) strategy is the most efficient choice.

The second parameter to adjust was the number of directions considered by a node localized on the estimated boundary of cloud, k_d. The length of stabilization phase was set as before to k_s = 4. The obtained results (Fig. 8) for $k_d \in \{4, 8, 16\}$ shows that influence of this parameter on the traveled distance is minimal. However, greater number of considered directions (8 and 16) increases quality of detection. Because of no difference between results for $k_d = 8$ and $k_d = 16$, the $k_d = 8$ is deemed as better as less computation is needed to take single decision.

Finally length of stabilization phase (parameter k_s) was

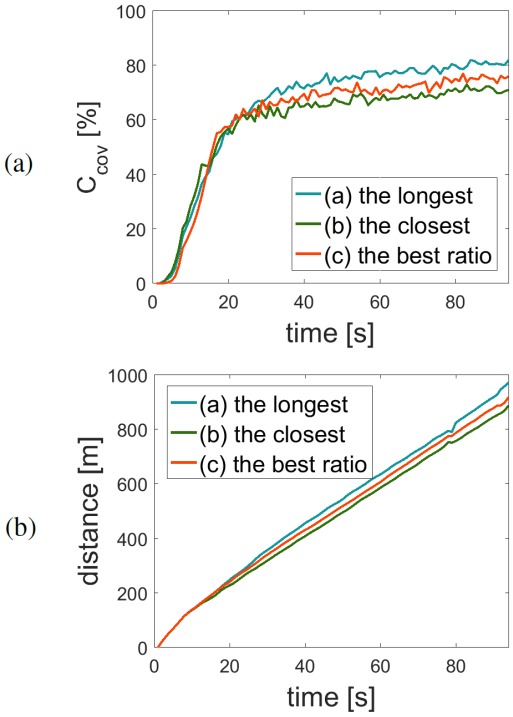


Fig. 7: Cloud detection quality and distance traveled if nodes take decision with greedy approach configured with $k_d = 8$ and $k_s = 4$ and different strategies for choosing edge.

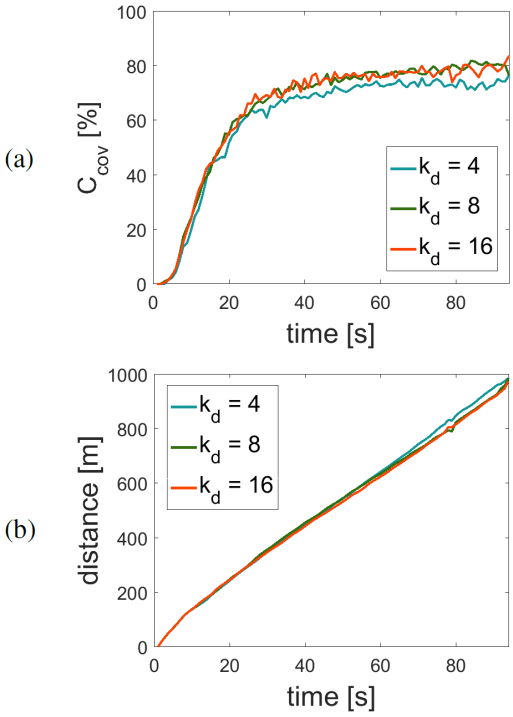


Fig. 8: Cloud detection quality and distance traveled if nodes take decision with greedy approach configured with $k_s = 4$, choosing the longest edge and different values of k_d.

considered. As Figure 9 depicts, the stabilization mechanism has great influence on distance traveled — for $k_s = 4$ the increase of traveled distance is about 45%, whereas decrease in quality of finally detected area covered by cloud is minimal. This value is assumed to be the best in further considerations. According to the results disabling stabilization mechanism can be justified for cases when faster detection of covered area is critical.

The result of deploying MANET composed of nodes taking decision with greedy approach ($k_s = 4$, $k_d = 8$ and choosing the middle of the longest edge) at the end of simulation is presented on Figure 10.

B. Centroid repulsion

Then centroid repulsion approach was examined. In the first experiment the influence of stabilization mechanism and its length (parameter k_s) was tested (Fig. 11), with $\epsilon_j^i = 0$. Significant influence of the stabilization mechanism on the quality of cloud shape detection can be observed only in a very beginning of simulation, when the cloud grows rapidly — as expected the longer stabilization period is the smaller area covered by gas is detected. Comparing the distance traveled for different configurations it appears that $k_s > 4$ does not reduce the energy consumption enough to justify weaker cloud detection. For $k_s = 4$ the reduce of distance traveled is around 50% and this value is used in further considerations.

The second experiment was conducted to examine the influence of repelling nodes from each other. The results of implementing aforementioned strategies (Fig. 12) indicate significant improvement in cloud detection quality if repelling is turned on ($\epsilon_j^i > 0$). However, activating repelling from other nodes in stabilization phase (strategies (ii) and (iii)) causes serious increases of distance traveled by nodes. Hence strategy (i) — repelling from other nodes only when repelling from centroid is active ($\epsilon_g^i > 0$) — can be deemed as conciliation between energy and quality aspects. In the future the possibility of changing strategy during the network deployment to increase quality of detection without significant increase of the distance traveled should be examined.

In Figure 13 the topology of network created by nodes that takes decisions according to centroid approach ($k_s = 4$ and repelling from others nodes only when repelling from centroid is active - strategy (i)) is presented.

C. Comparison

The detection of the exact boundary of the cloud is impossible due to the limited number of nodes. Hence, it can be assumed that both methods detects the area covered by gas satisfactorily — the nodes are placed uniformly on the boundary, thus detected area is close to the maximum. The greedy approach performs slightly better in the middle phase of the cloud detection (steps 15-60) in terms of detection

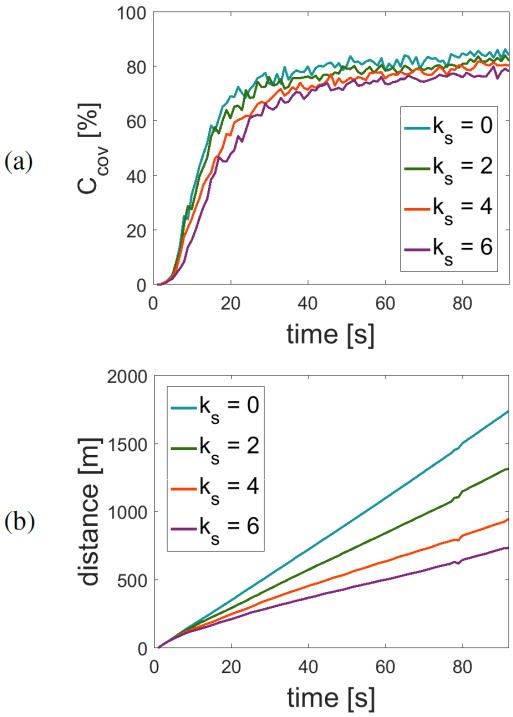


Fig. 9: Cloud detection quality and distance traveled if nodes take decision with greedy approach configured with $k_d = 8$, choosing the longest edge and different values of k_s.

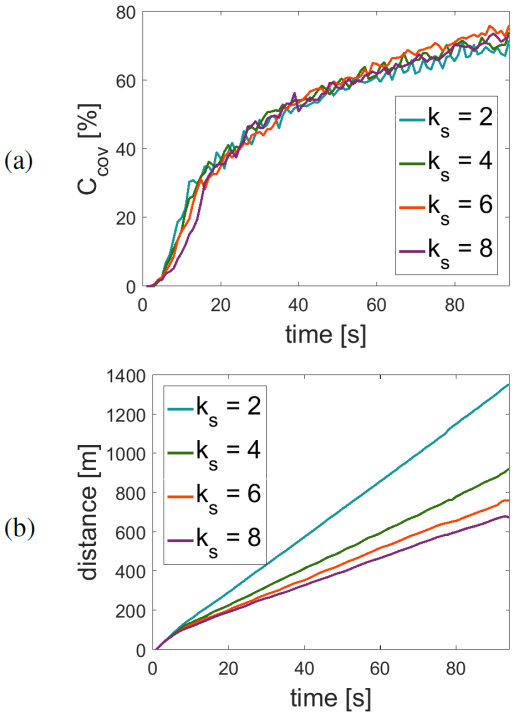


Fig. 11: Cloud detection quality and distance traveled if nodes take decision with centroid approach configured with no repelling from other nodes and different values of k_s.

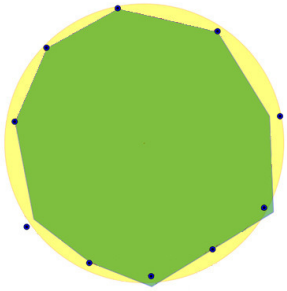


Fig. 10: Topology of MANET in the end of simulation — greedy approach with $k_d = 8$, choosing the longest edge strategy and $k_s = 4$.

quality (Fig. 14). In respect of energy saving there is no significant difference between both methods.

VI. Conclusions

The application of MANET for detecting and tracking cloud of heavy gas was proposed. Two decision-making methods, based on greedy approach and centroid repulsion, were proposed. Both methods proved to be satisfactory in detecting and tracking gas cloud in a simple scenario. Different configurations of methods where tested and compared to adjust parameters values in order to decrease energy consumption and increase quality of cloud's shape detection.

We believe that method based on repulsion from centroid should be used as the basis for future research as this method allow straightforward integration with other decision modules (e.g. module to avoid collisions or tracking the boundary), especially if they are based on PFM model as well.

The future research should concentrate on few areas. Firstly, more complex environment should be considered, e.g. with obstacles, slopes and changing wind direction, to examine method in the case of irregular cloud shape. The possibility to dynamically adjust values of discussed parameters during MANET deployment is also worth considering.

Secondly, scenarios without external communication system should be considered. Thus, the method should be extended with connectivity maintenance mechanism which would allow multi-hop communication between each pair of nodes. As presented approaches create ring topology, which can be ineffective and error prune in terms of communication, creating some internal-cloud communication layer or network clusterization should be investigated.

References

[1] F. Scargiali, E. D. Rienzo, M. Ciofalo, F. Grisafi, and A. Brucato, "Heavy gas dispersion modelling over a topographically complex mesoscale: A {CFD} based approach," *Process Safety and Environmental Protection*, vol. 83, no. 3, pp. 242 – 256, 2005. doi: http://dx.doi.org/10.1205/psep.04073. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S095758200571243X

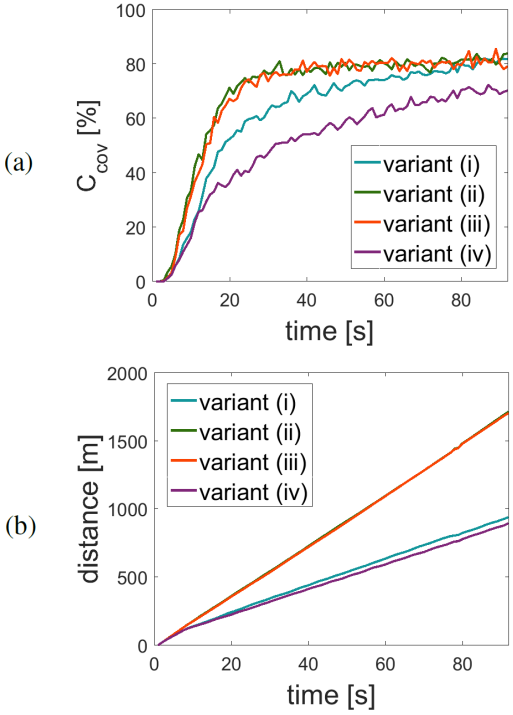


Fig. 12: Cloud detection quality and distance traveled if nodes take decision with centroid approach configured with $k_s = 4$ and different strategies of repelling from other nodes.

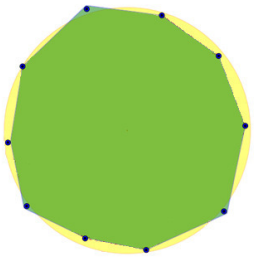


Fig. 13: Topology of MANET in the end of simulation — centroid approach with repelling from others nodes only when repelling from centroid is active and $k_s = 4$.

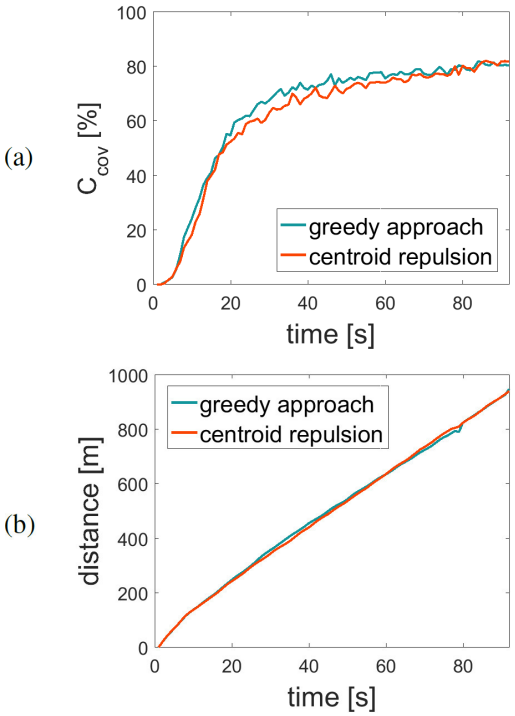


Fig. 14: Comparison of cloud detection quality and distance traveled if nodes take decision with greedy approach ($k_d = 8$, choosing the longest edge and $k_s = 4$) and centroid repulsion approach ($k_s = 4$ and repelling from others nodes only when repelling from centroid is active - variant (i)).

[2] R. Jones, B. Wills, and K. C., "Chlorine gas: An evolving hazardous material threat and unconventional weapon," *Western Journal of Emergency Medicine: Integrating Emergency Care with Population Health*, vol. 11, no. 2, pp. 151–156, May 2010. doi: 10.1109/TIE.2012.2196010

[3] C. C. Yockey, B. M. Eden, and R. B. Byrd, "The McConnell missile accident. Clinical spectrum of nitrogen dioxide exposure," *JAMA*, vol. 244, no. 11, pp. 1221–1223, Sep 1980.

[4] N. B. Charan, C. G. Myers, S. Lakshminarayan, and T. M. Spencer, "Pulmonary injuries associated with acute sulfur dioxide inhalation," *Am. Rev. Respir. Dis.*, vol. 119, no. 4, pp. 555–560, Apr 1979.

[5] P. J. Baxter, M. Kapila, and D. Mfonfu, "Lake nyos disaster, cameroon, 1986: the medical effects of large scale emission of carbon dioxide?" *BMJ*, vol. 298, no. 6685, pp. 1437–1441, 1989. doi: 10.1136/bmj.298.6685.1437

[6] G. J. Fitzgerald, "Chemical warfare and medical response during world war i," *American journal of public health*, vol. 98, no. 4, p. 611, 2008. doi: 10.2105/AJPH.2007.11930

[7] R. Saladi, E. Smith, and A. Persaud, "Mustard: a potential agent of chemical warfare and terrorism," *Clinical and experimental dermatology*, vol. 31, no. 1, pp. 1–5, 2006.

[8] M. Markiewicz, "Mathematical modeling of heavy gas atmospheric dispersion over complex and obstructed terrain," *Archives of Environmental Protection*, vol. Vol. 36, no. 1, pp. 81–94, 2010.

[9] Y.-N. Lien, H.-C. Jang, and T.-C. Tsai, "A manet based emergency communication and information system for catastrophic natural disasters," in *29th IEEE International Conference on Distributed Computing Systems Workshops, 2009. ICDCS Workshops '09.*, June 2009. doi: 10.1109/ICDCSW.2009.72. ISSN 1545-0678 pp. 412–417.

[10] Y.-N. Lien, L.-C. Chi, and C.-C. Huang, "A multi-hop walkie-talkie-like emergency communication system for catastrophic natural disasters," in *39th International Conference on Parallel Processing Workshops (ICPPW), 2010*, Sept 2010. doi: 10.1109/ICPPW.2010.77. ISSN 1530-2016 pp. 527–532.

[11] M. Aloqaily, S. Otoum, and H. Mouftah, "A novel communication system for firefighters using audio/video conferencing/sub-conferencing in standalone manets," in *5th International Conference on Computer Science and Information Technology (CSIT), 2013*, March 2013. doi: 10.1109/CSIT.2013.6588764 pp. 89–98.

[12] J. Kim, D. Kim, S. Jung, M. Lee, K. Kim, C. Lee, J. Nah, S. Lee, J. Kim, W. Choi, and S. Yoo, "Implementation and performance evaluation of mobile ad hoc network for emergency telemedicine system in disaster areas," in *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*, Sept 2009. doi: 10.1109/IEMBS.2009.5333889. ISSN 1557-170X pp. 1663–1666.

[13] E. Kulla, R. Ozaki, A. Uejima, H. Shimada, K. Katayama, and N. Nishihara, "Real world emergency scenario using manet in indoor environment: Experimental data," in *Ninth International Conference on Complex, Intelligent, and Software Intensive Systems (CISIS), 2015*, July 2015. doi: 10.1109/CISIS.2015.49 pp. 336–341.

[14] T. Aurisch and J. Tölle, "Relay placement for ad-hoc networks in crisis and emergency scenarios," in *Proceedings of the Information Sys-*

tems and Technology Panel Symposium (IST-091), Bucharest, Romania, vol. 11, 2009.

[15] A. Martín-Campillo, R. Martí, S. Robles, and C. Martínez-García, "Mobile agents for critical medical information retrieving from the emergency scene," in *7th International Conference on Practical Applications of Agents and Multi-Agent Systems (PAAMS 2009)*, ser. Advances in Intelligent and Soft Computing, Y. Demazeau, J. Pavón, J. Corchado, and J. Bajo, Eds. Springer Berlin Heidelberg, 2009, vol. 55, pp. 30–39. ISBN 978-3-642-00486-5. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-00487-2_4

[16] R. Martí, S. Robles, A. Martín-Campillo, and J. Cucurull, "Providing early resource allocation during emergencies: The mobile triage tag," *Journal of Network and Computer Applications*, vol. 32, no. 6, pp. 1167 – 1182, 2009. doi: http://dx.doi.org/10.1016/j.jnca.2009.05.006. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1084804509000769

[17] Y. Hayakawa, K. Mori, Y. Ishida, K. Tsudaka, T. Wada, H. Okada, and K. Ohtsuki, "Development of emergency rescue evacuation support system in panic-type disasters," in *Consumer Communications and Networking Conference (CCNC), 2012 IEEE*, Jan 2012. doi: 10.1109/CCNC.2012.6181047 pp. 52–53.

[18] T. Tsunemine, E. Kadokawa, Y. Ueda, J. Fukumoto, T. Wada, K. Ohtsuki, and H. Okada, "Emergency urgent communications for searching evacuation route in a local disaster," in *Consumer Communications and Networking Conference, 2008. CCNC 2008. 5th IEEE*, Jan 2008. doi: 10.1109/ccnc08.2007.267 pp. 1196–1200.

[19] T. Nakamura, K. Kogo, J. Fujimura, K. Tsudaka, T. Wada, K. Ohtsuki, and H. Okada, "Development of emergency rescue evacuation support system (eress) in panic-type disasters: Disaster detection by positioning area of terminals," in *42nd International Conference on Parallel Processing (ICPP), 2013*, Oct 2013. doi: 10.1109/ICPP.2013.111. ISSN 0190-3918 pp. 931–936.

[20] M. Zhong and C. Cassandras, "Distributed coverage control and data collection with mobile sensor networks," *Automatic Control, IEEE Transactions on*, vol. 56, no. 10, pp. 2445–2455, Oct 2011. doi: 10.1109/TAC.2011.2163860

[21] M. T. Thai, R. Tiwari, R. Bose, and A. Helal, "On detection and tracking of variant phenomena clouds," *ACM Trans. Sen. Netw.*, vol. 10, no. 2, pp. 34:1–34:33, Jan. 2014. doi: 10.1145/2530525. [Online]. Available: http://doi.acm.org/10.1145/2530525

[22] V. Subramanian, A. Umbarkar, and A. Doboli, "Decentralized detection and tracking of emergent kinetic data for wireless grids of embedded sensors," in *Conference on Adaptive Hardware and Systems (AHS), 2012 NASA/ESA*. IEEE, 2012. doi: 10.1109/AHS.2012.6268650 pp. 198–204.

[23] Y. Liu and M. Li, "Iso-map: Energy-efficient contour mapping in wireless sensor networks," in *27th International Conference on Distributed Computing Systems, 2007. ICDCS '07.*, June 2007. doi: 10.1109/ICDCS.2007.115. ISSN 1063-6927 pp. 36–36.

[24] J.-H. Kim, K.-B. Kim, C. S. Hussain, M.-W. Cui, and M.-S. Park, "Energy-efficient tracking of continuous objects in wireless sensor networks," in *Ubiquitous Intelligence and Computing*. Springer, 2008, pp. 323–337.

[25] S. Duttagupta, K. Ramamritham, and P. Ramanathan, "Distributed boundary estimation using sensor networks," in *IEEE International Conference on Mobile Adhoc and Sensor Systems (MASS), 2006*, Oct 2006. doi: 10.1109/MOBHOC.2006.278571 pp. 316–325.

[26] H. Hong, S. Oh, J. Lee, and S.-H. Kim, "A chaining selective wakeup strategy for a robust continuous object tracking in practical wireless sensor networks," in *IEEE 27th International Conference on Advanced Information Networking and Applications (AINA), 2013*. IEEE, 2013. doi: 10.1109/AINA.2013.131 pp. 333–339.

[27] K. Matsuo, K. Goto, A. Kanzaki, T. Hara, and S. Nishio, "Overhearing-based efficient boundary detection in dense mobile wireless sensor networks," in *IEEE 15th International Conference on Mobile Data Management (MDM), 2014*, vol. 1. IEEE, 2014. doi: 10.1109/MDM.2014.34 pp. 225–234.

[28] G. Keung, B. Li, Q. Zhang, and H.-D. Yang, "The target tracking in mobile sensor networks," in *Global Telecommunications Conference (GLOBECOM 2011), 2011 IEEE*, Dec 2011. doi: 10.1109/GLOCOM.2011.6134188. ISSN 1930-529X pp. 1–5.

[29] S. Srinivasan, S. Dattagupta, P. Kulkarni, and K. Ramamritham, "A survey of sensory data boundary estimation, covering and tracking techniques using collaborating sensors," *Pervasive and Mobile Computing*, vol. 8, no. 3, pp. 358–375, 2012. doi: 10.1016/j.pmcj.2012.03.003

[30] J. Brink and E. Pebesma, "Plume tracking with a mobile sensor based on incomplete and imprecise information," *Transactions in GIS*, vol. 18, no. 5, pp. 740–766, 2014. doi: 10.1111/tgis.12063

[31] T. Sun, H. Pei, Y. Pan, and C. Zhang, "Robust adaptive neural network control for environmental boundary tracking by mobile robots," *International Journal of Robust and Nonlinear Control*, vol. 23, no. 2, pp. 123–136, 2013. doi: 10.1002/rnc.1816. [Online]. Available: http://dx.doi.org/10.1002/rnc.1816

[32] Z. Jin and A. Bertozzi, "Environmental boundary tracking and estimation using multiple autonomous vehicles," in *46th IEEE Conference on Decision and Control, 2007*, Dec 2007. doi: 10.1109/CDC.2007.4434857. ISSN 0191-2216 pp. 4918–4923.

[33] A. Joshi, T. Ashley, Y. R. Huang, and A. L. Bertozzi, "Experimental validation of cooperative environmental boundary tracking with on-board sensors," in *American Control Conference, 2009. ACC'09.* IEEE, 2009. doi: 10.1109/ACC.2009.5159837 pp. 2630–2635.

[34] J. Brink, "Boundary tracking and estimation of pollutant plumes with a mobile sensor in a low-density static sensor network," *Urban Climate*, pp. –, 2014. doi: http://dx.doi.org/10.1016/j.uclim.2014.07.002. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S2212095514000492

[35] S. Susca, F. Bullo, and S. Martínez, "Monitoring environmental boundaries with a robotic sensor network," *Control Systems Technology, IEEE Transactions on*, vol. 16, no. 2, pp. 288–296, 2008. doi: 10.1109/TCST.2007.903395

[36] J. Clark and R. Fierro, "Mobile robotic sensors for perimeter detection and tracking," *ISA transactions*, vol. 46, no. 1, pp. 3–13, 2007. doi: 10.1016/j.isatra.2006.08.001

[37] D. Marthaler and A. L. Bertozzi, "Collective motion algorithms for determining environmental boundaries," *Autonomous Robots, special issue on Swarming, submitted for publication*, 2003.

[38] I. Triandaf and I. B. Schwartz, "A collective motion algorithm for tracking time-dependent boundaries," *Mathematics and Computers in Simulation*, vol. 70, no. 4, pp. 187 – 202, 2005. doi: http://dx.doi.org/10.1016/j.matcom.2005.07.001. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0378475405001850

[39] S. Srinivasan, "Contour estimation using collaborating mobile sensors," in *In DIWANS '06: Proceedings of the 2006 workshop on Dependability issues in wireless ad hoc networks and sensor networks*. ACM, 2006. doi: 10.1145/1160972.1160986 pp. 73–82.

[40] M. Kemp, A. L. Bertozzi, and D. Marthaler, "Multi-uuv perimeter surveillance," in *Proceedings of*, 2004. doi: 10.1109/AUV.2004.1431200 pp. 102–107.

[41] S. Subchan, B. A. White, A. Tsourdos, M. Shanmugavel, and R. Zbikowski, "Dubins path planning of multiple uavs for tracking contaminant cloud," in *Proceedings of the 17th World Conference on the International Federation of Automatic Control, Seoul, Korea*, 2008. doi: 10.3182/20080706-5-KR-1001.00964 pp. 6–11.

[42] B. White, A. Tsourdos, I. Ashokaraj, S. Subchan, R. Żbikowski *et al.*, "Contaminant cloud boundary monitoring using network of uav sensors," *Sensors Journal, IEEE*, vol. 8, no. 10, pp. 1681–1692, 2008. doi: 10.1109/JSEN.2008.2004298

[43] A. Sinha, A. Tsourdos, and B. White, "Multi {UAV} coordination for tracking the dispersion of a contaminant cloud in an urban region," *European Journal of Control*, vol. 15, no. 3–4, pp. 441 – 448, 2009. doi: http://dx.doi.org/10.3166/ejc.15.441-448. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0947358009709999

[44] D. W. Casbeer, R. W. Beard, T. W. McLain, S.-M. Li, and R. K. Mehra, "Forest fire monitoring with multiple small uavs," in *American Control Conference, 2005. Proceedings of the 2005*. IEEE, 2005. doi: 10.1109/ACC.2005.1470520 pp. 3530–3535.

[45] S. Srinivasan, K. Ramamritham, and P. Kulkarni, "Ace in the hole: Adaptive contour estimation using collaborating mobile sensors," in *International Conference on Information Processing in Sensor Networks, 2008. IPSN'08.* IEEE, 2008. doi: 10.1109/IPSN.2008.38 pp. 147–158.

[46] E. Niewiadomska-Szynkiewicz, A. Sikora, and J. Kołodziej, "Modeling mobility in cooperative ad hoc networks," *Mobile Networks and Applications*, vol. 18, no. 5, pp. 610–621, 2013. doi: 10.1007/s11036-013-0450-2

[47] A. Sikora, E. Niewiadomska-Szynkiewicz, and M. Krzyszton, "Simulation of mobile wireless ad hoc networks for emergency situation awareness," in *Federated Conference on Computer Science and Information Systems (FedCSIS), 2015*. IEEE, 2015. doi: 10.15439/2015F52 pp. 1087–1095.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1085–1088
DOI: 10.15439/2016F286
ACSIS, Vol. 8. ISSN 2300-5963

Juraj Miček, Ondrej Karpiš, Veronika Olešnaníková
University of Zilina
Univerzitna 8215/1
010 26 Zilina, Slovakia
Email: {Juraj.Micek, Ondrej.Karpis, Veronika.Olesnanikova}@fri.uniza.sk

Mobile sensor elements based on robotic platform YROBOT

***Abstract*—During the years 2014-2015 a kit called Yrobot was developed at the Department of technical cybernetics. The kit is a mobile robotic platform designed mainly to support technical education at high schools and universities. Since the kit was presented at conferences RAAD 2014 and EDERC 2014 for the first time, several expansion modules and interesting applications were developed. One application is presented in this paper. The application uses the mobility of wireless sensors to map the chosen area. As an example, we realized RSSI measurements and visualized them as a function of sensor position. Obviously, one can find many similar tasks: mapping of temperature or gas concentration in given space and the like. Our ultimate goal is to analyze the possibilities of using RSSI measurements for indoor localization. However, at this stage of research, we were focused just on acquiring the data and their subsequent visualization.**

I. Introduction

IN PREVIOUS years, simple mobile system Yrobot was developed at the Department of technical cybernetics. This system was designed especially for teaching IT subjects in the high schools. The concept of the system, its features and functions, as well as first experience of deployment of the system in teaching, were presented at the conference RAAD in 2014 [2, 3]. Original author's intent was as following: "By using a simple technical device to increase the motivation of high school students in studying technical fields, particularly in information technologies." Thanks to the long-term support of the Volkswagen Slovakia Foundation, Yrobot kits were delivered to the high schools. Our team prepared detailed textbooks also [1]. Based on the Yrobot platform, competition for high school students "Program the robot - Yrobot Cup" was organized and other supporting activities were carried out. Note that the robotic platform has the nature of open source hardware and thus, we suppose the development of other applications (hardware and software modules) directly by high schools and universities students. In addition to the typical educational mission, the platform allows verification of various methods of information processing. The data can be obtained from the real environment using different sensor modules.

Many modules were developed to sense:

- temperature and humidity,
- parameters of magnetic fields,
- acceleration,
- illumination,
- sound (20 Hz to 4 kHz),
- concentrations of CO2, CO, NOx,
- distance (using ultrasound),
- optical reflectivity of the surface.

Connection of sensors with mobile platform brings many solutions that often go beyond the simple application tasks.

The mobile platform is formed by a simple MCU ATmega 16, a pair of DC motors and supporting electronics. The block structure of the platform is shown in Fig.1.

Wireless connectivity of mobile robots can further expand set of tasks that can be successfully solved by Yrobot platform. Therefore, we developed a communication module providing wireless communication between the mobile platforms.

II. RF Communication Modules

We assume that wireless connectivity of mobile robots extend significantly the application potential of the platform. Recall that Yrobot was originally developed as an autonomous device capable of solving simple tasks based on the status of its sensors (line following, obstacles avoiding, area browsing). Wireless communication changes the autonomous Yrobot platform to the cooperative system of multiple robots for solving problems using data fusion. Providing connectivity allows transition from autonomous to multi-robotic system and the robust solutions of complex challenges. To allow effective communication between elements of the system a variety of communication technologies (protocols and network topologies) can be used. First, we have decided to implement three separate communication modules operating in the ISM 2.4 GHz band.

In table I. are shown main parameters of selected communication technologies.

All developed modules are connected via connectors of the motherboard to the microcontroller on Yrobot. The motherboard provides power supply to the communication modules. Communication with modules is serial, either asynchronous (UART) or synchronous (SPI). The module contains circuitry to ensure power supply, logic signals level shift and selection of communication line UART/SPI. LEDs are used to signalize status of communication module. The buttons on the module allow restarting of the module, firmware change or change of the operation mode.

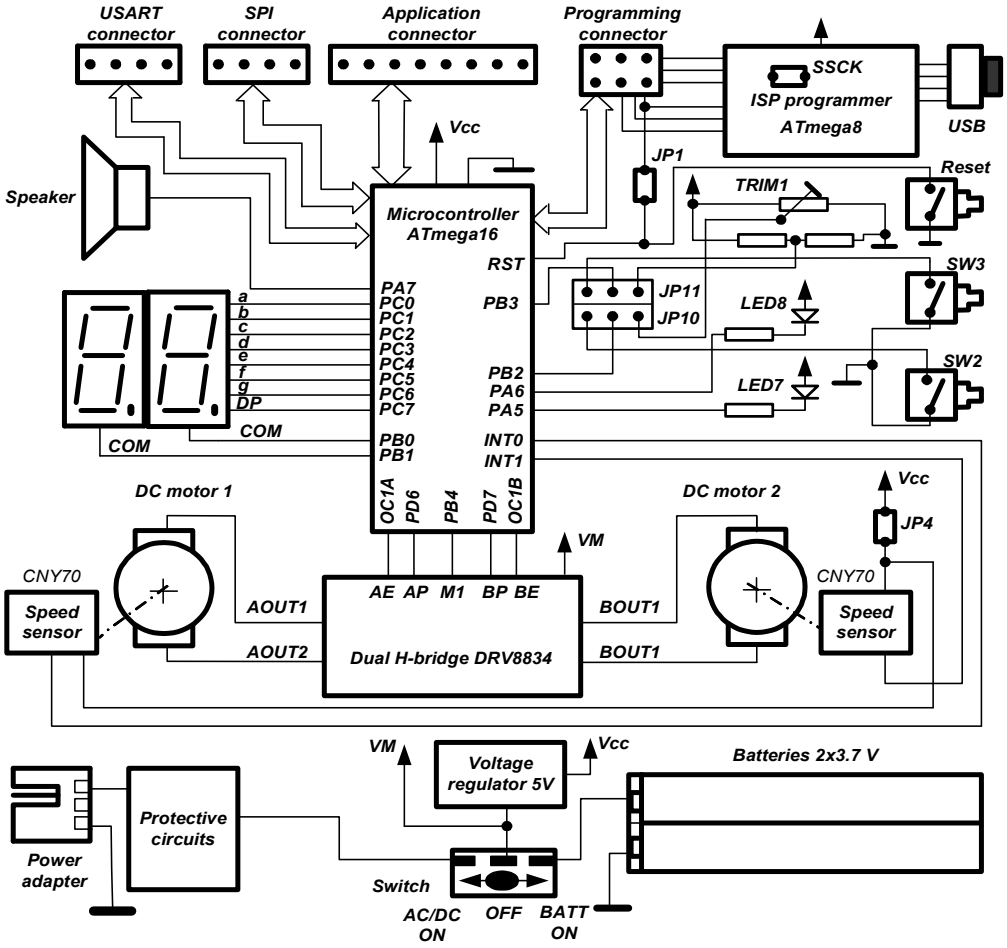


Fig 1. Block structure of Yrobot mobile platform

III. Communication Module Y-WiFi

As a first module, we developed communication module Y-WiFi. The essential element of this system is WizFi250 communication module. The WizFi250 module has an integrated LNA, printed antenna, and connector for external antenna. It offers secure communication based on protocols WEP, WPA, and WPA2. It is controlled by a set of AT commands.

Module Y-WiFi includes all other components that allow easy connection to the mobile Yrobot platform or to a personal computer. The module Y-WiFi can be connected to the PC via the USB port. FT323RL circuit was used to convert UART to USB. Block diagram of the module is shown in Fig. 2.

The module contains two buttons S1 and S2 that allow selection of basic operation mode and restarting of the module. Status of the WizFi250 module is displayed using two LEDs – LED1 is on when the connection is established and LED 2 is on in data mode. Status of the FT323RL circuit is likewise indicated by LED3 (pulsing when data are transmitted) and LED4 (pulsing when data are received). LED5 is used to indicate the power-supply.

Jumper JMP1 allows selection between application or BOOT mode in which can be updated module firmware.

Table I.
Communication technologies

Network technology	Standard	Frequency band	Embedded module	Max bit rate
WiFi	802.11 b/g/n	2.4 GHz	WizFi250	65 Mbps
Zigbee	802.15.4	2.4 GHz	JN5168M0	1Mbps
			MRF24J40	250 kbs
Bluetooth v. 2.0 EDR	802.15.1	2.4 GHz	BTM182	3 Mbps

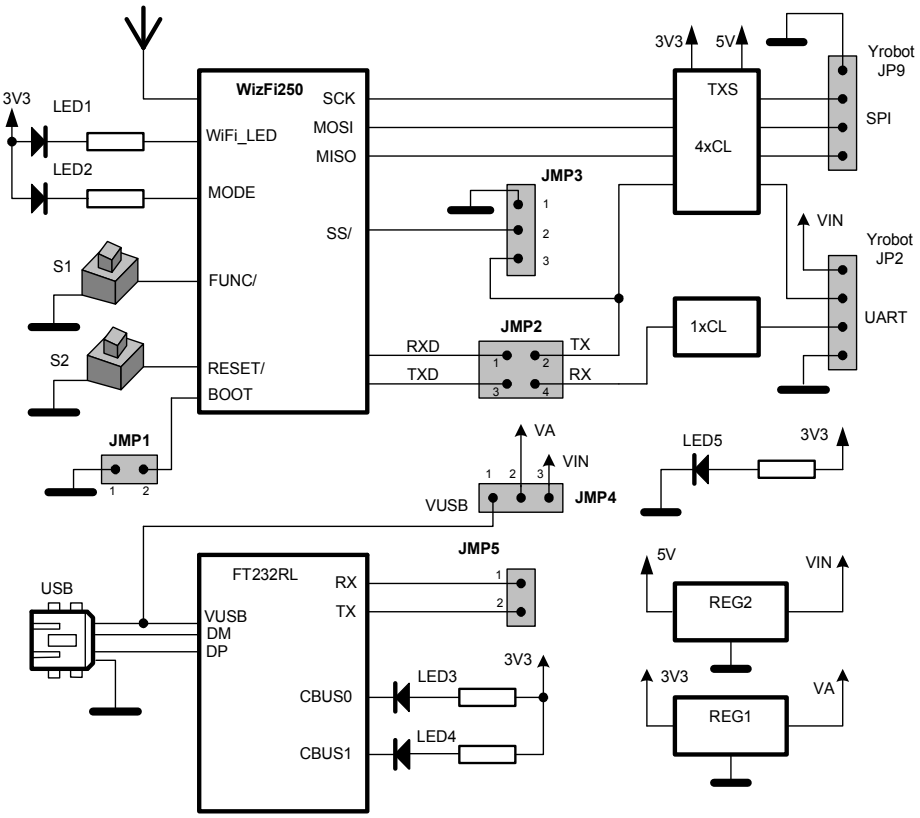


Fig 2. Block structure of Yrobot mobile platform

Jumpers JMP2 and JMP5 select connection of communication module WizFi250 to a PC (via USB) or to the Yrobot platform. It is also possible to connect a PC to the Yrobot. Jumper JMP3 is used when communicating with the module using SPI interface. Finally, with a jumper JMP4 we choose source of power supply for the module – PC USB or Yrobot batteries.

IV. Experiment

Using the described technical means, we conducted an experiment to determine the signal strength map in the 2.4 GHz band. The experiment was carried out in the hallway of our workplace. WiFi internet connection in the hallway and in adjacent rooms is provided by Access Point Planet WAP-4000 with the following parameters: standard 802.11g, channel 7, authentication: WPA-PSK, transmitter power set to a minimum (-12 dB). Transmitter power during the measurement was set to a minimum in order to magnify the signal attenuation in places distant from the AP. The measurement was carried out at the time when the hallway was empty, as the presence of people could influence the results.

Yrobot was programmed so that it can be controlled via a notebook. The Y-WiFi module was configured in Station Mode with a fixed IP address. The Y-WiFi module had an activated TCP server that listens on port 5000. Yrobot movement was based on orders received by the server. The speed of rotation of the robot's wheels was set at about 1 revolution per second. This corresponds to a speed of 0.2 m/s. After each half turn of the wheel (e.g. every 0.1 m), the strength of the signal (RSSI) was measured and sent to the client (notebook).

Control notebook was also connected to a WiFi network and using the Realterm application it was connected to the Y-WiFi module. Realterm application was used for the reason that it allows to send pressed buttons to the server and record the received data to a file simultaneously. Fig. 3 shows the robotic platform Yrobot with Y-WiFi module used in the experiment.

Fig. 4 depicts a plan view of the hallway and the route of the robot during the experiment. The distance of adjacent tracks was 0.3 m. The whole experiment was divided into three parts – the two short passages and a larger central hall were measured independently. After measurements, the three sets of data were combined into one. RSSI was measured about 1900 times during the experiment. The data were interpolated using Matlab in order to obtain values with a spatial resolution of 0.1 m. The measured values are shown in Fig. 5. The map is consistent with our assumptions that the weakest signal is in the areas that are shaded by walls.

Note that during the measurement the antenna of Y-WiFi module was only about 7 cm above the ground. The situation in different height may be different.

The acquired map can be partially used for determining the position of the robot based on RSSI measurements. Obviously, the estimation of the position of the robot based on a single measurement is not possible. Significantly more

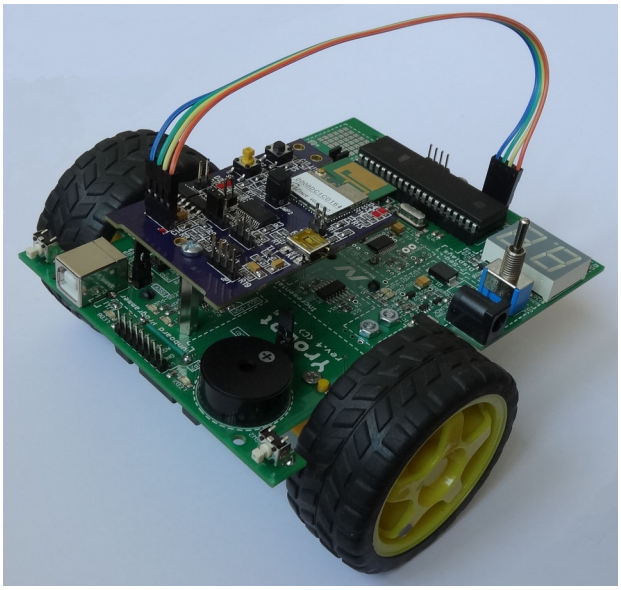


Fig 3. Mobile platform Yrobot with Y-WiFi module

precise localization should be possible with a map obtained by a directional antenna that can be rotated in different directions. Such a system is currently under development.

V. Conclusion

Extension of the Yrobot platform with network modules significantly expands the variety of applications that can be realized with it. Based on the experiment with indoor RSSI measurements it can be stated that the Yrobot system enhanced by communication modules with wireless connectivity provides more functionality and allows development of many interesting applications such as precise localization of mobile systems using signal strength measurements.

In the next step we want to integrate the mobile platform into IoT environment using ThingWorx suite. We would like to expand the possibilities of the Yrobot platform further by developing of the additional modules for technologies such as RFID, NFC, Z-Wave and a chosen proprietary communication systems in the ISM bands (eg. RFM70).

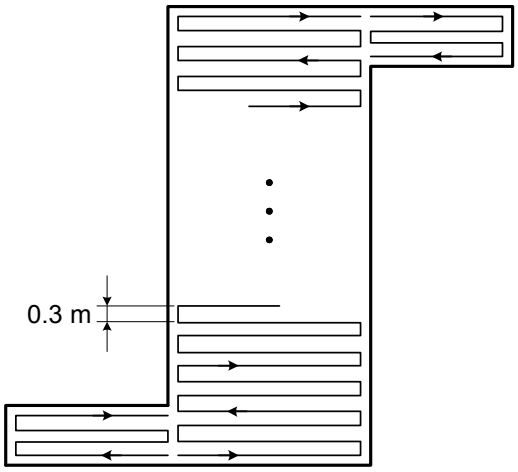


Fig 4. Plan view of the hallway and a robot route

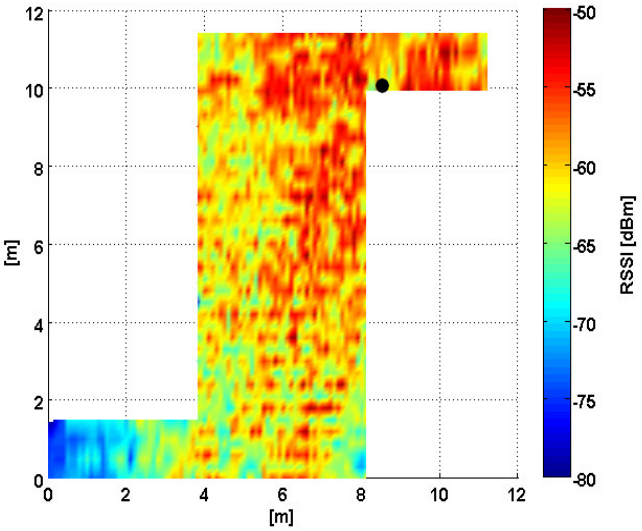


Fig 5. Signal strength map. The black point is AP.

References

[1] J. Miček et. al., *Sprievodca po svete Yrobota*, University of Žilina, 2015.

[2] M. Kochláň, M. Hodoň, "Open hardware modular educational robotic platform – Yrobot", in *Proc. 23rd International Conference on Robotics in Alpe-Adria-Danube Region (RAAD)*, Slovakia, 2014. http://dx.doi.org/10.1109/RAAD.2014.7002246

[3] J. Miček, O. Karpiš, "Audio communication subsystem of multi-robotic system YROBOT", in *Proc. 23rd International Conference on Robotics in Alpe-Adria-Danube Region (RAAD)*, Slovakia, 2014. http://dx.doi.org/10.1109/RAAD.2014.7002263

[4] J. Miček, O. Karpiš, M. Kochláň, "Audio-communication subsystem module for Yrobot - a modular educational robotic platform", in *Proc. 6th European Embedded design in education and research (EDERC)*, Milano, 2014, pp. 60-64. http://dx.doi.org/10.1109/EDERC.2014.6924359

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1089–1092
DOI: 10.15439/2016F262
ACSIS, Vol. 8. ISSN 2300-5963

Unicast Routing on VANETs

Boubakeur Moussaoui‡¶, Hacène Fouchal∓, Marwane Ayaida∓ and Salah Mermiz ¶
‡ Laboratoire D'électronique et des Télécommunications Avancées (ETA),
Université Mohamed Bachir El Ibrahimi de Bordj Bou Arreridj, 34031, Algeria
∓ Centre de recherche CReSTIC,
Université de Reims Champagne-Ardenne
51687 REIMS Cedex 2, France
¶ Université de Constantine -Abdelhamid Mehri- Algeria
Moussaoui.bkr@gmail.com, {hacene.fouchal, marwane.ayaida }@univ-reims.fr, s_merniz@hotmail.com

***Abstract*—Greedy routing in VANETs requires some geographical informations, such as the source location and the destination location. The first one could be obtained using some localization devices like GPS receiver. However, the second one is provided by a location service. This later has a high overhead especially if it is implemented over V2V (vehicle to vehicle) communications. Many location services are well known as HLS, RLS, GLS.**

This paper is interested in reducing this overhead by using some Road Side Units (RSU) already deployed along the roads. We propose here a location service called "improved Reactive Location Service (iRLS)", which is an extension of the RLS service. The major difference is that RLS assumes only V2V communications and iRLS takes profit of a wireless backbone based on RSUs to catch the destination's position. This allows to reduce the overhead instead of flooding requests and also makes the communication faster since we will not have to wait to the request to reach the destination before receiving the response and then starting sending data. In our proposal, the closest RSU will reply with the actual location.

In order to show the contribution of our approach, we have conducted some simulations that prove that iRLS outperforms any geographic protocol by using the V2I communications in terms of end-to-end delay which is one of the most important parameter. We considered also the ratio of packet received correctly by the destination vehicle (PDR), our protocol improves significantly this second parameter, and ensures more than 20% of packets received correctly

***Index Terms*—VANETs; Location-based Services; Geographic Routing Protocols, .**

I. Introduction

VANETs (Vehicular Ad-hoc NETworks) are a special case of MANETs (Mobile Ad-hoc NETworks). Therefore, the routing protocols used for MANETs could be adapted for MANETs. However, the topology-based protocols in MANETs are not useful for VANETs because of the high mobility rate. Then, geographic routing protocols are more suitable for such networks since they are more scalable.

The geographic routing protocols need as an input the node's location as well as the destination's location. The node's position is easy to obtain. However the destination's position is obtained from a location service. This service is responsible to reply to requests like : "Where is the node X?". "Reactive Location Service" (RLS) is a well known location service. When a node needs to find the position of another node, it floods a request. When this request reaches the destination, this latter replies with its current position using the same path for the request delivery. Therefore, RLS uses only V2V communication without considering the V2I (vehicle to infrastructure) ones. For this purpose, we propose a new location service denoted "improved Reactive Location Service" (iRLS). This service uses mainly the RSU network to reply directly to the location request in order to avoid previous flooding. This allows to gain in term of overhead and delay as demonstrated by the simulations that have been conducted using the well known NS-3 simulator.

The paper is organized as follows. The section II presents some works related to the location-based service and the geographic routing protocols. Section III described our iRLS proposal. Section IV introduces the simulations conducted with our proposal. Finally, section V concludes the paper and presents some future works.

II. Related works

In this section, we will describe some related works, published these last two decades. All these works focused on routing in VANET environments where most of solutions used V2V communications and/or V2I communications. They do not consider position changes during the dissemination process. Both the source node and the destination node are vehicles in general case. To deal with the high moving speed and dynamic changes of the topology, RSUs can provide a lot of benefits in order to achieve efficient routing of messages over the network. In order to propose a realistic solution,, we do not need to deploy an RSU at each road intersection. An overview on routing for mobile and vehicular ad hoc networks is well detailed in [1].

In [2], authors propose an improved version of DGRP (Directional Greedy Routing Protocol). In addition to the information used by DGRP (position, speed, direction) provided by a GPS sensor, they consider and evaluate link stability. As a result, they reduce links breakage, enhance reliability of a used path and ensure a high packet delivery ratio.

Another proposed work which assists routing in VANETs using base station was presented in [3]. Roads are composed of segments, each of them is governed by a base station. This later sends a response to a route request packet by choosing the shortest path using RSUs. But if we do not

have any available path, the RSU checks if there is enough free bandwidth to grant the request. They have used two mechanisms, using Fast Learning Neural Networks, to prevent link failure and to predict bandwidth consumption during handoffs. An alternative path will be established, before that a broken link occurs.

Authors in [4] have proposed an infrastructure-assisted routing protocol. The main benefits behind this approach is to reduce the routing overhead and improve the end-to-end performance. A backbone network ensures connectivity by using RSUs, an enhancing search of the shortest path to reach the destination is done by this infrastructure. Authors present an extended protocol to the topology-aware GRS routing protocol. At any intersection, anchor nodes in the GRS system compute the shortest path to reach the destination using the Djikstra algorithm without considering the road density or any other parameter which can ensure connectivity of roads.

Roadside-Aided Routing (RAR) has been proposed in [5]. It is another study which prefers V2I communication to improve the search for a stable routes over VANETs networks.

A. Location-based services

Location-based services can be classified into two classes : "Flooding-based" and "Rendez-vous-based". The first class is composed of reactive and proactive services. In the proactive flooding-based location-based service, every node floods its geographic information through all the network periodically. Thus, all the nodes are able to update their location tables. Since this approach uses flooding and may surcharge the network by location update messages, several techniques to reduce the congestion were used. One of them is to tune the update frequency with the node mobility (the more node is moving fast, the higher update location frequency is used).

Therefore, the update frequency decreases with the distance to the node. The second idea is, a node with high mobility sends more update location packets. As a result, there are less packets than a simple flooding scheme without affecting the network performances. For the second group (i.e the reactive flooding-based location-based service), the location response is sent when receiving a location request. This avoids the overhead of useless location information of some nodes updated and never used. But, it adds high latencies not suitable in VANETs. One of these known services is Reactive Location Service (RLS) [6].

In the second class (rendez-vous-based location service), all the nodes agree on a unique mapping of a node to other specific nodes. The geographic information are disseminated through the elected nodes called the "location servers".

Thus, the location-based services consists of two components :

1) Location Update : A node has to recruit location servers (chosen from other nodes) and needs to update its location through theses servers. The location servers are responsible of storing the geographic data of the relating nodes.
2) Location Request : When a node needs to know the location of another node, it broadcasts a location request. The location server will replay as soon as it receives this request.

B. Geographic Routing Protocols

Routing protocols algorithms must choose some criteria to make routing decisions, for instance the number of hops, latency, transmission power, bandwidth, etc. The topology-based routing protocols suffer from heavy discovery and maintenance phases, lack of scalability and high mobility effects (short links). Although, geographic routing are suitable for large scale dynamic networks. The first routing protocol using the geographic information is the *Location-Aided Routing (LAR)* [7]. This protocol used the geographic information in the route discovery. This latter is initiated in a *Request Zone*. If the request doesn't succeed, it initiates another request with a larger *Request Zone* and the decision is made on a routing table. The first real geographic routing protocol is the *Greedy Perimeter Stateless Routing (GPSR)* [8]. It is a reactive protocol which forwards the packet to the target's nearest neighbor (Greedy Forwarding approach) until reaching the destination. Therefore, it scales better than the topology-based protocols, but it does still not consider the urban streets topology and the existence of obstacles to radio transmissions. Another geographic routing protocol is the *Geographic Source Routing (GSR)* [9]. It combines geographical information and urban topology (street awareness). The sender calculates the shorter path (using Djikstra algorithm) to the destination from a map location information. Then, it selects a sequence of intersections (anchor-based) by which the data packet has to travel, thus forming the shortest path routing. To send messages from one intersection to another, it uses the greedy forwarding approach. The choice of intersections is fixed and does not consider the spatial and temporal traffic variations. Therefore, it increases the risk of choosing streets where the connectivity is not guaranteed and losing packets.

In [10], authors propose an improved version of DGRP (Directional Greedy Routing Protocol). In addition to the information used by DGPR (position, speed, direction) provided by GPS, they consider and evaluate link stability. As a result, they reduce links breakage, enhance reliability of a used path and ensure a high packet delivery ratio.

III. Improved Location-based Service

To better understand our proposal, we present a simple scenario in the Figure 1. We suppose that the vehicle V1 has data to send to the vehicle V6, which is not in its direct neighborhood. If V1 has a valid and fresh route in its routing table, it sends immediately the data without any routing service requirement, otherwise it broadcast a request to find a suitable route to use. Since V1 is in the RSU4's range, V1 sends a request to RSU4 asking for the new V6's location. RSUs are connected together and exchange vehicles' positions when needed. After receiving the response from RSU4, V1 starts sending data using the greedy approach, where it selects

the closest neighbor to the destination V6 (here V8). Then, data packets travel through V8, V7 and finally to reach V6. Therefore, destination's location could be found more rapidly than using V2V communications when a destination node is far away from the source node;

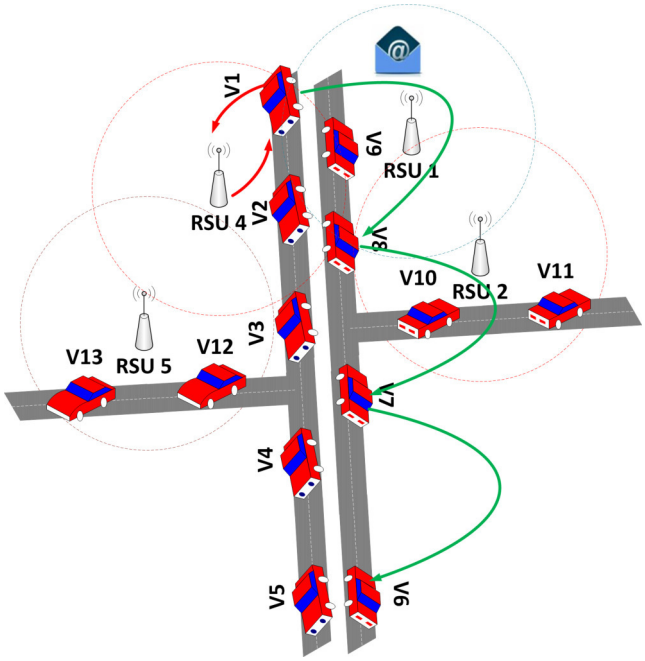


Figure 1. An example of a scenario with iRLS

The algorithm 1 presents more details how the iRLS service works. When starting, the source s looks if it is not in the range of a RSU, then it works like RLS. It floods a request to find the location of the destination d. When the request reaches the node Z that has this information, the latter replies directly to s with d's new position. Otherwise, the node s sends the request in unicast way to this RSU (RSUX). RSUX forwards this request to all connected RSUs. When RSUY receives the request and it has already this information, it replies to RSUX. RSUX then transmits the d's position to node s. Therefore, s has all required information to start sending data. These data are forwarded using the greedy approach until reaching the destination. By reducing the V2V search phase, the available limited bandwidth of our links will be used to transmit the effective data rather than the transmission of control packets. Indeed, our protocol ensures a higher PDR than the standard protocol (see Figure 3).

In the case where a node S has no RSU in its neighborhood, it first sends a request to find the closest RSU. This request is run in a setup step before starting the routing process.

But in some cases, there is not enough vehicle density and the initial request could not be sent (caused by a lack of connectivity). We suggest to resend the packet after a defined time (timeout) until a reply is received from a close RSU.

Algorithm 1 iRLS Location Service

```
1:  function iRLS(s, d, Data)
2:      Initialization : s source, d destination
3:      if (source ∉ RSUX's range) then
4:          s broadcasts requests to find d's position
5:          neighbors floods the request
6:          node Z has d's location and replies directly to s
7:          s sends location request to RSU to find d's position
8:      else
9:          RSU sends request to other RSUs
10:         RSUY has d's location and replies to RSUX
11:         RSU forwards response to s
12:     end if
13:     next ← s
14:     while (next != d) do
15:         choose n the best next hop using greedy approach
16:         s sends data to n
17:         next ← n
18:     end while
19:     Data reaches destination d
20: end function
```

IV. Simulations

A. Working Environment

The simulations were performed using the Ns-2 simulator 2.33 [11]. The geographic routing protocol used is the Greedy Perimeter Stateless Routing (GPSR) [8]. The chosen area is a 2x2 km^2 of a real map representing part of the French city *Reims*. This area is extracted from Open Street Map [12]. The MAC layer used is 802.11p [13]. The parameters used in the simulation are summarized in the Tab. I.

At each simulation, every node initiates 4 CBR traffics of 100 packets with a size of 128 KB to 4 random destination nodes with a second of interval between each sent message. The CBR traffic simulates for example an audio or a video streaming. It may be used in security applications, such as viewing the video stream from a camera located on a bus by the police car or the security agent vehicle. Also, this traffic could be used in entertainment applications to connect to the Internet or to play online video games.

B. Experimentation Results

Our experimentations have provided the following figures. On each figure, we show the network behavior with usual RLS (denoted RLS curve) and the behavior of our proposed solution (denoted iRLS curve for improved RLS). Figure 2 shows the delay to send a message depending on the size of the network from 10 nodes to 100 nodes. we observe that the delay to send a message from a node s to d decreases with iRLS even if the number of nodes is higher. The gain is around 10 per cent. In Figure 3, the results highlight that our protocol can achieve a higher PDR than the RLS one. In the density 100 vehicles, it is clear that taking the same scenario of simulation, our protocol guarantee the same ratio but when using RLS this ratio decreases.

Table I
THE SIMULATION PARAMETERS

Parameters	Value
Channel type	Channel/WirelessChannel
Propagation model	Propagation/TwoRayGround
Network interface	Phy/WirelessPhyExt
MAC layer	802.11p [13]
Interface queue type	Queue/DropTail/PriQueue
Link layer	LL
Antenna model	Antenna/OmniAntenna
Interface queue length	512 packets
Ad-hoc routing protocol	GPSR
Location-based service	RLS
Location cache maximum age	4, 8, 12, 16, and 22 s
Area	2x2 km^2
Number of nodes	10-150
Simulation time	150 s
GPRS beacon interval	0,5 s
CBR traffic	4 x 100 packets / node
CBR packet size	128 KB
CBR sent interval	1 s

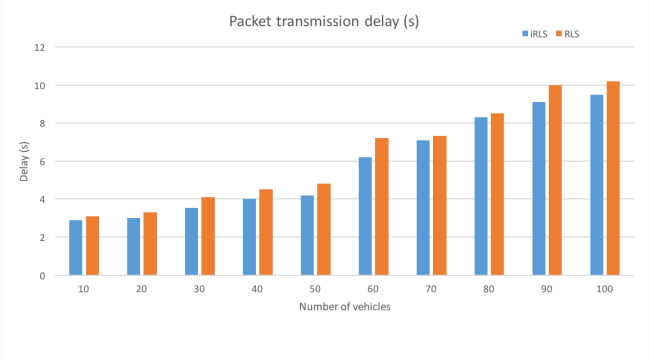


Figure 2. The average delay for packet transmission

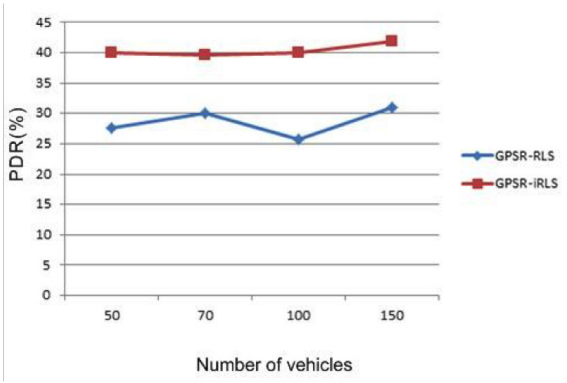


Figure 3. Impact of the number of vehicles on the achieved PDR average

These results show that our protocol outperform by more than 10% in general case the use of RLS. All these result can be explained by the best consumption of the network resources like bandwidth. This observed result is quite predictable since the RLS process uses the V2V communications (a RSU is considered as a vehicle also) then the request is sent through the whole network without any efficiency. Contrary to RLS, iRLS takes advantage of the RSU backbone which provide a fast connection between the initial RSU and the closest to the destination one.

V. Conclusion & Future Works

This paper presents the iRLS location service. This location service takes advantage of deployed RSUs on roads in order to ensure fast replies of destination lookup . RSUs are already wirelessly connected. Therefore, they can exchange data of the nodes position. This enhances the network routing performances. As a consequence, network performances such as the end-to-end delay are better than in the original RLS service. The presented work is realistic since it is applicable in any deployed C-ITS (cooperative Intelligent Transport System) and offers the ability to route packet from a node to another with interesting performances.

As future works, we intend to use the RSUs backbone to send real data. We intend also to study the scalability of such a mechanism. It could be interesting also to evaluate our protocol with another kind of protocols (Topology based protocol), with bandwidth metrics.

Acknowledgement

This work is partially supported by the EC SCOOP project (INEA/CEF/TRAN/A2014/1042281).

References

[1] R. Rana, S. Rana and K. C. Purohit, "A review of various routing protocols in VANET," International Journal of Computer Applications, vol. 96-18, 2014.
[2] K. Prasanth, D. K. Duraiswamy, K. Jayasudha and D. C. Chandrasekar, "Improved packet forwarding approach in Vehicular ad hoc networks using RDGR algorithm", arXiv preprint arXiv:1003.5437, 2010.
[3] C. J. Huang, Y. T. Chuang, . Y. J. Chen, D. X. Yang and I.F. Chen, "QOSaware roadside base station assisted routing in vehicular networks." Engineering Applications of Artificial Intelligence, 22(8), 1292-1301.
[4] G. J. Borsetti, "Infrastructure-assisted geo-routing for cooperative vehicular networks," IEEE Vehicular networking conference (VNC), 2010, pp. 255-262.
[5] Y. Peng, Z. Abichar and J.M. Chang, "Roadside-aided routing (RAR) in vehicular networks." IEEE International Conference on ICC'06.
[6] Käsemann, M., Füßler, H., Hartenstein, H., Mauve, M. (2002). *A reactive location service for mobile ad hoc networks*. Universität Mannheim/Institut für Informatik, Technical report REIHE INFORMATIK 14/2002. 14/2002
[7] Y.-B. Ko and N. H. Vaidya, "Location-aided routing (LAR) in mobile ad hoc networks," Wireless Networks, vol. 6, no. 4, pp. 307–321, Jul. 2000.
[8] B. Karp and H. T. Kung, "GPSR: Greedy Perimeter Stateless Routing for Wireless Networks," in Proceedings of the 6th annual international conference on Mobile computing and networking (MobiCom'00), New York, NY, USA, 2000, pp. 243-254.
[9] C. Lochert, H. Hartenstein, J. Tian, H. Fuessler, D. Hermann, and M. Mauve, "A routing strategy for vehicular ad hoc networks in city environments," in In Proceedings of the IEEE Intelligent Vehicles Symposium, 2003, pp. 156-161.
[10] K. Prasanth, D. K. Duraiswamy, K. Jayasudha, and D. C. Chandrasekar, "Improved packet forwarding approach in vehicular ad hoc networks using rdgr algorithm," arXiv preprint arXiv:1003.5437, 2010.
[11] NS2 : Network Simulator. [Online]. Available: http://nsnam.isi.edu/nsnam/
[12] OpenStreetMap. [Online]. Available: http://openstreetmap.fr/
[13] NS2 802.11p implementation. [Online]. Available: http://dsn.tm.uni-karlsruhe.de/english/Overhaul_NS-2.php/

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1093–1096
DOI: 10.15439/2016F567
ACSIS, Vol. 8. ISSN 2300-5963

Case-study of Localization via WSN Using Distributed Compressed Sensing

Veronika Olešnaníková, Michal Kochláň, *IEEE Student Member*, Róbert Žalman
University of Žilina
Faculty of Management Science and Informatics,
Univerzitná 8215/1 Žilina 010 26
Email: {Michal.Kochlan, Robert.Zalman, Veronika.Olesnanikova}@fri.uniza.sk

***Abstract*—Distributed compressed sensing task can be parallelized into several nodes that is highly suitable for using in Wireless Sensor Networks. Localization is one of the critical tasks solved in wireless systems. This paper investigates the possibilities of localization using compressed sensing implemented on wireless nodes and aggregation node. The presented case study simulates the application scenario of a target deployed in the field. This target is being localized by the wireless sensor network based on the emitted acoustic signal. Several types of the emitted signals have been used during the simulation runs. The emphasis was put on the properties of the reconstruction process such as compression ratio and minimization of the reconstruction error.**

I. Introduction

THE PROBLEM of target localization, in general, can be defined as finding an object in the space. Localization process in wireless sensor networks aims to localize a target based on the sensor data from the spatially distributed wireless nodes. For a single target and/or source localization in wireless sensor networks, there are various methods. For outdoor localization, this can include localization in traffic monitoring (vehicles, aircraft, bicycles, etc.). In indoor environment persons as well as animals and object can be tracked. Scientific literature refers to two core ways for estimation of the target location:

- Angle of Arrival (AOA) [1];
- Time Difference of Arrival (TDOA) [2-5].

In this paper, we are focused on the algorithms for single-source localization. The literature recognizes two basic groups of these algorithms:

- Energy Decay Model-based Localization Algorithms (EDMLA);
- Model Independent Localization Algorithms (MILA).

A. Decay Model-Based Method

The following formula shows the decay model which is in detail described in [6-8]. The received signal strength at i-th wireless node in time instant t can be expressed as follows:

$$y_i(t) = g_i \frac{E(t)}{d_{ik}^2(t)} + n_i(t), \tag{1}$$

where g_i means *gain factor* of i-th sensor. $E(t)$ is the *energy of the received signal in 1 meter distance from the wireless node*, and d_{ik} is the *Euclidean distance between i-th sensor and the target*. Moreover, n_i represents *measurement noise with zero mean value and Gaussian probability distribution* with variance σ_i^2, i.e.$N(0, \sigma_i^2)$.

B. Model-Independent Method

Authors in [9] describe a kernel averaging approach which does not need information about energy decay model. On the other hand in [10], the authors propose novel model-independent localization method and employed a distributed sorting algorithm. The research relies on the fact that the nodes closer to the target can measure higher RSSI (received signal strength indicator). Assuming that the wireless nodes know their rank, the distance estimates can be calculated from the respective probability density functions. RSSI indicator at i-th sensor in time instant t is as follows:

$$y_i(t) = g_i \sum_{k=1}^{K} \frac{E_k(t)}{d_{ik}^{\alpha}(t)} + \epsilon_i(t), \tag{2}$$

where $d_{ik}(t)$ is the *distance between the i-th sensor and k-th target*. K is the *number of targets*. g_i is *gain of i-th sensor*. $\epsilon_i(t)$ is random variable with mean value equal to μ_i and variance given as σ_i^2. $E_k(t)$ is the *energy of the received signal in 1 meter distance from k-th target*. α represents *attenuation exponent*.

The target localization task can be performed by the distributed nodes of a wireless network. An interesting approach arises from combination of compressed sensing and wireless sensor networks using distributed compressed sensing.

II. Distributed Compressed Sensing

A wireless sensor network (WSN) is formed by numerous spatially distributed devices (nodes) that process sensor data. Each node has a power source e.g. in form of battery thus having a limited lifetime. Since the energy consumption is a critical point, low power and efficient signal processing units are used in wireless nodes. Thus, wireless sensor nodes have limited computing and communication capabilities. Although, these nodes have low individual computing power, they can cooperate so that the computational power of the whole network allows performing advanced signal processing tasks [11]. Having the ability of advanced processing tasks leads to higher degree of robustness and greater versatility in low-lost scenario. This represents one of the most attractive reasons

978-83-60810-90-3/$25.00

why WSNs are used for wide range of remote sensing and environmental monitoring applications [12].

From the point of signal processing theory, a major challenge in WSNs is effective design of set of sensor-local signal processing operations and strategies suitable for inter-sensor communication and networking in order to address the desired trade-off among energy consumption, simple design, and overall system performance [13]. This trade-off shows, for example, in sensor lifetime maximization and effective battery utilization when reducing communication bandwidth. This can be achieved by each sensor by locally compressing the observed data and thus low rate inter-sensor communication is required. Such techniques can be represented by distributed compressed sensing (DCS).

Typical DCS scenario comprises numerous sensors measuring individually sparse signals, which are correlated among each other [14]. It should be noted that the signals are sparse in a certain basis. Each sensor individually encodes its signal by transforming it into another, incoherent basis (for example a random one). Then the sensor broadcasts only a few of the resulting coefficients to the aggregation node [14]. One of the advantages of DCS is that it does not require collaboration among the sensors when obtaining and processing the signal. Moreover, random projections in DCS are universal [15]. This means that any sparse basis can be used, which allows the same encoding strategy to be applied in different scenarios. This contributes to the robustness of the solutions based on DCS, i.e. the measurement stream from each sensor has equal priority. This is different from Fourier or wavelet transforms. It should be also mentioned that random measurements allow a progressively better recovery of the data, that means single measurement or more can be lost without the effect on the entire recovery process.

The problem of DCS illustrates the described example that follows. Let's have a network of n nodes, where each node has a piece of information given by x_j where $j = 1, \ldots, n$. Let's assume that each piece of information x_j is a scalar quantity. Together, the scalar quantities form a data vector $\boldsymbol{x} = [x_1, \ldots, x_n]^T$, which is called *networked data*. This underlines the fact that the data is distributed across the network and that the data may be shared over the network [17].

In wireless sensor networks, n can be a large number, thus having the networked data large as well. Therefore, the process of data acquisition at a single point is daunting. However, let's imagine that it is possible to create highly compressed version of vector $\boldsymbol{x}$ in a decentralized fashion. Scientific literature states several decentralized compressed sensing strategies. One strategy relies on the correlations among the a priori known data at different nodes [21]. In such case, a technique called *distributed source coding* known as Slepian-Wolf coding can be utilized as a compression scheme that allows none or little collaboration among the nodes. However, in lots of application scenarios the prior knowledge of the data correlations is not known. This situation supports research in collaborative signal processing within the sensor networks as well as data compression techniques [20]. It can be quite challenging to propose and implement an effective algorithm of distributed and collaborative processing for wireless sensor network [22]. Such algorithms rely to a great extend on specific prior knowledge and the relation of the expected signal correlations. The success of the implementation of such algorithms lies in sophisticated communication pattern and good processing capabilities of a sensor node.

The mentioned projections in standard multi-hop wireless networks utilizing compressed sensing can be expressed as vector $\boldsymbol{y}$, which components y_i can be calculated as follows:

$$y_i = \sum_{j=1}^{n} A_{i,j} x_j. \tag{3}$$

The components y_i are able to be computed in a decentralized fashion and efficiently because each value of the compressed data is represented as a simple linear combination of the values obtained at each node [23].

Basically, there are two simple steps in the computation and transmission of each compressed data sample y_i, where $i = 1, \ldots, k$ [23]:

1) Let's consider n sensor nodes in the wireless sensor network. Each of the sensors as its index $j = 1, \ldots, n$. Each node n_j computes locally properties $A_{i,j}$ and x_j so that the measured data are being multiplied with the corresponding element of the measurement matrix. The measurement matrix can be distributively created as local (at node) realization of $A_{i,j}$ using a pseudo-random number generator initialized by the node identifier, e.g. integers $j = 1, \ldots, n$. Having these node identifiers, particular node can simply calculate the vectors $\{A_{i,j}\}_{i=1}^{k}$, where sensors are indexed as $j = 1, \ldots, n$.
2) The local sensor node variables $A_{i,j}x_j$ are being continuously combined and transmitted over the sensor network using so called *randomized gossip*. The *randomized gossip* represent a decentralized algorithm, which computes linear functions such as:

$$y_i = \sum_{j=1}^{n} A_{i,j} x_j. \tag{4}$$

To summarize compressed sensing, one could say that it is a technique for signal processing and signal representation, where the signal can be sensed only by such number of samples, which corresponds to the signal sparsity in some base. The overall nature of compressed sensing matches the nature of event-driven control presented in the previous sections. Event-driven signal representation and reconstruction mechanism also leads to signal sampling and its reconstruction by as many (little) samples as are truly needed.

III. Case Study and Numerical Results

The idea of the localization system is to identify the position of the target by multiple WSN nodes deployed in the area. To decrease the consumption of the nodes the compressed sensing algorithms are supposed to be used. In order to use DCS the transmitted signal by the target has to follow special

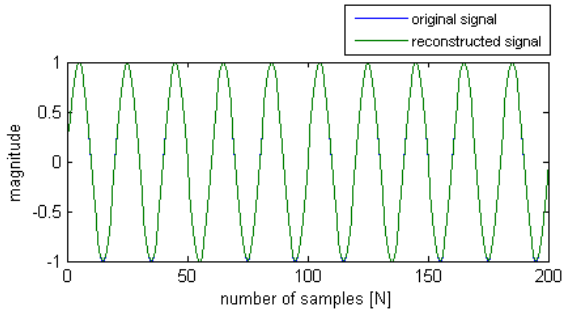


Figure 1. Continuous signal before and after reconstruction

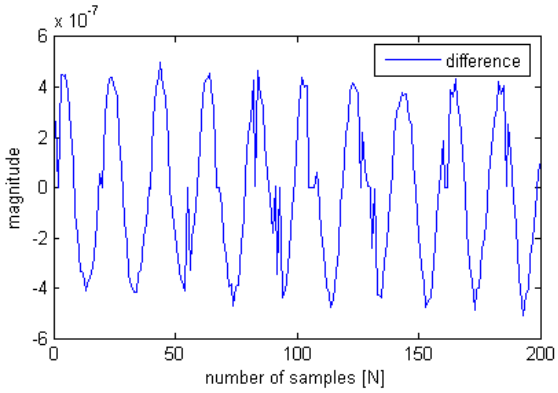


Figure 2. Difference of original and reconstructed continuous signal, cr=100

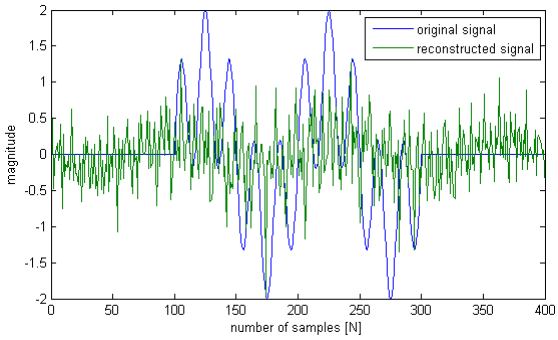


Figure 3. Burst signal (BURST01) before and after reconstruction

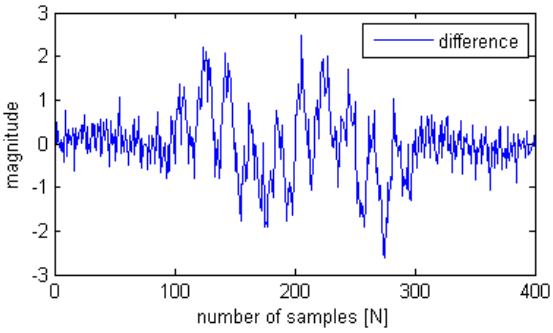


Figure 4. Difference of original and reconstructed burst signal, cr=100

requirements, e.g. certain sparsity in the frequency domain. This paper investigates signals in acoustic domain, i.e. in rage 20Hz up to 20kHz.

For the simulation purposes, the Matlab version 2013b was used. Reconstruction of original signal was performed by using the L1-magic library. The main goal of this article was to inspect the properties of different signals which could be used in the localization tasks. These signals should have suitable parameters for the reconstruction in order to use compressed sensing. All used signals are sparse in the frequency domain.

In the first simulation case, the signal was continuously transmitted. The signal has the sinusoid shape with the frequency of 100 Hz. The reconstruction using the distributed compressed sensing performs well with compression ratio (cr - see equation (5)) up to 200. The compression ratio is expressed by the following formula:

$$cr = \frac{s_a}{\chi}, \tag{5}$$

where cr is the compression ratio, s_a represents the number of all samples in the original signal and χ is the number of randomly selected compressed sensing coefficients. The following Fig. 1 shows the reconstruction of harmonic signal 100 Hz. Despite the fact that in the reconstruction of such a signal the compression ratio can be relatively high, this class of signals are not suitable for localization purposes. The reason is that we are not able to identify the same particular part of the received signal, e.g. the start of the same period.

The second simulation scenario is based on the transmitting the signal in bursts. By modification of the above-mentioned continuous signal, we are able to detect the start of the burst, thus, all nodes are able to determine the time of arrival of the received signal.

The Fig. 3 depicts the reconstruction of the burst signal with the same compression ratio as in the Fig. 1. The burst signal is compounded of the carrier signal and a secondary frequency. Carrier signal has a frequency of the 100Hz and secondary modulated signal has a frequency of 500Hz. The burst contains two periods of the carrier and ten periods of the secondary signal. Having this compression ratio the reconstruction results are insufficient for localization purposes.

Proper change of the parameters of the compression ratio leads to reconstruction improvement. This is demonstrated on the Fig. 5, along with difference between the original and the reconstructed signal on the Fig. 6. Compression ratio was decreased from 100 to 20.

Sumarization of the parameters used in the simulation is shown in Table I. It is obvious that continuous signals are very suitable for compressed sensing and can be reconstructed using high compression ratio, however, it is difficult to use them for localization purposes. The reconstruction error was calculated as a mean value based on differences between the original and reconstructed signals (see Fig. 2, 4, 6).

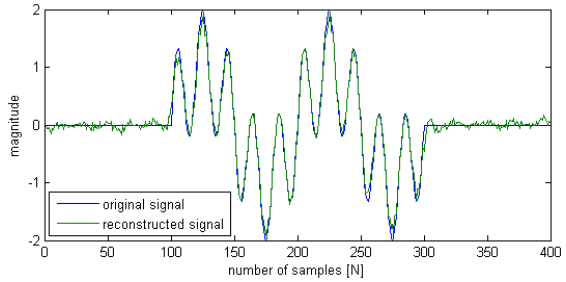


Figure 5. Burst (BURST02) signal before and after reconstruction

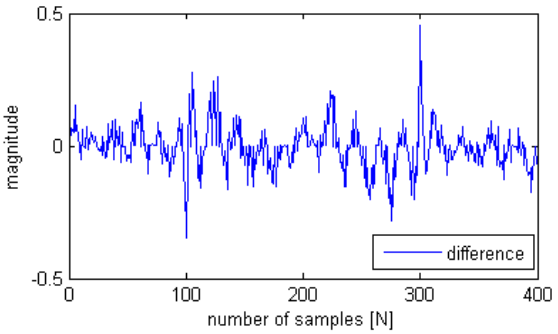


Figure 6. Difference of the original and reconstructed burst signal, cr = 20

Table I
PARAMETERS AND THE RESULTS OF THE SIMULATIONS

Type of the signal	Frequency	Number of periods	cr	Error
CONTINUOUS	100Hz	10	100	$9.52 \cdot 10^{-8}$
BURST01	100Hz	10	100	0.5638
BURST02	100Hz	10	20	0.0536

IV. CONCLUSION

This contribution investigates suitable signals for target localization using WSN and DCS. The results show, that continuous periodic signals can be reconstructed easily with high compression ratio using DCS. However, these signals are not target-identifiable. Changing the form of the transmitted signal broadcast from the target enables the sensor network to detect it. The changed signal has the form of burst, which enables to detect the start of the beacon signal. Based on the time of the arrival and the ability to detect the start of the beacon signal, the network is able to localize the target.

REFERENCES

[1] L. M. Kaplan, Q. Le, and P. Molnar, "Maximum likelihood methods for bearings-only target localization", in Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 5, pp. 3001 – 3004, Salt Lake City, Utah, USA, May 2001.

[2] Y. Weng, W. Xiao, and L. Xie, "Total least squares method for robust source localization in sensor networks using TDOA measurements", International Journal of Distributed Sensor Networks, vol. 2011, Article ID 172902, 8 pages, 2011.

[3] X. Qu and L. Xie, "Source localization by TDOA with random sensor position errors-part I: static sensors", in Proceedings of the 15th International Conference on Information Fusion, pp. 48-53, Singapore, July 2012.

[4] X. Qu and L. Xie, "Source localization by TDOA with random sensor position errors-part II: mobile sensors", in Proceedings of the 15th International Conference on Information Fusion, pp. 54-59, Singapore, July 2012.

[5] K. C. Ho, "Bias reduction for an explicit solution of source localization using TDOA", IEEE Transactions on Signal Processing, vol. 60, no. 5, pp. 2101-2114, 2012.

[6] D. Blatt and A. O. Hero, "Energy-based sensor network source localization via projection onto convex sets", IEEE Transactions on Signal Processing, vol. 54, no. 9, pp. 3614-3619, 2006.

[7] K. Deng and Z. Liu, "Weighted least-squares solutions of energy-based collaborative source localization using acoustic array", International Journal of Computer Science and Network Security, vol. 7, no. 1, pp. 159-165, 2007.

[8] Q. Shi and C. He, "A new incremental optimization algorithm for ML-based source localization in sensor networks", IEEE Signal Processing Letters, vol. 15, pp. 45-48, 2008.

[9] M. G. Rabbat, R. D. Nowak, and J. Bucklew, "Robust decentralized source localization via averaging", in Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '05), vol. 5, pp. V1057-V1060, Philadelphia, Pa, USA, March 2005.

[10] D. Ampeliotis and K. Berberidis, "Energy-based modelindependent source localization in wireless sensor networks," in Proceedings of the 16th European Signal Processing Conference, Lausanne, Switzerland, August 2008.

[11] I.F. Akyildiz and W. Su and Y. Sankarasubramaniam and E. Cayirc, "Wireless sensor networks: a survey", in Computer Networks, 2002, ISSN: 1389-1286

[12] A. Mainwaring and D. Culler and J. Polastre and R. Szewczyk and J. Anderson, "Wireless sensor networks for habitat monitoring", in Proceedings of the 1st ACM international workshop on Wireless sensor networks and applications,2002

[13] M. A. Razzaque and Ch. Bleakley and S. Dobson, "Compression in wireless sensor networks: A survey and comparative evaluation", in ACM Transactions on Sensor Networks (TOSN), 2013

[14] C. Caione and D. Brunelli and L. Benini, "Distributed compressive sampling for lifetime optimization in dense wireless sensor networks", in Industrial Informatics, IEEE Transactions on, 2012

[15] K. Hayashi and M. Nagahara and T. Tanaka, "A user's guide to compressed sensing for communications systems", in IEICE transactions on communications, 2013

[16] S. Foucart and H. Rauhut, "An Invitation to Compressive Sensing", 2013

[17] Y. C. Eldar and G. Kutyniok, "Compressed sensing: theory and applications", 2012

[18] M. Fornasier and H. Rauhut, "Handbook of mathematical methods in imaging, Compressive sensing", p. 187 - 228, 2011, ISBN 978-0-387-92920-0

[19] M. Fornasier and H. Rauhut, "Compressive sensing", 2011, ISBN: 9780387929200

[20] M. Elad, "Sparse and redundant representations: from theory to applications in signal and image processing", 2010, ISBN: 9781441970107

[21] A. C. Fannjiang and T. Strohmer and P. Yan, "Compressed remote sensing of sparse objects",2010

[22] E. J. Candes and M. B. Wakin, "An Introduction to Compressive Sampling" ISSN: 1053-5888

[23] A. Cohen and W. Dahmen and R. DeVore, "Compressed Sensing and Best k-term Approximation", 2009

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1097–1102
DOI: 10.15439/2016F386
ACSIS, Vol. 8. ISSN 2300-5963

Uniform Inbuilt Wireless Sensor Node for Working Conditions Monitoring

Denis Spirjakin, Alexander M. Baranov
Moscow Aviation Institute
(National Research University),
Moscow, Russia
Email: denis.spirjakin@gmail.com

***Abstract*—All companies strive to eliminate work accidents. But still a lot of professions are exposed to different hazardous conditions and many work injuries and even deaths occur everyday. Monitoring of working conditions is a very important part of labor protection. Combining wireless sensor networks with wearable technology is possible to significantly improve safety delivery capability of such systems and add new functionality to them. In this work we present the design results of the uniforms inbuilt wireless sensor node for working conditions monitoring. The node is able to signalize about employee presence in relation to working facilities, and to monitor the atmosphere for temperature and combustible gases concentration. It consists of light-weighted distributed pieces which are built in to clothes and has low power consumption. The average power consumption of the node is low enough for several weeks autonomous lifetime.**

I. Introduction

While safety standards for industrial plants become tougher, the demand for continuous monitoring of employee state and working conditions including environmental conditions is rising. At the same time modern wearable technology tends to build electronics in to clothes [1] providing new quality in healthcare [2], [3], military [4] and other spheres.

Variety of wireless sensor networks were developed recently. These networks consist of small nodes and are equiped with transceivers, microprocessors and sensors [5], [6]. They can be used in different areas of life (security, military, home automation, etc.). But the most frequent area is environmental [7] and human [8] monitoring.

Incorporating wireless sensor networks with wearable technology is possible to significantly improve safety delivery capability of monitoring systems and add new functionality to them. Such “smart” uniforms can provide in-situ monitoring of people, facilities and environmental conditions and in case of an emergency situation assist special services and employees.

It is expected that in the nearest future wireless sensor networks will connect computer networks and physical world. That will lead to tight integration of real and virtual worlds [9] where communication will go between people and devices – Internet of Things [10].

In this work we present the design results of the uniform inbuilt wireless sensor node for working conditions monitoring. The node is able to control employee presence in relation to working facilities, and to monitor the atmosphere for temperature and combustible gases concentration. It consists of light-weighted distributed pieces which are built in to clothes and has low power consumption. The average power consumption of the node is low enough for several weeks autonomous lifetime.

The paper is organized as follows: at first we overview the node in Section II. In Section III we describe the sensing circuit of the node. Section IV is dedicated to data transmission and presence detection principles which are used in this work. The power consumption of the node is discussed in Section V. Finally, we provide concluding remarks in Section VI.

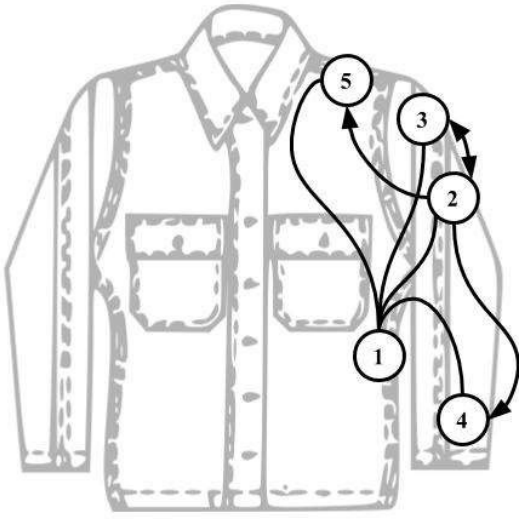


Fig. 1. Block diagram of the node: a power source (1), a sensor module (2), a trasceiver (3), a light indicator (4) and a buzzer (5).

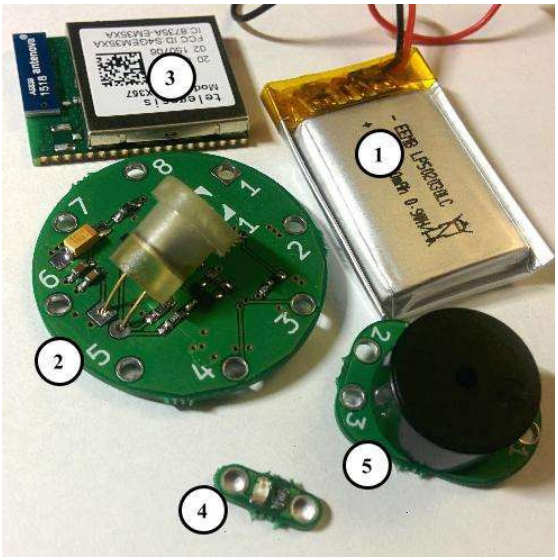


Fig. 2. The node pieces: a power source (1), a sensor module (2), a trasceiver (3), a light indicator (4) and a buzzer (5).

Fig. 3. The assembled node: the light indicator (1), the sensor module (2) and the buzzer (3).

II. Node Overview

The diagram of the uniform inbuilt wireless sensor network node is presented in Fig. 1. The node consists of several distributed pieces wich are built in to a jacket: a power source (1), a sensor module (2), a trasceiver (3), a light indicator (4) and a buzzer (5). The node pieces are presented in Fig. 2.

Since the node is wearable and built in to clothes, all its pieces are light and small. At the same time, the pieces are placed according to their functions and distributed around the clothes to make wearing comfortable.

The power source of the node is a lithium-polymer battery. Its nominal voltage is 3.6 V and the capacity is 250 mAh. The weight of the battery is about 5 g and the dimensions are 30×20×5 mm. The battery is placed in the bottom of the jacket and connected to a ribbon cable to provide power to other pieces.

The sensor module provides sensing of the atmosphere temperature and combustible gases concentration. It also controls the light indicator and the buzzer to maintain alarm signals and send and receive data using the transceiver.

This module is based on ATxmega16E5 microcontroller. It uses built-in temperature sensor to measure the atmosphere temperature and the commercial catalytic gas sensor for combustible gases concentration measuring. The gas sensor is manufactured by NTC IGD (Russia) and its power consumption is 200 mW in continuous measurement mode. Therefore, the catalytic gas sensor is very power hungry. To decrease its power consumption, the special measuring algorithm is used. More details about the measuring circuit and the algorithm are presented in Section III.

The sensor module is placed on the left shoulder and connected to the power supply ribbon cable and cables to the light indicator, the buzzer and the transceiver. The light indicator consists of light emitting diode and current limiting resistor. It placed on the left arm to be on the line of sight of the wearer. The buzzer is placed on the left shoulder near the neck. The task of light indicator and the buzzer is to perform visual and audial alarm signals to wearer when dangerous condition occurs.

Except visual and audial signals, information about dangerous conditions is transmitted to other devices in a wireless sensor network. For this purpose the transceiver module Telegesis ETRX3 is used. The module provides IEEE 802.15.4/Zigbee standards compatible protocol and is controlled by UART interface using AT-style commands set.

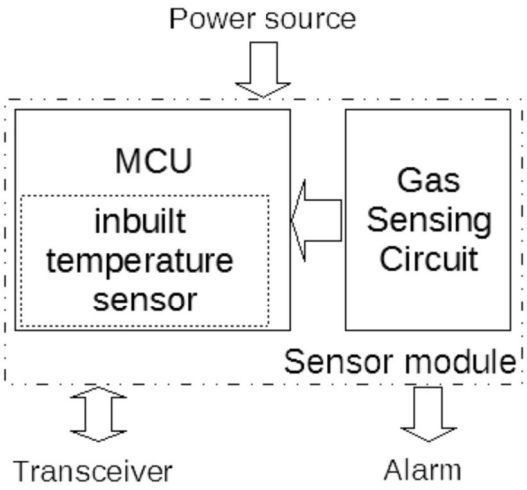


Fig. 4. Block diagram of the sensor module.

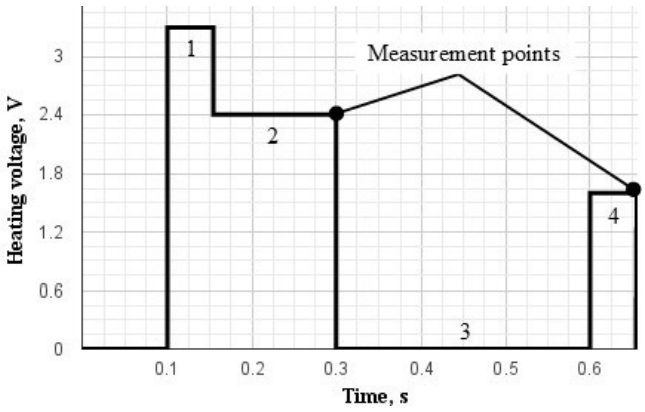


Fig. 5. Gas sensor heating profile.

The transceiver is also used to perform employee's presence tracking. The algorithm to detect wearer presence is described in Section IV.

The node which is assembled on the jacket is presented in Fig. 3.

III. Sensor Module

Sensor module consists of the MCU with inbuilt temperature sensor and the gas sensing circuit. To measure gas concentration the catalytic sensor is used. The block diagram of the sensor module is presented in Fig. 4.

Catalytic gas sensor circuits are usually based on the Wheatstone bridge circuit. That circuit consists of two resistors and two sensors, one active and one reference. Measurement process includes heating of the sensors (up to 450C for methane) which is the main part of sensor's power consumption. The power consumption value of the Wheatstone bridge circuit is about 200 mW. That's too much and not appropriate for autonomous operation of the node.

Excluding the reference sensor decreases the power consumption. The compensation of the atmosphere humidity and temperature which is usually performed by the reference sensor in this case is fulfilled by using the specific multistage heating pulse. This method was offered and discussed in [11]-[13].

In this work we use four stages for every multistage pulse (1-4 regions in the diagram, Fig. 5).

In the first and the second stages the sensor is heated up to the catalysis external diffusion region (about 450C) and the partial evaporation of surface water is performed. During the third stage heating is not applied. In the final fourth stage the sensor is heated up to the beginning of the catalysis kinetic region (about 200C). After this pulse, the element cools down to ambient temperature. Heating voltages for the stages are 3.3 V, 2.4 V, 0 V and 1.6 V respectively.

The measurement result is the difference between sensor voltages at two different temperatures (measurement points in the diagram).

The gas sensor measuring circuit for the multistage pulse method is presented in Fig. 6. The circuit is controlled by the microcontroller and consists of a sensor R5, a MOSFET key VT1 which controls heating, two range switching MOSFET keys VT2 and VT3 which scale up output signal based on sensor resistance, and an output low pass filter R7-C1 which is connected to microcontroller's inbuilt ADC.

Different voltage levels of the multistage heating pulse are generated by applying pulse-width modulated signal to the heating key. As it was shown in [14] PWM heating allows to significantly decrease sensor power consumption and therefore use a catalytic gas sensor in battery powered autonomous devices.

Range switching keys maintain the scale of circuit output signal inside the input range of ADC converter. Two ranges are used since two measurements per cycle are performed.

To match safety standards [15], [16] the measuring pulses are performed once per 20 seconds. Between pulses to spare the power the node is switched to sleep mode.

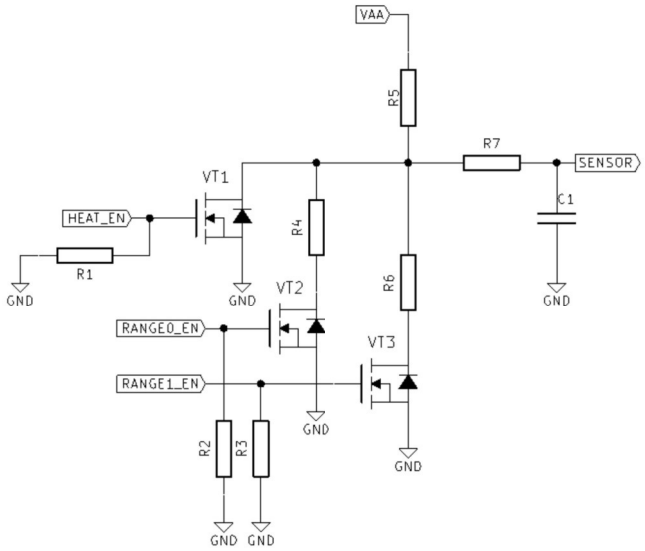


Fig. 6. The gas sensor measuring circuit for the multistage pulse method with PWM heating.

Table I. Reliable service area versus transceiver output power.

Output power, dBm	Distance, m
-43	0.4
-26	3
-20	5
-17	7
-14	10

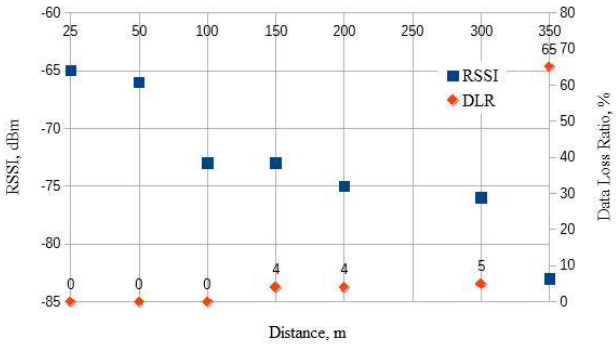


Fig. 7. Relative signal strength versus transmission distance and related data loss ratio.

The node is configured to activate alarm based on two threshold values of methane concentration which are 0.5 and 1 % vol. If the concentration is less than first threshold, there is no relation from the node. If the concentration is going higher than 0.5 % vol. methane, the node provides the light and short sound alarm. When the concentration is more than 1 % vol. the sound alarm becomes longer. In both cases the alarm signal is also sent over the wireless network to the data sink node.

IV. Data Transmission and Presence Detection

Data transmission is performed by using the transceiver module Telegesis ETRX3. The module is IEEE 802.15.4/Zigbee compatible. The communication with module is performed by UART interface using AT-style command set.

To communicate by wireless network the node should join it before. After it is joined, it has parent node where the data will be sent to later routing. If the connection with parent node is lost, the node should rejoin the network to be able to send any messages.

The transmission distance depends on the signal power. The transceiver module is able to change its output power. Therefore, the signal transmission distance can be regulated. This feature is used to implement the node presence detection.

Since the node is connected to the network its presence is already detected by parent node. If the connection is lost the node is trying to rejoin the network every work cycle which is 20 seconds. Attempts are continued till the node is joined the network and at that moment the presence will be detected by new parent node.

The detection radius is based on the transmission power of the node. Reliable service distances versus transceiver output power are shown in Table I. The service area is considered as

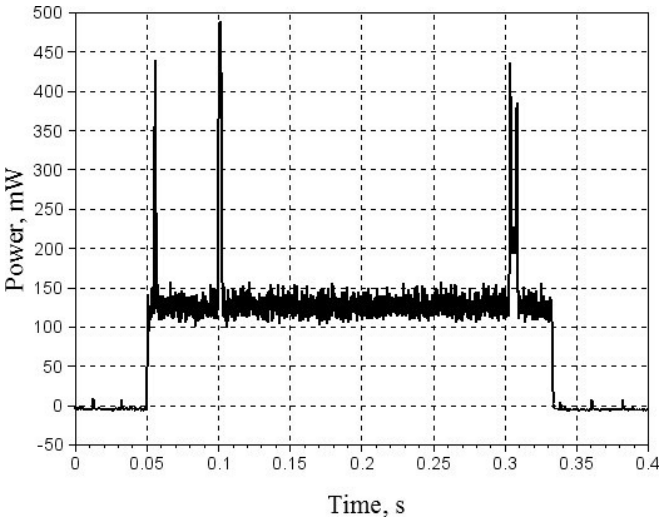


Fig. 9. The node power consumption during data transmission.

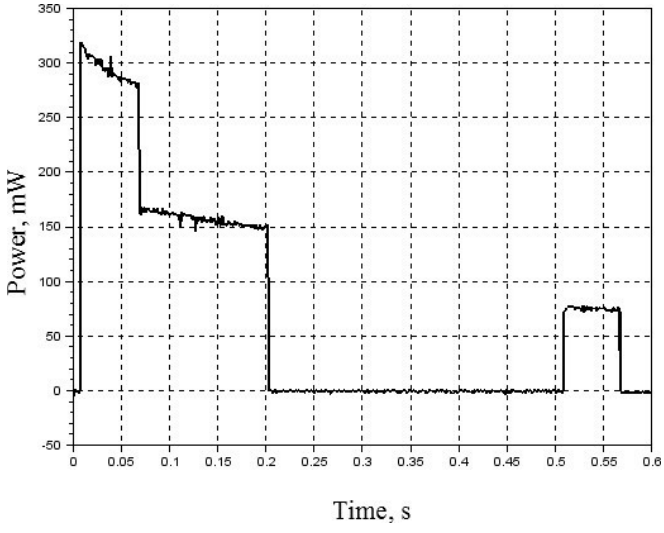


Fig. 8. The node power consumption during measurements.

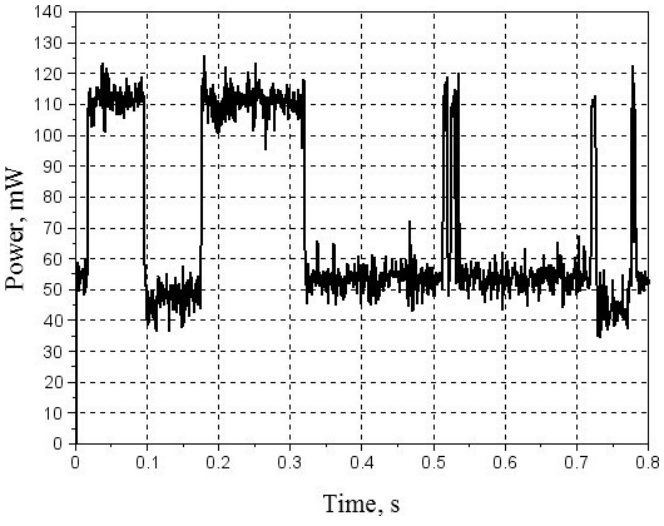


Fig. 10. The node power consumption during joining the network.

reliable if the RSSI value is more than -70 dBm which corresponds to zero data loss ratio (Fig. 7). In this work, for more precise detection, the transceiver output power is reduced to -26 dBm.

At the same time, when the node is going to alarm state it's necessary to ensure that the message have the best chances to be delivered. Relative signal strength versus transmission distance and related data loss ratio are presented in Fig. 7. The transceiver output power in that case is +8 dBm. As it's shown in the figure the reliable service area of the transceiver is 100 meters since in this distance there are no data losses. At this distance the RSSI value is -70 dBm.

Therefore, during alarm data transmission the output power of the transceiver is setted to +8 dBm which is the maximum output power of the transceiver in its boost mode.

V. Power Consumption

As it was said before, the work cycle of the node is 20 seconds. During the cycle the node goes through following stages: the multistage measurement pulse, the data transmission and presence detection and the power saving mode.

The power consumption of the node during the multistage pulse is presented in Fig. 8.

As it's shown in the figure the power consumption value during first stage of the multistage pulse is 320-280 mW. For second stage the value is 170-150 mW. And during fourth stage it's less than 80 mW. The average power consumption for whole multistage pulse is 77 mW and it's legth is about 0.6 s.

The multistage pulse is followed by data transmission to assure that the wireless connection is established in common situation or to deliver data during alarm mode. The average power consumption in this stage is about 125 mW for alarm mode (when transceiver output power is +8 dBm) and about 70 mW otherwise. The node power consumption during data transmission is shown in Fig. 9.

The data transmission takes no more than 0.5 s. But if the wireless connection isn't established it takes about 1s to rejoin the network. The node power consumption during joining the network is presented in Fig. 10. The average power consumption of the node in this mode is about 70.5 mW.

Since the cycle length is 20 s, the average power consumption for whole cycle is about 8.56 mW in worst case of an alarm situation while the connection to the network was lost. And the average power consumptions is about 2.66 mW in general situation when the connection is established.

Providing 900 mWh power supply the autonomous lifetime of the node will be about 105 hours in worst case and about 338 hours in general situation. Assuming that a work day has 8 hours, that's more than 13 work days long autonomous lifetime. Therefore, the average power consumption of the node is low enough for more than two weeks autonomous lifetime.

VI. Conclusion

In this work the uniform inbuilt wireless sensor node for working conditions monitoring is presented. The node consists of the sensor module (with the catalytic methane sensor, the temperature sensor and the MCU), the ZigBee transceiver, the light indicator and the buzzer. All these pieces are low power, light-weighted and built in to the clothes. The node is able to signalize about employee presence in relation to working facilities, and to monitor the atmosphere for temperature and combustible gases concentration. Due to the average power consumption optimization, the autonomous lifetime of the node is more than two weeks. The reliable service area of data transmission is 100 meters.

References

[1] Stoppa Matteo, Alessandro Chiolerio. "Wearable electronics and smart textiles: a critical review." Sensors 14.7 (2014): 11957-11992. http://dx.doi.org/10.3390/s140711957

[2] Chan Marie, et al. "Smart wearable systems: Current status and future challenges." Artificial intelligence in medicine 56.3 (2012): 137-156. http://dx.doi.org/10.1016/j.artmed.2012.09.003

[3] Sultan Nabil. "Reflective thoughts on the potential and challenges of wearable technology for healthcare provision and medical education." International Journal of Information Management 35.5 (2015): 521-526. http://dx.doi.org/10.1016/j.ijinfomgt.2015.04.010

[4] Scataglini, Sofia, Giuseppe Andreoni, and Johan Gallant. "A Review of Smart Clothing in Military." Proceedings of the 2015 workshop on Wearable Systems and Applications. ACM, 2015. http://dx.doi.org/10.1145/2753509.2753520

[5] Pei Zhou, Gongsheng Huang, Linfeng Zhang, Kim-Fung Tsang, Wireless sensor network based monitoring system for a large-scale indoor space: data process and supply air allocation optimization, Energy and Buildings, Volume 103, 15 September 2015, Pages 365-374. http://dx.doi.org/10.1016/j.enbuild.2015.06.042

[6] Hang Shen, Guangwei Bai, Routing in wireless multimedia sensor networks: A survey and challenges ahead. Review Article, Journal of Network and Computer Applications, Volume 71, August 2016, Pages 30-49. http://dx.doi.org/10.1016/j.jnca.2016.05.013

[7] A. Somov, A. Baranov, D. Spirjakin, A. Spirjakin, V. Sleptsov, R. Passerone, Deployment and evaluation of a wireless sensor network for methane leak detection, J. Sensors and Actuators A. 202 (2013) 217-225. http://dx.doi.org/10.1016/j.sna.2012.11.047

[8] Alessandro Redondi, Marco Chirico, Luca Borsani, Matteo Cesana, Marco Tagliasacchi, An integrated system based on wireless sensor networks for patient monitoring, localization and tracking, Ad Hoc Networks, Volume 11, Issue 1, January 2013, Pages 39-53. http://dx.doi.org/10.1016/j.adhoc.2012.04.006

[9] Dimitris Kelaidonis, Andrey Somov, Vassilis Foteinos, George Poulios, Vera Stavroulaki, Panagiotis Vlacheas, Panagiotis Demestichas, Alexander Baranov, Abdur Rahim Biswas, Raffaele Giaffreda, Virtualization and Cognitive Management of Real World Objects in the Internet of Things. In: IEEE International Conference on Green Computing and Communications (GreenCom), pp.187-194, IEEE Press (2012). http://dx.doi.org/10.1109/GreenCom.2012.37

[10] Miorandi, D., Sicari, S., De Pellegrini, F., Chlamtac, I.: Internet of Things: Vision, Applications and Research Challenges. J. Ad Hoc Networks 10, 1497-1516 (2012). http://dx.doi.org/10.1016/j.adhoc.2012.02.016

[11] A.Somov, A.Baranov, D.Spirjakin, R. Passerone, Circuit design and power consumption analysis of wireless gas sensor nodes: One-sensor versus two-sensor approach, IEEE Sensors Journal, 14 (6) (2014) 2056- 2063. http://dx.doi.org/10.1109/JSEN.2014.2309001

[12] Denis Spirjakin, Alexander M. Baranov, Vladimir Sleptsov. "Design of smart dust sensor node for combustible gas leakage monitoring." In Computer Science and Information Systems (FedCSIS), 2015 Federated Conference on (pp. 1279-1283). IEEE. http://dx.doi.org/10.15439/2015F172

[13] Alexander Baranov, Denis Spirjakin, Saba Akbari, Andrey Somov, "Optimization of power consumption for gas sensor nodes: A survey." Sensors and Actuators A 233 (2015) 279–289. http://dx.doi.org/10.1016/j.sna.2015.07.016

[14] Spirjakin D., Baranov A. M., Somov A., & Sleptsov, V. "Investigation of Heating Profiles and Optimization of Power Consumption of Gas Sensors for Wireless Sensor Networks." Sensors and Actuators A: Physical (2016). http://dx.doi.org/10.1016/j.sna.2016.05.049

[15] Standard GOST R EN 50194-1-2012, Signalizators for the detection of combustible gases in domestic premises, 2000.

[16] Standard EN 50194-2000, Electrical apparatus for the detection of combustible gases in domestic premises. Test methods and performance requirements, 2000.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1103–1106
DOI: 10.15439/2016F592
ACSIS, Vol. 8. ISSN 2300-5963

Using wireless acceleration sensor for system identification

Peter Šarafín
University of Žilina
Faculty of Management Science and Informatics,
Univerzitná 8215/1, Žilina 010 26
Email: peter.sarafin@fri.uniza.sk

Juraj Miček, Jana Milanová
University of Žilina
Faculty of Management Science and Informatics,
Univerzitná 8215/1, Žilina 010 26
Email: {juraj.micek, jana.milanova}@fri.uniza.sk

***Abstract*—In practical applications, we often encounter problems controlling weakly damped resonant systems. These are devices which often include inertia masses and flexible connecting elements. These devices are mainly gantry cranes, mechatronic systems, elevators, filling lines for the food industry and many others. One approach to improving the transition process in the control of these weakly damped systems is a method of shaping control signals. This method starts to be used in the control of systems with flexible elements in the 90s of the twentieth century. Over the next twenty years, we meet with successful applications, especially in the control of positioning systems. When we are talking about the theoretical description of input shaping today, we meet mainly with two basic approaches. The first is based on the selection of a proper sequence of pulses in the time domain. The second is based on the design of such discrete shaper, which compensates the effect of the complex poles of a controlled system causing residual vibration. Irrespective of the shaper design method, we must know either the systems oscillations and controlled system damping, or the location of the complex poles causing the vibration.**

I. Introduction

THE AIM of the input shaper is to adjust the control signals of the weakly damped system in order to eliminate residual vibrations of the system. With its proposal, we try to minimize transition time simultaneously. To illustrate this, let us have oscillatory system with the transfer function:

$$F(s) = \frac{1}{T^2 s^2 + 2bTs + 1}, b \in < 0, 1), \tag{1}$$

where b is the damping of the system and T is the time constant.

The task is to propose such methods of the control signal adjustment, that the transition process fits stated requirements. The possible system arrangement is shown in figure 1. It is evident that the shaper acts as a serial correction element that is known from the classical theory of automated control.

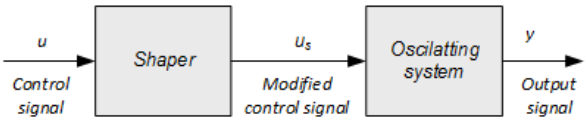


Fig. 1. The input shaping process

For the shaping of the control signal, two approaches can be used:

1) Signal shaping in the time domain, e.g. direct generation of the control signals with suitable properties [1], [2].
2) Using discrete shaping members [6], [9].

Appropriately chosen shaper suppress residual vibrations of the system. Note that the suppression ratio of residual vibrations is often expressed as the ratio of the amplitude of the output signal with the shapers input signal to the amplitude of the output signal without the shaper [4], [7], [8].

II. Proposal of an input shaper in the time domain

The rate of suppression of residual vibrations can be expressed as a function of the angular velocity ω and the proportional damping factor b of the system as:

$$V(\omega, b) = e^{-b\omega t_n} \sqrt{(C(\omega, b))^2 + (S(\omega, b))^2}, \tag{2}$$

while:

$$\begin{aligned} C(\omega, b) &= \sum_{i=1}^{n} A_i \cdot e^{b\omega t_i} \cdot cos(\omega \cdot t_i \sqrt{1 - b^2}), \\ S(\omega, b) &= \sum_{i=1}^{n} A_i \cdot e^{b\omega t_i} \cdot sin(\omega \cdot t_i \sqrt{1 - b^2}), \end{aligned} \tag{3}$$

where A_i is the amplitude, t_i is the time of occurrence of the ith pulse and n is the number of input shapers pulses. Ideally, with complete suppression of the vibrations, the relation (2) has to be equal to zero for its own circular speed and relative system damping. It is obvious that for $\omega = \omega_0$ is given:

$$C(\omega_0) = 0, S(\omega_0) = 0, \tag{4}$$

If the shaped output is to have the same final value as the unshaped (normalized shaper), then it is given that the sum of the amplitudes of all pulses has to be equal to one:

$$\sum_{i=1}^{n} A_i = 1. \tag{5}$$

Note that if the pulse amplitudes A_i are not limited, then in terms of minimizing the total time of transition t_n, it will acquire an infinitely large value. When practical solutions, we most often encounter with two restrictive conditions:

978-83-60810-90-3/$25.00

3) The pulse amplitudes can take values from the range of $\pm A_{max}$,
4) The pulse amplitudes can take only non-negative values $A_i \geq 0$.

It is obvious that if amplitudes of all the pulses are non-negative, then respecting conditions (5), it must belong to the interval $0 \leq Ai \leq 1$. With respect to those limitations, by solving the equations (2) we get for $min(t_n)$ the solution describing positive ZV shaper whose design is often stated in matrix form [3], [5], [6], [7]:

$$\begin{bmatrix} A_i \\ t_i \end{bmatrix} = \begin{bmatrix} \frac{1}{1+K} & \frac{K}{1+K} \\ 0 & 0.5 \cdot T_D \end{bmatrix}, \tag{6}$$

where:

$$T_D = \frac{2 \cdot \pi \cdot T}{\sqrt{1-b^2}}, K = e^{-\frac{b\pi}{\sqrt{1-b^2}}}. \tag{7}$$

ZV shaper (zero-vibration) (6) may be described in the time domain by a relation:

$$y(t) = A_1 \cdot \delta(t) + A_2 \cdot \delta(t - t_2). \tag{8}$$

ZV shaper described, however, is quite sensitive to changes in the parameters of the controlled system, in particular, to the changes in its own circular speed. For this reason, more robust input shapers have been developed. The most famous is ZVD (zero vibration derivate) shaper. In deriving the ZVD shaper, it is assumed that the derivative of the function $V(\omega, b)$ is equal to zero.

$$\begin{bmatrix} A_i \\ t_i \end{bmatrix} = \begin{bmatrix} \frac{1}{(1+K)^2} & \frac{2K}{(1+K)^2} & \frac{K^2}{(1+K)^2} \\ 0 & 0.5 \cdot T_D & T_D \end{bmatrix} \tag{9}$$

From (9) it is clear that the increase of robustness is penalized by the increase of the transition time. Transition time rose from $0.5 \cdot T_D$ to T_D.

III. Proposal of discrete input shaper

The problem of suppression of residual vibrations in the control of weakly damped systems can be successfully resolved with the appropriate design of discrete systems (correction elements) regulating the spectrum of the control signal to suppress residual vibration of the system. It is clear that the task can be solved by the appropriate placing the zeros of the z-transfer function characteristic element to the z-plane points corresponding to the field effect system. As the controlled system is characterized by a continuous transfer function, it is necessary to find a suitable transformation of the continuous system to a discrete equivalent. The discrete shaper will then be the inverse of the discrete equivalent of the continuous system. If the designed shaper compensates only selected complex transfer poles, then it is sufficient to place zeroes to the appropriate poles of the shaper. To find the discrete equivalent of the continuous system $F(s)$ defined by poles which cause oscillation of the system, it is suitable to use the transform between s and z plane by using the relation:

$$z = e^{sT_v}, \tag{10}$$

where T_v is the sampling period.

Recall that we want to compensate only complex poles, therefore we consider the transfer in the form:

$$F(s) = \frac{1}{\prod_{i=1}^{N}(T_i^2 s^2 + 2b_i T_i s + 1)}, \tag{11}$$

where N represents the number of pairs of complex conjugate poles to be compensated, T_i stands for the time constant and b_i is the damping coefficient of the ith subsystem. The poles of the system with transfer (11) are:

$$p_{i,1,2} = -\frac{b_i}{T_i} \pm j\frac{\sqrt{1-b^2}}{T}, i = 1, 2, 3, ..., N \tag{12}$$

In applying the transformation equation (10), for the poles of discrete equivalent of continuous system (11) will apply:

$$p_{di,1,2} = r_i e^{\pm j\phi_i} = e^{-\frac{T_v b_i}{T_i}} e^{\pm j\frac{T_v\sqrt{1-b_i^2}}{T_i}}, i = 1, 2, ..., N \tag{13}$$

To compensate for the effect of selected poles of the continuous system, the discrete shaper has to contain zeroes in such points in the z-plane that correspond to poles of the discrete equivalent ($p_{i1,2}$) (Fig. 2).

If we place the zeroes of the shapers z-transfer function into the points corresponding to the position of poles, then transfer function shaper will be in the form:

$$F(z) = \prod_{i=1}^{N}(z - z_{i1})(z - z_{i2}), \tag{14}$$

alternatively, when considering a 2N-multiple pole at the origin of the plane:

$$\begin{aligned} F(z) &= \textstyle\prod_{i=1}^{N} C_i(1 - z_{i1}z^{-1})(1 - z_{i2}z^{-1}), \\ F(z) &= \textstyle\prod_{i=1}^{N} C_i(a_{i2}z^{-2} + a_{i1}z^{-1} + a_{i0}), \end{aligned} \tag{15}$$

where

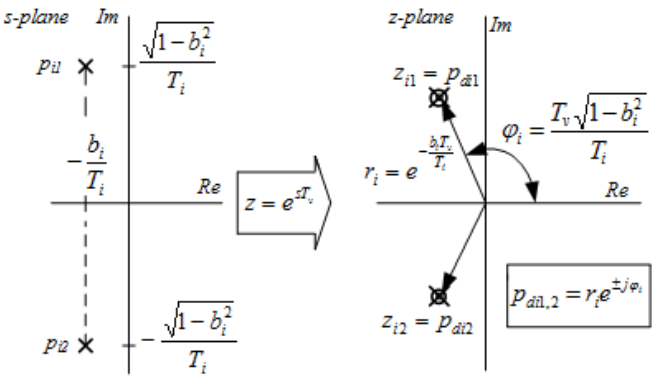


Fig. 2. Poles representation in the s-plane and z-plane

$$\begin{aligned} C_i &= \prod_{i=1}^{N} \frac{1}{a_{i0}+a_{i1}+a_{i2}}, \\ a_{i0} &= 1, a_{i1} = -(z_1+z_2), a_{i2} = z_1 \cdot z_2, \end{aligned} \tag{16}$$

alternatively:

$$\begin{aligned} a_{i0} = 1, a_{i1} = -2r_i cos(\phi i), a_{i2} = r_i^2, \\ r_i = e^{-\frac{T_{vi} b_i}{T_i}}, \phi_i = \frac{T_{vi}\sqrt{1-b_i^2}}{T_i}. \end{aligned} \tag{17}$$

When defining the gain C_i, we assume that consistentely with the continuous system transfer(11) is the shaper normalized (shaper gain is equal to one). From (13) it is clear that the position of poles of the discrete equivalent of the continuous system (11) $p_{di,1,2}$ can be changed in the z-plane by selecting the sampling period T_v. This is illustrated in figure 3.

The curves on the right side of figure 3 illustrate a possible $p_{d1,2}$ poles location of the discrete equivalent of a continuous system, depending on the selection of sampling period T_v. The left side of figure 3 shows the poles of the original continuous oscillating system with the transfer function (1). It is clear that the value T_v can be theoretically chosen in the range of 0 to ∞. Recall that the settling time of system output is proportional to the sampling period T_v. It follows that if we focus on achieving good system dynamics shaper-system, we are trying to make T_v minimal during the design of the shaper. Obviously, the choice of T_v minimum value is related to the energy options of the actuator and the constraints of system input. More specifically, the input shaper synthesis and the problem of choice of sampling frequency was discussed in the paper [10].

IV. The Problem of Oscillating Subsystem Identification

As already indicated, the basic assumption of the successful suppression of residual vibrations of a transition process by input shaper is the knowledge of the controlled oscillating subsystem parameters. Due to the rapid adaptation, it is advisable to devote to the identification of only that part of the system, which causes vibration. The proposed solution is illustrated in figure 4. This example does not discuss the identification of the drive of subsystem M, but only analyze the dependence of the pivot position $x(t)$ and the load position $y(t)$.

In order to verify the proposed comprehensive solutions, wireless acceleration sensors have been developed in our department. The sensors were used to measure the displacement $x(t)$ and the load movement $y(t)$. Based on the input to the oscillating system section $x(t)$ and the corresponding output $y(t)$, parameters T and b of the oscillating section were identified with the transfer function:

$$F(s) = \frac{1}{T^2 s^2 + 2bTs + 1} \tag{18}$$

V. Case Study

In general, it is common to face to the problem of identification in control applications. In order to be able to appropriately modify the control signal and thus the system response to this signal, it is important to know the transfer function of the system. For this purpose, the modules providing collection and data transfer were designed.

A. Description of used nodes

When designing the node, microcontroller ATmega168 was chosen as an element controlling the data collection and transfer. Considering that the system frequency and damping have to be known in the identification, we decided to retrieve data through LSM303DLHC module that includes a 3D digital linear acceleration sensor and a 3D digital magnetic sensor. Data obtained from the accelerometer is sent via I2C serial bus interface to the microcontroller. The microcontroller uses a timer/counter for timing each data collection process, and it ensures equidistance sampling ($fs = 400Hz$). The output of the accelerometer is represented by three 16-bit words, where each represents one of the axes x, y, and z. Gathered data is further sent through the wireless module RFM70. The proprietary module operates at a 2.4GHz frequency, and can be configured as a transmitter and as a receiver. Communication with RFM70 is ensured via the SPI interface. The data from the nodes is sent at each subsequent reading of the accelerometer, which means that we have a real-time data.

In order to get the data to the PC for processing, the node consisting of a microcontroller ATmega8 was proposed. Its role is to ensure the data reception from two transmitters via wireless module RFM70 and subsequently to send the data to

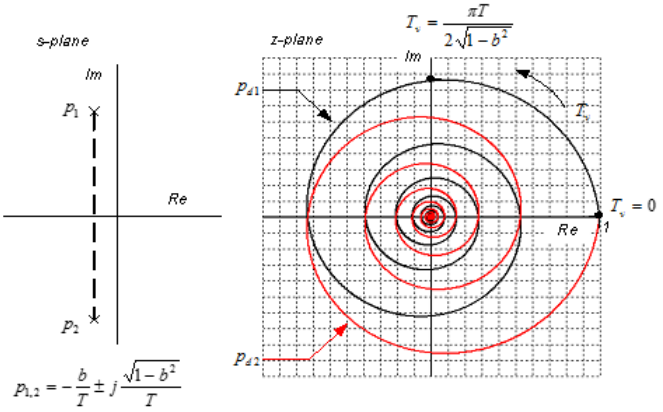


Fig. 3. Locus of the poles of discrete equivalent of the continuous system

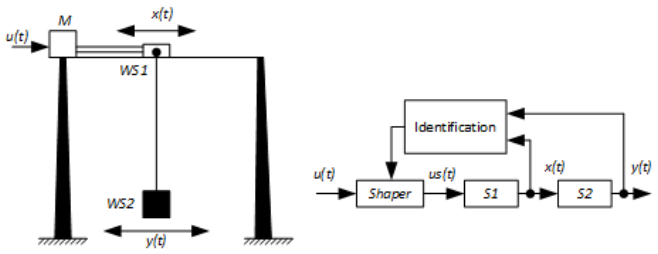


Fig. 4. Oscillating subsystem identification

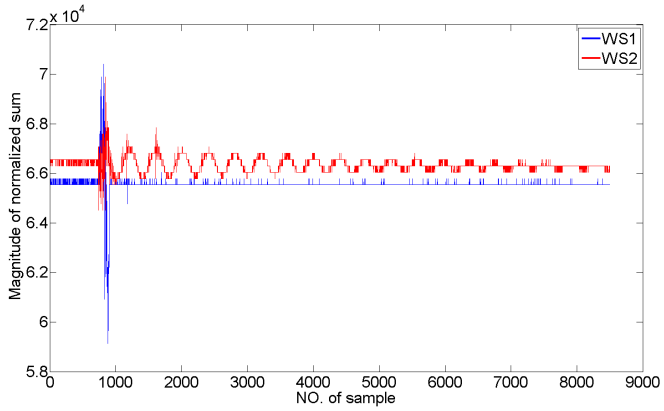


Fig. 5. Step response of weakly damped system

the PC via the UART interface. Here the data is visualized and further processed.

B. The measured results

Data obtained from the 3D accelerometers is sent via the RS232 interface to the PC where it is rendered over Matlab environment in the graph in real-time. In addition to displaying operation, the received data is recorded to the file, which forms the basis for further processing. Considering the change of the sensor position is not recorded only in one accelerometer axis, the data measured at the same time is summed. The data obtained through sensors $WS1$ and $WS2$ is represented in figure 5.

C. System identification

In order to identify the system, it is important to describe the output signal. As can be seen in figure 5, the output signal has damped oscillatory character after excitation (Fig. 6) [11]. As the input signal, the output signal is stabilized at the original level.

From the data obtained, it is approximated (determined) the system equation. After the system excitation, we get the local minimum $K-y_2$ and the local maximum y_1-K. These values are in equation 19 represented as x_1 and x_2.

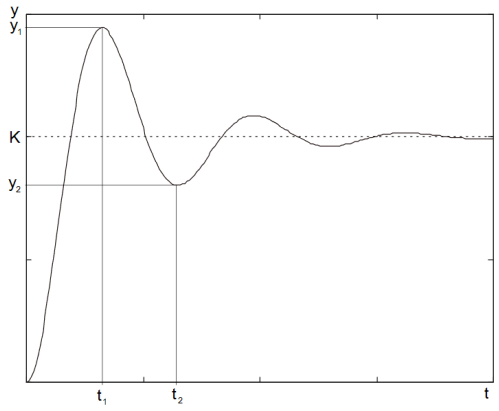


Fig. 6. Step response of weakly damped system

$$b = ln(\frac{x_1}{x_2}) \cdot \frac{2}{T}, \tag{19}$$

where b represents the damping of the system output. Period T, and thus the natural frequency of the system is defined by a difference between the times when there was a local minimum t_2 and the local maximum t_1. Substituting the obtained parameters in equation 18, we get the system step response (Eq. 20).

$$F(s) = \frac{1}{0.98^2 s^2 + 2 \cdot 0.1702 \cdot 0.98 s + 1} \tag{20}$$

VI. Conclusion

In this paper, conventional methods of input shaping techniques were reviewed. To propose suitable input shaper the controlled system should be identified. For this reason, nodes providing data from the accelerometer and wireless communication were constructed. These devices are small so they can be used also in other applications (e. g. to identify weaknesses of construction proposal and getting its mathematical model).

As the results show, the identification of second order system is not computationally demanding but captured data suffer from unwanted noise. To improve the process of identification data gathered from the accelerometer should be appropriately filtered.

There is a need to propose new nodes involving the gyroscope coming to the fore. This fact will result in the future work that will be focused on data fusion, particularly combining data from accelerometer and gyroscope.

References

[1] J. Fortgang, V. Patrangenaru, W. Singhose, " Scheduling of Input Shaping and Transient Vibration Absorbers for High-Rise Elevators", in *Proceedings of the American Control Conference*, 2006

[2] N. C. Singer, W. Seering, "Per shaping Command Inputs to reduce Systems Vibration, Journal of Dynamic Systems", in *Measurement and Control*, vol. 112, 1990

[3] P.Hubinský, P. Hauptle, "Reducing Oscillation during Positioning of a Servomechanism having Flexibility", in *Journal of Electrical Engineering* Vol.63, No.4, 2012, ISSN 1335-3632

[4] M. J. Robertson, W. E. Singhose, "Multi-Level Optimization Techniques for Designing Digital Input Shapers", in *Proceedings of the American Control Conference*, Arlington 2001

[5] E. Biediger, J. Lawrence, W. Singhose, "Improving Trajectory Tracking for Systems with Unobservable Modes Using Command Generation", in *Proceedings of the American Control Conference*, Portland 2005

[6] L. Y. Pao, C. F. Cutforth, "On Frequency-Domain and Time-Domain input Shaping for Multi-Mode Flexibile Structures", in *ASME* vol. 125, 2003

[7] N. Singer, W. Singhose, W. Seering, "Comparison of Filtering Methods for Reducing Residual Vibration", in *European Journal of Control*, 1999

[8] K. Kozak, J. H uey, W Sighose, "Performance Measures for Input Shaping", in *Proceedings of the IEEE Conf. on Control Applications*, Istanbul, Turkey, 2003

[9] T. D. Tuttle, W. P. Seering, "A Zero-placement Technique for Designing Shaped Inputs to Suppress Multiple-mode Vibration", in *Proceedings of the American Control Conference*, Baltimore 1994

[10] J. Miček, J. Juríček, "Discrete shaper of control signals", in *International journal of engineering research in Africa*. - ISSN 1663-3571. - Vol. 18 (2015), s. 65-74.

[11] M. Fikar, J. Mikleš, "Identifikácia systémov", in *https://www.kirp.chtf.stuba.sk/fikar/research/ident/ident.pdf* - ISBN 80-227-1177-2.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1107–1110
DOI: 10.15439/2016F576
ACSIS, Vol. 8. ISSN 2300-5963

The multi-topology converter for the solar panel

Samuel Žák
University of Žilina
Faculty of Management Science and Informatics,
Univerzitná 8215/1, Žilina 010 26
Email: samuel.zak@fri.uniza.sk

Peter Šarafín, Peter Ševčík
University of Žilina
Faculty of Management Science and Informatics,
Univerzitná 8215/1, Žilina 010 26
Email: {peter.sarafin, peter.sevcik}@fri.uniza.sk

***Abstract*—In this paper, we propose voltage converter with high efficiency over wide input voltage. This converter is suitable for the solar panel for WSN applications where the only power source is a solar cell that outputs highly variable voltage. The aim is to achieve this by using multiple converter topologies in parallel. Use of such converter has a meaning in renewable resources that in the long term operation significantly change their output voltage. The efficiency of particular topologies is estimated in simplified loss model, which is later experimentally tested.**

I. Introduction

EFFICIENT ENERGY production from alternative sources is one of the main problems that must be solved to make the use of solar or wind power came to the fore. The greatest weakness of renewable energy sources is their fluctuating power availability [1]. Existing power converters are tuned for operation at peak power of the particular source. The task of this paper is to design a converter that can efficiently take power from renewable source even while their energy potential is lower. To do this we need to examine the efficiency of commonly used topologies to determine their suitable operating point. This would allow creating control unit able to operate multiple topologies in parallel while maintaining better total efficiency than single topology converter.

II. The model of topologies

There are multiple topology designs for DC-to-DC converters. This paper will consider topologies with potentially lowest power loss, disregarding their other properties (output noise, transient response, etc.). Favorable candidates are elementary step-up and step-down converters (Fig. 1 and 2). They contain least components and offer relatively high efficiency while V_{in}/V_{out} ratio is close to one [2]. Harvestable power sources, however, offer a wide range of operating voltage. In such cases, other topologies may be preferable. Effective conversion with high V_{in}/V_{out} difference is a convenient property of transformers. Integrating transformer into mentioned topologies will result in fly-back transformer converter (Fig. 3) [3].

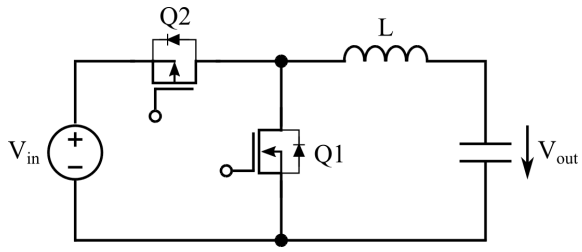


Fig. 1. Step-down converter basic topology

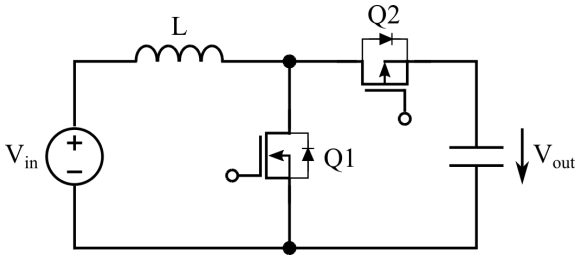


Fig. 2. Step-up converter basic topology

Our application for sensor nodes assumes solar cell as a power source. This means an available voltage will heavily fluctuate from low during dawn and dusk to high during midday. To find out which topology is most effective for given input voltage we need a model of losses for each of them. Simplified loss model of buck and boost converters can be described as shown in equation 1. Loss of the two topologies differentiate by RMS current calculation of particular branches. Table I contains the list of squared currents with their respective duty cycle D, where I_{RMS_L}, $I_{RMS_{Q2}}$ and $I_{RMS_{Q1}}$ represent effective current through inductor and transistors respectively [4].

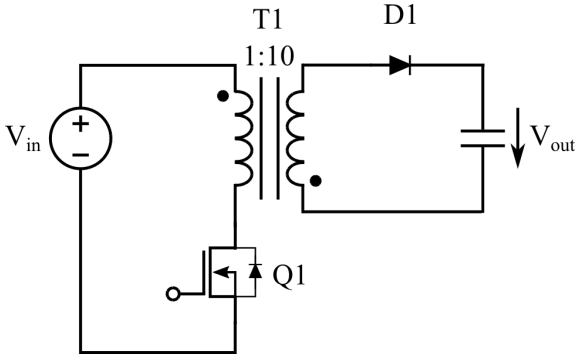


Fig. 3. Fly-back transformer topology

TABLE I
COMPUTATIONS OF CURRENTS WITH RESPECT TO THE DUTY CYCLE

	I_{RMS_L}	$I_{RMS_{Q2}}$	$I_{RMS_{Q1}}$
Buck	I_{out}^2	$I_{out}^2 \cdot D$	$I_{out}^2 \cdot (1-D)$
Boost	$(\frac{I_{out}}{1-D})^2$	$(\frac{I_{out}}{1-D})^2 \cdot (1-D)$	$(\frac{I_{out}}{1-D})^2 \cdot D$

$$\begin{aligned} P_{loss} = {} & R_{DS(ON_{Q1})} \cdot I_{RMS_{Q1}} + Q_G \cdot V_{G_{Q1}} \cdot f + \\ & \frac{1}{2} V_{IN} \cdot I_{RMS_{Q1}} \cdot (\frac{Q_G}{I_G} + \frac{Q_G}{I_G}) \cdot f + \\ & R_{DS(ON_{Q2})} \cdot I_{RMS_{Q2}} + Q_G \cdot V_{G_{Q2}} \cdot f + \\ & R_L \cdot I_{RMS_L} \end{aligned} \quad (1)$$

where:

- $R_{DS(ON_{Q1})}$ - drain-source resistance of transistor Q_1
- Q_G - total gate charge
- V_G - gate voltage (same as output voltage)
- I_G - gate charge current
- R_L - inductor resistance at switching frequency
- f - switching frequency

Power loss in the fly-back circuit can be approximated as stated in equation 2 [5]. This loss model simplifies transformer loss only to its wiring loss. Core loss is omitted due to the anticipation of low power provided by solar cell during operation of this topology.

$$\begin{aligned} P_{loss} = {} & R_{DS(ON_{Q1})} \cdot I_{RMS_{Q1}} + Q_G \cdot V_{G_{Q1}} \cdot f + \\ & \frac{1}{2} V_{IN} \cdot I_{RMS_{Q1}} \cdot (\frac{Q_G}{I_G} + \frac{Q_G}{I_G}) \cdot f + \\ & R_{L1} \cdot I_{RMS_{L1}} + R_{L2} \cdot I_{RMS_{L2}} + \\ & V_{D1} \cdot I_{RMS_{D1}} \end{aligned} \quad (2)$$

where:

- R_{L1} - primary winding at f
- R_{L2} - secondary winding at f
- V_{D1} - diode forward voltage drop

Duty cycle computation for boost and fly-back topologies is described in equation 3 and 4 respectively [6].

$$D = 1 - \frac{V_{in}}{V_{out}} \quad (3)$$

$$D = \frac{\frac{V_{out}}{V_{in}}}{\frac{N_s}{N_p} + \frac{V_{out}}{V_{in}}} \quad (4)$$

A. Optimal operating point

Next step is to fill described models with parameters of real-world equipment measured at switching frequency (Tab. II). We have chosen components that will later be used for testing. However, these components are by no means ideal for the power converter, which will be reflected in converter's efficiency [7].

We will simulate load at 3.3V drawing 10mA constant current. To minimize unaccounted parasitic losses, we have chosen lowest operating frequency that allows continuous conduction mode of operation. Transistors will, therefore, operate at 100kHz. Figure 4 shows the efficiency of step-up topologies relative to the input voltage. The simulation shows ranges of input voltage where respective topologies achieve better efficiency than the others.

TABLE II
LIST OF COMPONENTS USED FOR THE TESTING

	Component
N-channel MOSFET	TN0604N3
P-channel MOSFET	LP0701N3
Inductor	Generic toroid coil
	L = 203uH R=5.35 OHM
Diode	IN5817
Transformer	Rp=1.16 OHM, Lp = 51.9uH
	Rs= 33.6 OHM, Ls = 5.116mH
	Turn ratio 1:20

III. DESIGN OF MULTI-TOPOLOGY CONVERTER

Utilizing optimal operating point given by particular topology can be achieved by power management system with multiple parallel topologies and suitable control algorithm. The parallel conjunction of buck and boost topologies require some modifications [8]. Main reason is body diode of switching transistors. In the case of boost topology (Fig. 2), body diode

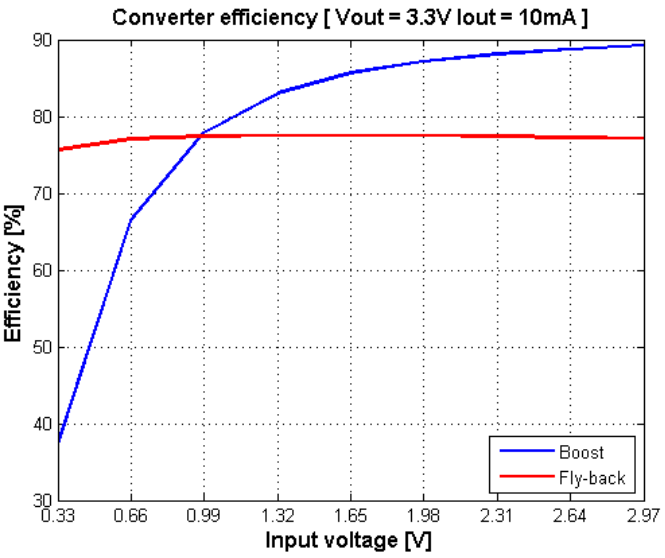


Fig. 4. Converter efficiency

of $Q2$ restricts the use of input voltage to value lower than the output voltage. If this restriction is not respected, body diode will conduct current making it impossible to maintain regulation on the output voltage. The similar limitation can also be observed on buck topology when the output voltage is greater than the input. This raises a need to add an extra transistor, which would be controlled to restrict current in cases where body diode is conductive. The more convenient solution is to add not one, but two transistors resulting a buck-boost topology where input and output voltages are separated by two transistors (Fig. 5). This modification will increase overall power loss due to extra resistance of constantly opened transistor - $R_{DS(ON)}$ of P-channel transistor.

This way, the properly designed control algorithm can step-up or step-down input voltage regardless of V_{in}/V_{out} ratio. Control signals needed for operation are described in Tab. III.

Choice of topologies and their total number depends on the used power source. For WSN node with a low power solar cell, we have chosen one fly-back and one buck-boost topology. For more power demanding appliances and sources it may be beneficial to have multiple buck-boost topologies due to the decrease of conductive losses [9]. Schematic of resulting converter for a solar cell is shown in figure 6.

IV. Experimental testing of the model

Verification of loss model will be performed on designed converter in boost operating mode. The converter will be tested for efficiency at the constant output voltage and current with a variable input voltage. Measured results are compared with loss model in Tab. IV.

Measurement confirms modeled data within an acceptable margin of error. As model estimated, this converter performs poorly while there is large difference between the input and the output voltage. This is due to high conduction losses that contribute about 90% of the total loss. Main reason is properties of used testing components. Better overall efficiency can be achieved by balancing conduction losses with switching losses (i.e. using transistors with lower drain-source resistance at cost of higher gate charge).

TABLE III
Control signals description

	Inactive	Boost	Buck
Q1	High / closed	Low / open	Switching at D
Q2	High / closed	Switching at (1-D)	Low / open
Q3	Low / closed	Low / closed	Switching at (1-D)
Q4	Low / closed	Switching at D	Low / closed

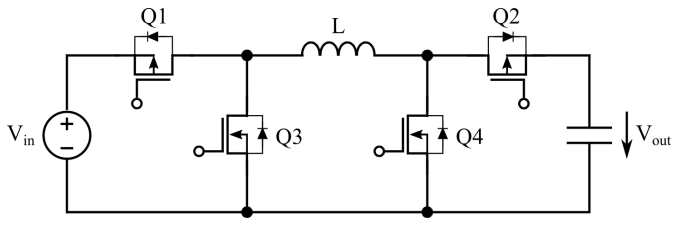


Fig. 5. Buck-boost topology

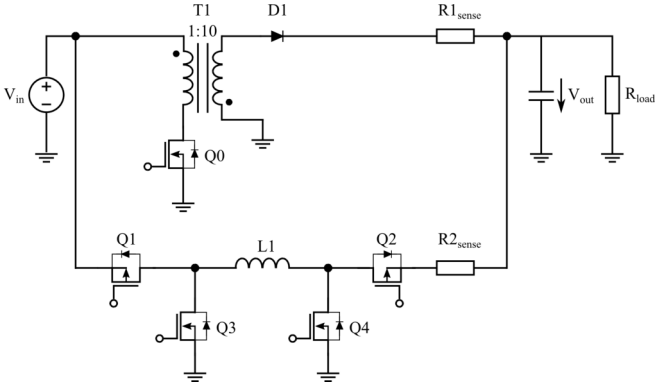


Fig. 6. Schematic of resulting converter for a solar cell

TABLE IV
Modelled and measured efficiency comparison

Input/output voltage ratio [%]	Modeled efficiency [%]	Measured efficiency [%]
10	37.32	34.81
20	66.44	65.10
30	77.90	75.94
40	82.99	81.73
50	85.63	84.33
60	87.16	85.64
70	88.12	87.60
80	88.77	88.05
90	89.22	88.67

This paper, however, shows relative efficiency between two voltage increasing topologies. The efficiency of both, boost and fly-back topologies can be improved with technologically advanced components. The elementary shape of their efficiency to input voltage function should, however, remain similar, which is given by topology operation. With boost topology, a high difference between input and output affects duty cycle. High duty cycle means a long time to charge inductor and a short time to discharge it. This creates high current peaks and therefore a significant loss. The fly-back converter may use transformer ratio to cope with high voltage difference and keep duty cycle in balance. This can also be seen from the computation of their respective duty cycle (Eq. 3 and 4).

V. Conclusion

We have evaluated voltage converter which utilizes multiple parallel topologies. Modeled data shows convenience of parallel operation of different topologies. As a result, fly-back and boost topologies are well suited for complementary operation where fly-back will be active at the lower input voltage and boost at higher. We also develop control algorithm selecting best available topology for devices with low computational power.

Implementation of modeled data into converter control may have a form of the lookup table where the device would measure all relevant parameters (input and output voltage, output current and current passing each active topology) and decide which topology (or their combination) to use. Experimental measurement confirmed our expectations about the efficiency of single topology derived from the model. To perfect efficiency for wider operating conditions, more topologies need to be examined.

REFERENCES

[1] M. Kochláň et al, "Control unit for power subsystem of a wireless sensor node", in *FedCSIS proceedings of the 2015 Federated conference on Computer science and information systems*, 2015, pp. 1249-1256, ISBN 978-83-60810-65-1.

[2] Mohan, Undeland and Robbins, "Power Electronics, Converters, Applications and Design." 2nd edition Wiley. ISBN 0–471–58408–8.

[3] R. D. Middlebrook and S. Cuk, "A General Unified Approach to Modeling Switching-Converter Power Stages", in *International Journal of Electronics*, Vol. 42, No. 6, pp. 521-550, June 1977.

[4] AN707, "Designing a High Frequency, Self–Resonant Reset Single Switch Forward Converter Using Si9118/Si9119 PWM/PSM" in *http://www.vishay.com/document/70824/70824.pdf*.

[5] AN1114, "Switch Mode Power Supply (SMPS) Topologies (Part I)" in *http://ww1.microchip.com/downloads/en/AppNotes/01114A.pdf*.

[6] AN607, "DC-to-DC Design Guide" in *http://www.vishay.com/docs/71917/71917.pdf*.

[7] G. W. Wester and R. D. Middlebrook, Low-Frequency Characterization of Switched Dc-Dc Converters, in *IEEE Transactions an Aerospace and Electronic Systems*, Vol. AES-9, pp. 376-385, May 1973.

[8] G.A. Rincón-Mora, Power Management ICs - A Top-Down Design Approach, ISBN: 1-4116-6359-4.

[9] M. Gildersleeve, G.A. Rincón-Mora et al, "A comprehensive power analysis and a highly efficient, mode-hopping DC-DC converter," in *IEEE Asia-Pacific Conference on ASIC*, 2002, pp. 153-156.

Information Technology for Management, Business & Society

IT4MBS is a FedCSIS conference area aiming at integrating and creating synergy between FedCSIS events that thematically subscribe to the disciplines of information technology and information systems. The IT4BMS area emphasizes the issues relevant to information technology and necessary for practical, everyday needs of business, other organizations and society at large. This area takes a sociotechnical view on information systems and relates also to ethical, social and political issues raised by information systems. Events that constitute IT4BMS are:

- ABICT'16—7th International Workshop on Advances in Business ICT
- AITM'16—14th Conference on Advanced Information Technologies for Management
- ISM'16 - 11th Conference on Information Systems Management
- KAM'16—22nd Conference on Knowledge Acquisition and Management
- UHH'16 - 2nd International Workshop on Ubiquitous Home Healthcare

7th International Workshop on Advances in Business ICT

ABICT focuses on Advances in Business ICT approached from a multidisciplinary perspective. It will provide an international forum for scientists/experts from academia and industry to discuss and exchange current results, applications, new ideas of ongoing research and experience on all aspects of Business Intelligence. ABICT will be also an opportunity to demonstrate different ideas and tools for developing and supporting organizational creativity, as well as advances in decision support systems.

We kindly invite contributions originating from any area of computer science, information technology and computational solutions for different applications areas, data integration and organizational implementation of ABICT, as well as practical ABICT solutions.

TOPICS

Topics include (but are not limited to):

- Advanced technologies of data processing, content processing and information indexing
- Analytics as a service
- Big Data: benefits and challenges
- Business Analytics
- Business applications of social networks
- Business data mining and knowledge discovery
- Business Intelligence
- Business Rules
- Business-oriented time series data mining, analysis, and processing
- Cloud based Business Intelligence
- Creativity Support Tools
- Customer Relationship Management, social Customer Relationship Management
- Data driven marketing
- Data Warehousing
- Decision support
- Digital Business Strategy
- Enterprise Device Management
- ICT technologies in enterprise management
- Information forensics and security, information management, risk assessment and analysis
- Information Systems Design
- Internet of Things
- Knowledge Management (for better Decision Support, Collaboration and Competitiveness)
- Legal text processing
- Leveraging ICT for Transforming Organization
- M2M Device Management, M2M Solutions
- Semantic Web and Ontologies in Business ICT
- Virtual Enterprise
- Web 2.0 and Web 3.0 in fusing Business Intelligence systems and Decision Support Systems
- Web-Based Data Management Systems

EVENT CHAIRS

- **Mach-Król, Maria,** University of Economics in Katowice, Poland
- **Olszak, Celina M.,** University of Economics in Katowice, Poland
- **Pełech-Pilichowski, Tomasz,** AGH University of Science and Technology, Poland

PROGRAM COMMITTEE

- **Abramowicz, Witold,** Poznan University of Economics, Poland
- **Andres, Frederic,** National Institute of Informatics, Tokyo, Japan
- **Badica, Amelia,** University of Craiova, Romania
- **Basiura, Beata,** AGH University of Science and Technology, Faculty of Management, Poland
- **Berio, Giuseppe,** Universite de Bretagne Sud, France
- **Brown, David,** Lancaster University Management School, UK, United Kingdom
- **Christozov, Dimitar,** American University in Bulgaria, Bulgaria
- **Gaweł, Bartłomiej,** AGH University of Science and Technology, Poland
- **Grabowski, Mariusz,** Krakow University of Economics, Poland
- **Hussain, Fehmida,** School of Science and Technology, Dubai
- **Kacprzyk, Janusz,** Institute of Computer Science, Polish Academy of Sciences, Poland
- **Khachidze, Manana,** Tbilisi State University, Georgia
- **Kiełtyka, Leszek,** University of Technology in Czestochowa, Poland
- **Kisielnicki, Jerzy,** University of Warsaw, Poland
- **Konikowska, Beata,** Institute of Computer Science, Poland
- **Korwin-Pawlowski, Michael L.,** Universite du Quebec en Outaouais, Canada
- **Kulczycki, Piotr,** Systems Research Institute, Polish Academy of Sciences, Poland
- **Loucopoulos, Peri,** Harokopio University of Athens, Greece

- **Madeyski, Lech,** Wrocław University of Technology, Poland
- **Michalik, Krzysztof,** University of Economics in Katowice
- **Milewski, Robert,** Medical University of Bialystok, Department of Statistics and Medical Informatics, Poland
- **Opiła, Janusz,** AGH University of Science and Technology, Poland
- **Owoc, Mieczyslaw,** Wroclaw University of Economics, Poland
- **Paliński, Andrzej,** AGH University of Science and Technology, Poland
- **Petrov, Oleksandr,** AGH University of Science and Technology, Ukraine
- **Petryshyn, Lubomyr,** AGH University of Science and Technology, Poland
- **Prasad, T. V.,** Chirala Engineering College, India
- **Pulvermueller, Elke,** University Osnabrueck, Germany
- **Reimer, Ulrich,** University of Applied Sciences St. Gallen, Switzerland
- **Rossi, Gustavo,** National University of La Plata, Argentina
- **Roztocki, Narcyz,** State University of New York at New Paltz, USA and Kozminski University, Poland
- **Salem, Abdel-Badeeh M.,** Ain Shams University, Egypt
- **Sankowski, Dominik,** University of Technology in Łódź, Poland
- **Sauer, Jurgen,** University of Oldenburg, Germany
- **Schroeder, Marcin,** Akita International University, Japan
- **Skalna, Iwona,** AGH University of Science and Technology, Faculty of Management, Poland
- **Soja, Piotr,** Cracow University of Economics, Poland
- **Stawowy, Adam,** AGH University of Science and Technology, Faculty of Management, Poland
- **Szpyrka, Marcin,** AGH University of Science and Technology, Poland
- **Teufel, Stephanie,** University of Fribourg, Switzerland
- **Tvrdikova, Milena,** VŠB Technological University of Ostrava, Faculty of Economics, Czech Republic
- **Zaliwski, Andrew,** University of Auckland
- **Zieliński, Jerzy S.,** University of Lodz, Poland
- **Zurada, Jozef,** College of Business University of Louisville, United States

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1115–1123
DOI: 10.15439/2016F328
ACSIS, Vol. 8. ISSN 2300-5963

Overview of Time Issues with Temporal Logics for Business Process Models

Krzysztof Kluza, Krystian Jobczyk, Piotr Wiśniewski, Antoni Ligęza
AGH University of Science and Technology
al. A. Mickiewicza 30, 30-059 Krakow, Poland
E-mail: {kluza,wpiotr,ligeza}@agh.edu.pl

***Abstract*—Process models can specify various aspects of business processes. In this paper, we present an overview of the existing solutions for describing time aspects of such models. We focus on Business Process Model and Notation and provide examples of representing time patterns in this notation. As temporal issues can be specified using temporal logics, we provide a short overview of selected temporal logics which can be used to specify the time patterns in business process models.**

***Index Terms*—BPMN, Business Processes, Temporal Logics, Temporal Issues, Time Patterns**

I. Introduction

BUSINESS Process Management (BPM) [1] is a modern approach to improving organization's workflow, which focuses on reengineering of processes to obtain optimization of procedures, increase efficiency and effectiveness by constant process improvement.

The key aspect of BPM is a Business Process (BP). A BP is usually described as a collection of related activities which transform different kinds of clearly specified inputs to produce a customer value, mainly considered as products or services and organizational goals, as output [2]. Such a process can be represented as a BP model [3]. However, there can be many representation methods for modeling processes. We give an overview of the existing representation methods for business processes, such as Petri nets, EPC, IDEF3, UML AD, YAWL and BPMN diagrams.

As temporal logics can have various applications related to knowledge management [4]–[6] and business process models [7], it is important to consider process representation from the perspective of time issues. Thus, we analyze the presented process representations in terms of time-related elements. The main focus is on the BPMN notation and time issues in this notation. We present how to represent the Allen's algebra relations using BPMN notation. As it was proven empirically in [8], representing relations using the 2D models is efficient in terms of understanding temporal aspects of time intervals.

Moreover, based on the existing time patterns from the literature, we describe how they can be used in BPMN as well as we provide a short overview of temporal logics which can support these time patterns.

The rest of this paper is organized as follows: Section II presents an overview of process modeling notations with with an emphasis on time or temporal aspects.. The most popular BPMN representation is analyzed in details in this context in Section III. In Section IV, we give an overview of temporal logics which can be used to support time patterns for process models. The paper is summarized in Section V.

The paper is supported by the AGH UST grant.

II. Business Process Representation

As process modeling is an essential part of BPM, in the following subsections, we present the most popular business process representations.

Although there are many process modeling languages, we focus here on the six visual and most successful ones: EPC, IDEF3, UML AD, Petri net, YAWL and BPMN. Processes in these languages consist of activities, which may be decomposed to subactivities. The order of activities defines the sequence of work. In the lower level, each activity transforms some inputs into outputs.

Table I presents a simple yet illustrative example of a car rental process [9] in the above mentioned notations. The process starts with a registering request, and then extra insurance can be added. When check-in is initiated, the customer can select a car; at the same time the driver's license is checked and the customer's credit card is charged. The process ends when the chosen car is supplied to the customer. This exemplary process contains only basic control flow elements, which can be represented in all of these languages. However, it is important to mention that the expressiveness of each of these languages is much higher than the required for this example.

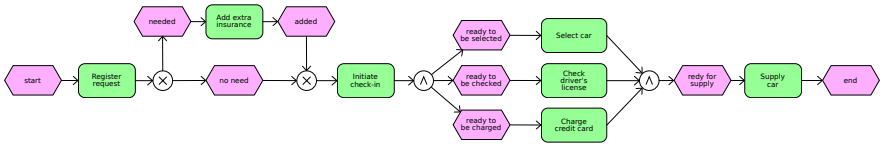
A. Process Modeling Languages

1) Event-driven Process Chain (EPC): Event-Driven Process Chain (EPC) is a simple graphical modeling language for modeling processes introduced by August-Wilhelm Scheer [10]. The EPC models are directed graphs visualizing the control flow [11]. Each EPC consists of events, functions and connectors, starts with at least one event and ends with at least one event, In EPC, each event triggers a function that leads to a new event. The notation supports three types of connectors (AND, XOR, OR) which can be used to model splits and joins.

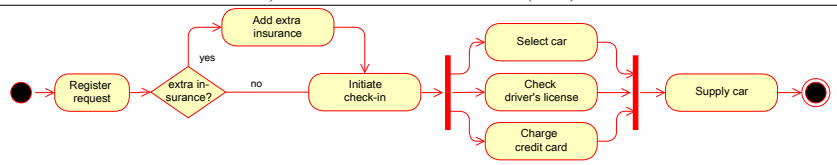
As events in EPC are rather passive elements which specify under what circumstances a process works or what is a result state of a process, such constructs are rather not suitable for time-related extensions.

978-83-60810-90-3/$25.00

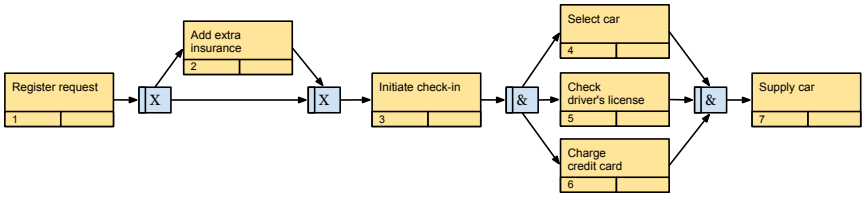
Table I
AN EXAMPLE OF CAR RENTAL PROCESS IN VARIOUS PROCESS MODELING NOTATIONS



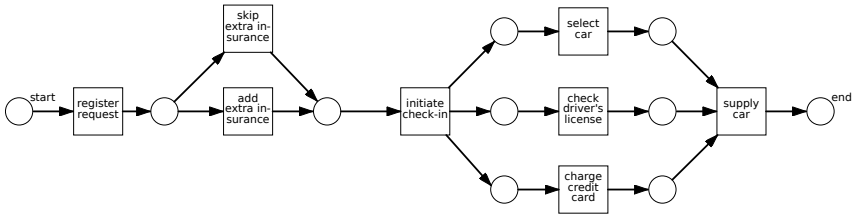
1) Event-driven Process Chain (EPC)



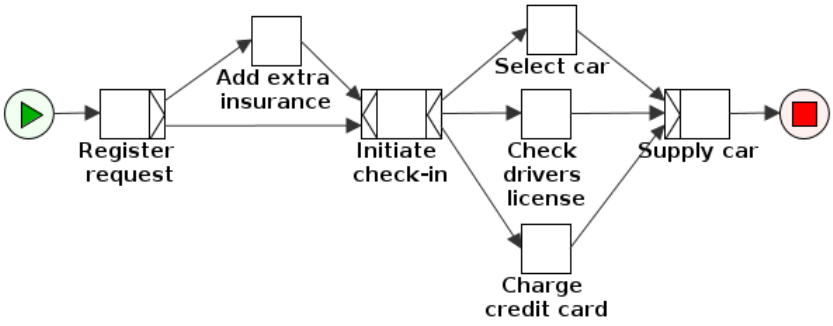
2) UML Activiti Diagram (UML AD)



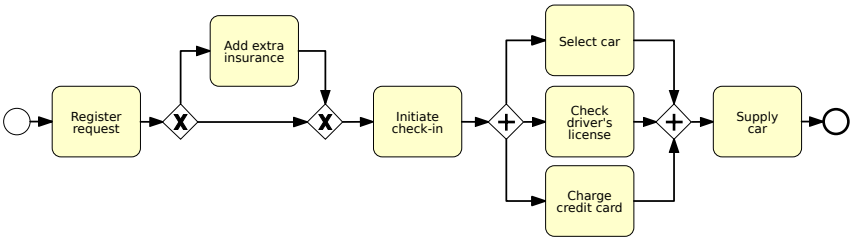
3) Integrated DEFinition Method 3 (IDEF3)



4) Petri net



5) Yet Another Workflow Language (YAWL)



6) Business Process Model and Notation (BPMN)

2) Integrated DEFinition Method 3 (IDEF3): Integrated Definition (IDEF) is a family of modeling languages that arose in the 1979s out of the U.S. Air Force in order to increase manufacturing productivity [12], [13]. The IDEF suite contains sixteen types of modeling languages in the field of systems and software engineering; however, the majority is still under development and only IDEF0-4 are commonly used in practice. IDEF3 is concerned with modeling the processes of a business or its systems [14], [15]. IDEF3 Process Flow diagram consists of such elements as Unit of Behavior (UOB) boxes, precedence links, junctions (AND, XOR, OR), referents, and notes. An IDEF3 process description organizes the network of relations between situations in a specified scenario from the process-centered perspective.

As in IDEF3, events are not modeled directly, there are no time events. However, IDEF3 specifies constraint precedence links which express simple temporal precedence relations between instances of one UOB and another UOB. They add constraints over and above the activation semantics of simple precedence and can indicate that an instance of the source UOB must be followed or preceded by an instance of the destination UOB [16].

3) UML Activiti Diagram (UML AD): The Unified Modeling Language (UML) [17] from the Object Management Group (OMG) constitutes a standardized notation for modeling software applications [18]. This multipurpose modeling language offers a variety of notations to capture different aspects of software [19], [20]. UML has become the dominant notation among software engineers and attempts to be a universal visual notation for software design. Activity diagram (AD) is a kind of the UML behavior diagrams for modeling dynamic aspects of the system, which focuses on the flow of activities involved in a single process and shows the dependencies among them. Although UML AD can be used for process modeling purposes, its complex nature makes it a barrier for non-technicians and it is not suitable for all aspects of this type of modelling. A simple process consists of a sequence of nodes modeled using control-flow and object-flow. The control-flow comprises two types of nodes: action nodes (activities to be performed or signals to be received/sent by the process) and control nodes, which model sequencing and parallel or alternative branching. The flow of data between nodes can be represented by associations of object nodes with activities.

In the case of UML AD, a time event trigger is supported (notated with an hour glass symbol). On the other hand, UML semantics do not dictate the amount of time between actions or events. However, additional timing constraint elements can be defined using customized stereotypes or specified in OCL (Object Constraint Language), especially as pre- and post-conditions for actions.

4) Petri net: Petri nets [21] offer a graphical notation for modeling processes that include choice, iteration, and concurrent execution. A classical Petri net is a directed graph composed of two types of nodes: places and transitions. An arc in the net may connect a place to a transition or vice versa, but no arc may connect two nodes of the same type. A transition can have a number of input and output places. As Petri nets have an exact mathematical definition of their execution semantics and a well-developed mathematical theory for process analysis, they are suitable for modeling, analysis and simulation of dynamic systems. As generally the execution of Petri nets is nondeterministic, they are well suited for modeling the concurrent behavior of distributed systems [22].

Although the original Petri nets do not model time issues explicitly, there are extensions, such as time Petri nets, which were used in analysis of concurrent systems where behavior was dependent on explicit values of time [23]. A detailed overview of the algorithms that allow for analysing time-dependent Petri nets can be found in the book [24].

5) Yet Another Workflow Language (YAWL): YAWL is a Petri Net based workflow language [25], [26]. Having formal semantics, it supports specification of the control flow and the data perspective of business processes. The language encompasses workflow patterns to guarantee language expressibility [27]. The YAWL language extends the class of workflow nets with multiple instances, OR-joins and cancelations. Workflow net (WF-net) [28] is a subset of Petri net used to model workflows. In a WF-net, there is a unique input place and a unique output place and every other place and transition are on a directed path from input to output place, and WF-nets introduce additional notations for joins and splits (AND and XOR). In the case of YAWL, in contrast to Petri nets and WF-nets, its syntax allows tasks to be directly connected, which helps compress the visual representation of a YAWL model.

Although in YAWL there are no separate time elements, any atomic task can be assigned a timer behaviour using a timer property. Such timer can be activated either when a task is enabled (i.e. is offered or allocated) or when it starts. What is more, there are also additional timer predicate expressions which can be used as flow predicates.

6) Business Process Model and Notation (BPMN): BPMN [29], adopted by the OMG group, is the most widely used notation for modeling BPs. The BPMN notation uses a set of models with predefined graphical elements to depict a process and how it is performed. Although the current BPMN 2.0 specification [29] defines three models to cover various aspects of Business Processes, in most cases, using only the Process Model is sufficient. Thus, in this paper we analyze the internal Business Process Model of BPMN, which can be compared to the previously presented approaches.

Table II presents an evaluative analysis of the described Business Process modeling languages (the summary prepared on the basis of [30]–[35]). The symbols in the tables have the following meaning: ❍ – not supported, ● – supported. ◗ – partially supported (not standardized or possible to present but not directly).

As one can see from the Table II, BPMN 2.0, a de facto industry standard for modeling processes, supports most of the listed elements. Thus, in the following sections we will focus on this notation and present time issues related to BPMN.

Table II
COMPARISON OF THE SUPPORTED ELEMENTS IN PROCESS MODELING LANGUAGES

		Business Process modeling languages					
		Petri net	**EPC**	**IDEF3**	**UML AD**	**YAWL**	**BPMN**
Year		1962	1992	1995	1997	2004	2004
Creator		C. Petri	A.-W. Scheer	U.S. Air Force	OMG	van der Aalst	OMG
Background		Academic	Academic	Industry	Industry	Academic	Industry
Standardised		N.A.	No	Yes	Yes	No	Yes
Metamodel		○	◗	○	●	○	●
Purpose	Formal methods	●	○	○	○	◗	○
	Graphical	◗	●	●	●	●	●
	Execution	○	○	○	○	●	●
Activities	Atomic	◗	◗	◗	●	●	●
	Subprocess	◗	◗	◗	●	●	●
Events		●	●	○	●	●	●
Gateways	AND	◗	●	●	●	●	●
	XOR	◗	●	●	●	●	●
	OR	○	●	●	●	●	●
	Complex	○	○	○	◗	◗	●
Participants	Internal	○	●	○	●	◗	●
	External	○	○	○	●	○	●
Data	Data flow	○	●	○	●	◗	●
	Data objects	○	◗	◗	●	◗	●
	Data repository	○	○	○	●	○	◗
Time-related	Events	○	○	○	●	◗	●
	Activities	○	○	◗	◗	◗	◗
	Gateways	○	○	○	◗	◗	◗

III. TIME ISSUES IN BUSINESS PROCESS MODELS

A. Time Representation in BPMN

If it comes to the time-related issues in BPMN, some of them can be represented directly in the following way:

- Timers are supported as Timer start and Timer intermediate events (see Figure 1) For such an element, one of the following properties can be specified:
  - time date – specifies a fixed date when trigger will be fired,
  - time duration – specifies how long the timer should run before it is fired, or
  - time cycle – specifies repeating interval, which can be useful for starting process periodically, or for sending multiple reminders for overdue user task.
- Event-based Subprocess are subprocesses which are started by an event, such as a timer (see Figure 2).
- Timer boundary events allow for following additional or alternative control flow when the timer is fired. This can either interrupt or not the associated task or subprocess (see Figure 3).
- Event-based gateway with timers works like an exclusive gateway (XOR) as both involve one path in the flow (see Figure 4). In the case of an event-based gateway, however, it is evaluated which event has occurred, not which condition is being met.

Apart from the time issues which are basically supported by the existing BPMN elements, some issues can be modeled indirectly using more complex combination of elements. In Table III, we presented the classic Allen's algebra relations [36] applied to BPMN models. On the left hand side there is a simple model with additional artefacts depicting time relations. On the right hand side, we present a refined model, compliant with the BPMN 2.0 specification [29] which should fulfil the presented relation (this can be also interpreted as a method of imposing some constraints related to Allen's algebra relations). As there are many equivalences in BPMN [37], the presented models are not the only one that are possible. One can also noticed that in some cases, the model is quite complex and requires additional BPMN elements.

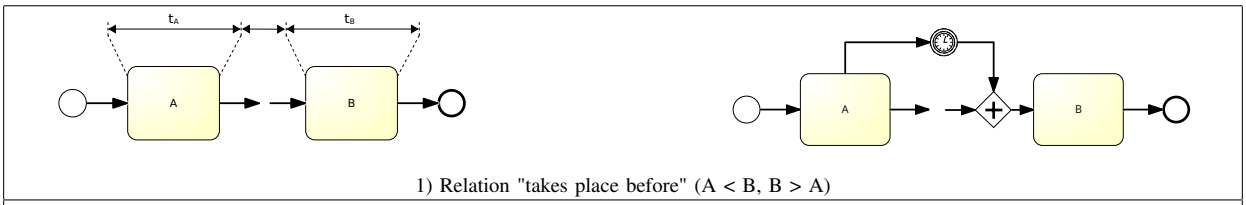
Simpler models for these relations can be found in [38] and [39]. However, they require additional non-standardized elements. Thus, they extend the BPMN notation.

A survey of some time-related aspects of process models conducted by Cheikhrouhou et al. can be found in [40], [41]. Their survey focused on the existing approaches to specifying and verifying temporal aspects of processes, not focusing on processes represented using the BPMN notation, but rather temporal constraints specification methods.

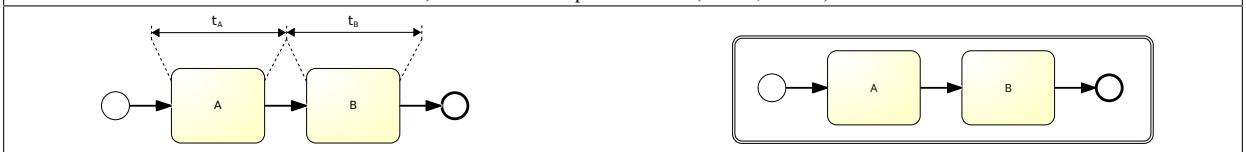
B. Time Patterns in Process Models

The objective of time patterns is to facilitate the analysis and comparison of Process-Aware Information Systems (PAIS). Their classification and selection criteria were presented in [42] and further developed in [38]. In [43], a dedicated tool for PAIS management and support was demonstrated. Time constraints related to specific patterns can be

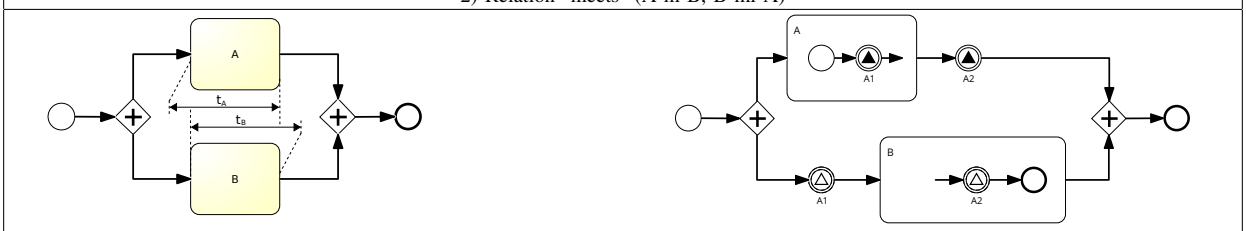
Table III
ALLEN'S RELATIONS IN BPMN: LEFT: STANDARD MODEL WITH TWO TASK (WITHOUT IMPOSED RESTRICTIONS), RIGHT: REFINED MODEL FULFILLING THE RELATION



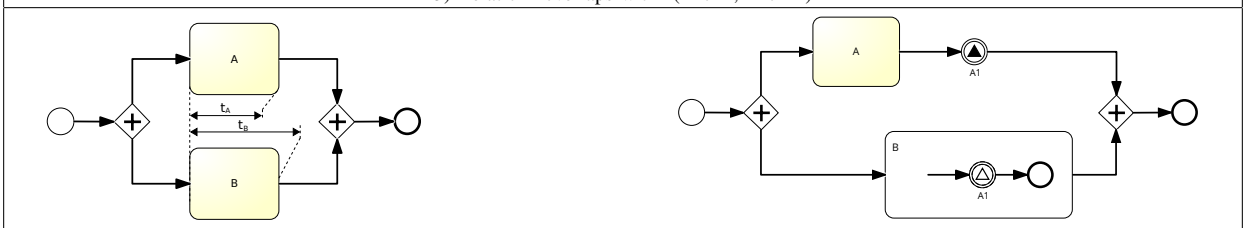
1) Relation "takes place before" (A < B, B > A)



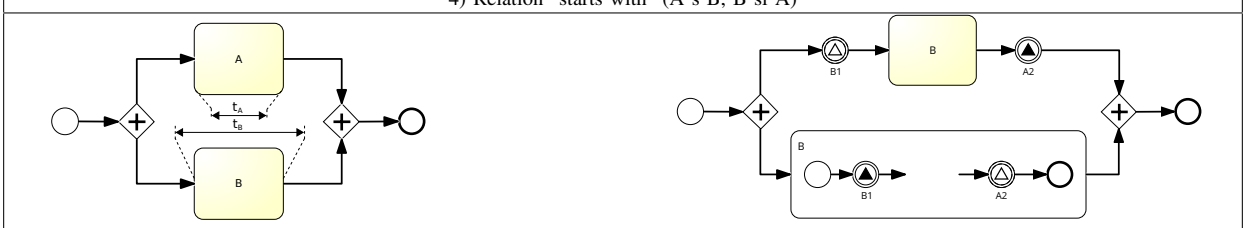
2) Relation "meets" (A m B, B mi A)



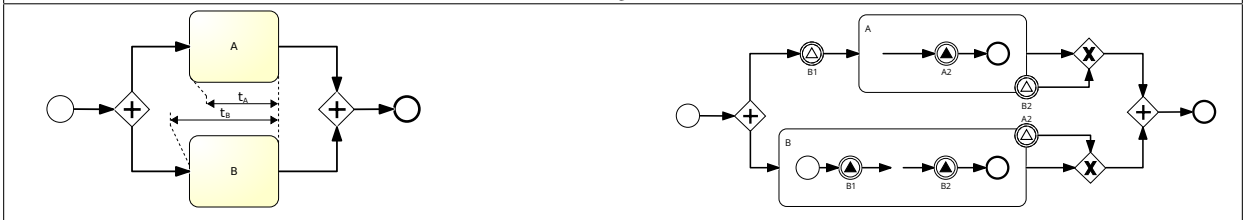
3) Relation "overlaps with" (A o B, B oi A)



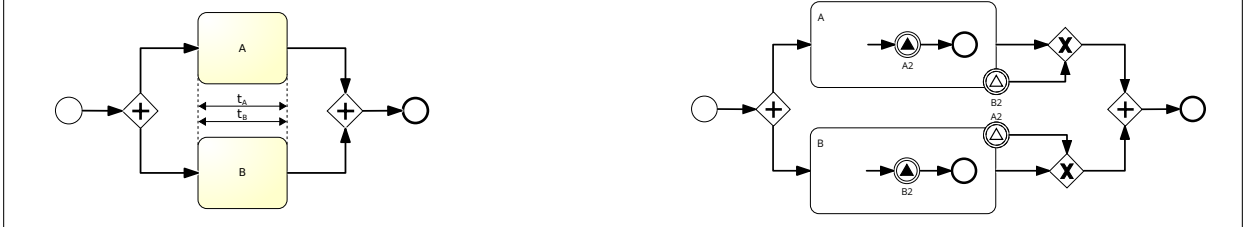
4) Relation "starts with" (A s B, B si A)



5) Relation "during" (A d B, B di A)



6) Relation "finishes" (A f B, B fi A)



7) Relation "is equal to" (A = B)

used in generating optimized plans for declarative process models [44], where a user determines the final purpose instead of an explicit task sequence. Another approach is to present temporal information along with resources using a modified Petri Net [45], where validity periods as well as maximal processing times are determined.

Thus, in [38], based on the representative set of business process models from different domains, 10 time patterns (divided to four categories) were identified and described:

I Duration and Time Lags
- TP1 Time Lags between two Activities
- TP2 Durations
- TP3 Time Lags between Arbitrary Events

II Restricting Execution Times
- TP4 Fixed Date Elements
- TP5 Schedule Restricted Elements
- TP6 Time-based Restrictions
- TP7 Validity Period

III Variability
- TP8 Time-dependent variability

IV Recurrent Process Elements
- TP9 Cyclic Elements
- TP10 Periodicity

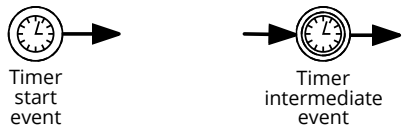


Figure 1. BPMN Timers

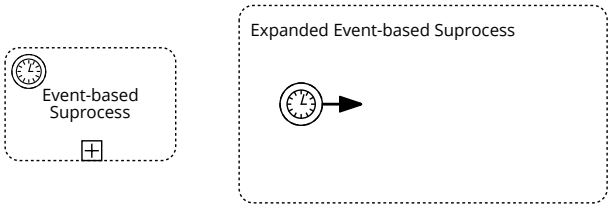


Figure 2. BPMN Event-based Subprocess

Each of these patterns can be used in various forms and applied to different elements (design choices). Thus, some of such choices of the patterns TP 1-4, 8 and 9 are supported by BPMN constructs as well:

- TP1 – supported by timers or combination of signals, as in diagrams in Table III,
- TP2 – supported by boundary timer events, as in Figure 3,
- TP3 – supported by timer intermediate event, as presented in Figure 1,
- TP4 – supported by timer start event, as in Figure 1 and 2,
- TP8 – supported by event-based gateway with timers, as in Figure 4,

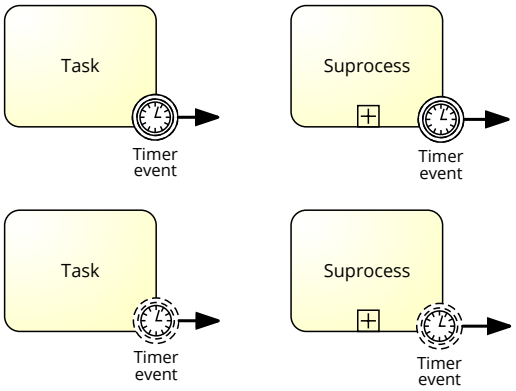


Figure 3. BPMN Timer boundary events

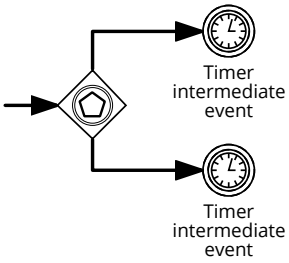


Figure 4. BPMN Event-based gateway with timers

- TP9 – supported either by a loop modeled using control flow with a timer intermediate event or by multi instance loop task or subprocess.

C. Analysis and Verification of Time Related Process Models

An important issue in process modeling is verification process models [46]–[48]. Among the existing papers concerning time issues in business processes, there are also papers related to analysis and verification of process and workflow models. There are several methods that can be used for validation of time related processes [49]–[56]. One of the existing methods concerns dynamic detection of temporal violations and providing possible solutions to a specific problem [49]. An analysis of time compatibility of web services with respect to time constraints, correlated with data and caused by message interaction between services, was conducted in [51]. As an extension to this solution a technique of analyzing and validating a set of processes was presented in [52]. It includes checking if a process choreography fulfils the declared time requirements and is based on Business Process Execution Language models which are then interpreted in the Fiacre formal verification language. Another method used to verify temporal constrains is based on new formal language, XTUS-Automata, which can be used to specify both relative and absolute time properties and also include deadlock verification [53]. This solution use temporal specification patterns, which are provided for different types of properties. In [54], a solution where an extended BPMN model is mapped onto timed automata and then verified using UPPAAL model checker is proposed. Temporal requirements

Table IV
COMPARISON OF THE EXISTING PROCESS VALIDATION METHODS

Authors	Du et al. [49]	Guermouche [51]	Guermouche et al. [52]	Kallel et al. [53]	Watahiki et al. [54]	Makni et al. [55]	Wong et al. [56]
Year	2011	2010	2012	2009	2011	2010	2009
Model	Time WF-net	Timed automata	Modified BPMN and Fiacre	XTUS automata	Extended BPMN and timed automata	Timed Petri Net	CSP algebra
Model checker	UPPAAL	UPPAAL	TINA	UPPAAL	UPPAAL	TINA	FDR
Graphical model	●	●	●	◑	●	●	❍
Deadlock detection	◑	●	●	●	●	❍	❍

represented by deadline constraints within inter-organizational workflows, where certain private processes remain unexposed for other users, can be verified using modified Petri Nets [55]. Another approach is base on building a semantic model of a business process, including the occurring delays, and its verification using CSP process algebra [56]. Table IV presents a summary of selected process validation techniques.

IV. TEMPORAL LOGICS FOR TIME PATTERNS IN PROCESS MODELS

Based on the presented time patters [38], we analyzed selected temporal logics [57], [58] in order to assess their suitability to represent these patterns.

A. Linear Temporal Logic

Linear Temporal Logic (LTL) [59] is a temporal logic system for representing of time linearity. Its language is obtained from standard propositional language (with the Boolean constant $\top$) by adding temporal-modal operators such as: *always in a past* (H), *always in a future* (G), *eventually in the past* (P), *eventually in the future* (F), *next and until* ($\mathcal{U}$) and *since* ($\mathcal{S}$) – co-definable with "until".

B. Computation Tree Logic

LTL is traditionally interpreted in models based on the point-wise time-flow frames $\mathcal{F} = \langle T, < \rangle$, so in a poit-wise semantics. If we also admit two additional quantifiers: A ('along all paths)' and E (read: 'there exists at least one path such that'), we obtain a new system – commonly called *Computation Tree Logic* (CTL) [59] as a branching-time logic.

The construction way of these systems demonstrates that these systems are rather suitable for describing situations, somehow, temporally 'open' in a past or in a future.

C. Halpern-Shoham logic

The logic of Halpern-Shoham (HS) – introduced in [60] forms a multi-modal system suitable to represent the 13 well-known Allen's relations between intervals [36], and constitutes a concurrent approach to temporal reasoning w. r. t. the Computational Tree Logic (CTL) or the Linear Temporal Logic LTL – the more traditional and pointwise approaches. More, precisely, HS forms a modal representation of temporal relations: **a**fter (or **m**eets, **l**ater), **b**egins (or **s**tart), **d**uring, **e**nd and **o**verlap and they are rendered in HS by corresponding modal operators: $\langle A \rangle$ for **a**fter, $\langle B \rangle$ for **b**egins, $\langle D \rangle$ for **d**uring, etc. The full syntax of HS-entities ϕ is defined by:

$$\phi := p | \neg\phi | \phi \wedge \phi | \langle X \rangle | \langle \bar{X} \rangle \,, \tag{1}$$

where p is a propositional variable and $\langle \bar{X} \rangle$ denotes a modal operator for the inverse relation. Such logic is more suitable for representing some temporally 'closed' events, actions and processes, their mutual temporal relations or time lags.

Table V presents the overview of these logics in terms of time pattern representation for process models.

V. SUMMARY

Process models can specify various aspects of business processes, among them the temporal ones. In this paper, we present the existing solutions for describing time aspects of process models. Although there are many notations for modeling processes, the main focus is on the BPMN notation. We provide several examples of representing time patterns in BPMN as well as discuss temporal issues with temporal logics for such specifications. Thus, the original contribution of this paper is threefold – we present:

1) the overview of business process modeling notations focusing on time-related elements,
2) the process models in BPMN notation which fulfil the Allen's relations,
3) the assessment of temporal logics in terms of using them for representing time patterns in process models.

The research presented in this paper is a proposal for further studies related to time issues in BPMN process models. Our future work will focus on practical assessment of process models with the time related logic specification and the possibility of validation and verification of such models [61] or compliance checking [62], especially with the existing tools which uses temporal knowledge [63], [64]. One of the possible directions is related to integration of timed process models with timed rules models for complex reasoning on contex data [65], [66]. Important issue is also practical design of timed models of processes and rules [67]. As we focus on analysis of a single process model, additional research for analyzing process models of processes changing over time can be an important issue [68].

Table V
SUITABILITY OF REPRESENTING TIME PATTERNS [38] USING TEMPORAL LOGICS AND ALLEN'S ALGEBRA

Pattern Type/ Logic type	**LTL/CTL**	**Allen's algebra**	**Halpern-Shoham logic**
Patter 1 (Duration and Time Lags)	◑	●	●
Pattern 2 (Duration)	◑	●	●
Pattern 3 (Time Lags between arbitrary Events)	◑	◑	◑
Pattern 4 (Fixed Data Element)	●	❍	❍
Pattern 5 (Schedule Restricted Element)	❍	●	●
Pattern 6 (Time Based Restrictions)	◑	❍	❍
Pattern 7 (Validity Period)	◑	●	●
Pattern 8 (Variability)	◑	●	●
Pattern 9 (Cyclic Elements)	◑	❍	❍
Pattern 10 (Periodicity)	❍	❍	❍

REFERENCES

[1] M. Weske, *Business Process Management: Concepts, Languages, Architectures 2nd Edition*. Springer, 2012.

[2] A. Lindsay, D. Dawns, and K. Lunn, "Business processes – attempts to find a definition," *Information and Software Technology*, vol. 45, no. 15, pp. 1015–1019, Dec 2003, elsevier.

[3] F. Hunka and R. Belunek, "Transaction based business process modeling," in *Computer Science and Information Systems (FedCSIS), 2015 Federated Conference on*. IEEE, 2015, pp. 1397–1402.

[4] M. Mach-Król, "Perspectives of using temporal logics for knowledge management," in *2012 Federated Conference on Computer Science and Information Systems (FedCSIS)*, 2012.

[5] M. Mach-Król and K. Michalik, "Selected aspects of temporal knowledge engineering," in *Computer Science and Information Systems (FedCSIS), 2014 Federated Conference on*. IEEE, 2014, pp. 1091–1096.

[6] M. Owoc and K. Marciniak, "Knowledge management as foundation of smart university," in *Computer Science and Information Systems (FedCSIS), 2013 Federated Conference on*. IEEE, 2013, pp. 1267–1272.

[7] W. M. P. Aalst, H. T. Beer, and B. F. Dongen, "Process mining and verification of properties: An approach based on temporal logic," in *On the Move to Meaningful Internet Systems 2005: CoopIS, DOA, and ODBASE: OTM Confederated International Conferences, CoopIS, DOA, and ODBASE 2005, Agia Napa, Cyprus, October 31 - November 4, 2005, Proceedings, Part I*, R. Meersman and Z. Tari, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 130–147.

[8] Y. Qiang, M. Valcke, P. De Maeyer, and N. Van de Weghe, "Representing time intervals in a two-dimensional space: an empirical study," *Journal of Visual Languages & Computing*, vol. 25, no. 4, pp. 466–480, 2014.

[9] W. M. P. van der Aalst, "Business process management: A comprehensive survey," *ISRN Software Engineering*, vol. 2013, 2013.

[10] A. W. Scheer, *Aris: Business Process Modeling*, 3rd ed. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2000.

[11] M. Rosemann and W. M. P. van der Aalst, "A configurable reference modelling language," *Information Systems*, vol. 32, no. 1, pp. 1–23, 2007.

[12] C. Menzel and R. J. Mayer, "The IDEF family of languages," in *Handbook on Architectures of Information Systems*, ser. International Handbooks on Information Systems, P. Bernus, K. Mertins, and G. Schmidt, Eds. Springer Berlin Heidelberg, 1998, ch. 10, pp. 209–241.

[13] C. Badiča, A. Badiča, and V. Litoiu, "A new formal IDEF-based modelling of business processes," in *Proc. of the 1st Balkan Conference in Informatics, Thessaloniki, Greece*, 2003, pp. 535–549.

[14] C. P. Menzel, R. J. Mayer, and D. D. Edwards, "IDEF3 formalization report," DTIC Document, Texas TX USA, Tech. Rep. KBSL-89-1007, October 1991.

[15] C. Badiča and C. Fox, "Hybrid DEF0/IDEF3 modelling of business processes: Syntax, semantics and expressiveness," in *Romanian-Austrian Workshop on Computer-Aided Verification of Information Systems: A Practical Industry-Oriented Approach. Timisoara, Romania*, 2004, pp. 20–22.

[16] R. J. Mayer, C. Menzel, M. Painter, P. S. deWitte, T. Blinn, and B. Perakath, "Information integration for concurrent engineering (IICE) IDEF3 process description capture method report," Knowledge Based Systems, Inc., Tech. Rep. AL-TR-1995-XXXX, 1995.

[17] OMG, "Unified Modeling Language (OMG UML) version 2.2. superstructure," Object Management Group, Tech. Rep. formal/2009-02-02, February 2009.

[18] J. Hunt, *Guide to the Unified Process featuring UML, Java and Design Patterns*. Springer, 2003.

[19] M. Fowler, *UML Distilled: A Brief Guide to the Standard Object Modeling Language*, 3rd ed. Addison-Wesley Professional, 2003.

[20] D. Pilone and N. Pitman, *UML 2.0 in a Nutshell*. O'Reilly, 2005.

[21] T. Murata, "Petri nets: Properties, analysis and applications," *Proceedings of the IEEE*, vol. 77, no. 4, pp. 541–580, 1989.

[22] M. Szpyrka, *Sieci Petriego w modelowaniu i analizie systemów współbieżnych*. Warszawa: Wydawnictwa Naukowo-Techniczne, 2008.

[23] B. Berthomieu and M. Diaz, "Modeling and verification of time dependent systems using time petri nets," *IEEE transactions on software engineering*, vol. 17, no. 3, p. 259, 1991.

[24] L. Popova-Zeugmann, *Time and Petri Nets*. Springer, 2013.

[25] A. H. M. ter Hofstede, W. M. P. van der Aalst, M. Adams, and N. Russell, Eds., *Modern Business Process Automation: YAWL and its Support Environment*. Springer, 2010.

[26] W. M. P. van der Aalst and A. H. M. ter Hofstede, "YAWL: Yet another workflow language," *Information Systems*, vol. 30, no. 4, pp. 245–275, 2005.

[27] W. M. P. van der Aalst and A. H. M. ter Hofstede, "Workflow patterns: On the expressive power of (petri-net-based) workflow languages," in *Proceedings of the Fourth International Workshop on Practical Use of Coloured Petri Nets and the CPN Tools, Aarhus, Denmark, August 28-30, 2002*, K. Jensen, Ed., University of Aarhus. DAIMI PB-560, Aug 2002, pp. 1–20.

[28] W. M. P. van der Aalst, "Verification of workflow nets," in *Application and Theory of Petri Nets 1997*, ser. Lecture Notes in Computer Science, P. Azema and G. Balbo, Eds. Springer Berlin Heidelberg, 1997, vol. 1248, pp. 407–426.

[29] OMG, "Business Process Model and Notation (BPMN): Version 2.0 specification," Object Management Group, Tech. Rep. formal/2011-01-03, January 2011.

[30] B. List and B. Korherr, "An evaluation of conceptual business process modelling languages," in *Proceedings of the 2006 ACM symposium on Applied computing*. ACM, 2006, pp. 1532–1539.

[31] W. M. P. van der Aalst, L. Aldred, M. Dumas, and A. H. M. ter Hofstede, "Design and implementation of the YAWL system," in *Advanced Information Systems Engineering*, ser. Lecture Notes in Computer Science, A. Persson and J. Stirna, Eds. Springer Berlin Heidelberg, 2004, vol. 3084, pp. 142–159.

[32] J. Recker, M. Rosemann, M. Indulska, and P. F. Green, "Business process modeling – a comparative analysis," *Journal of the Association for Information Systems*, vol. 10, no. 4, 2009.

[33] F.-R. Lin, M.-C. Yang, and Y.-H. Pai, "A generic structure for business process modeling," *Business Process Management Journal*, vol. 8, no. 1, pp. 19–41, 2002.

[34] W. Wang, H. Ding, J. Dong, and C. Ren, "A comparison of business process modeling methods," in *Proceedings of the IEEE International Conference on Service Operations and Logistics, and Informatics, 2006. SOLI '06*, 2006, pp. 1136–1141.

[35] O. Svatoš, "Conceptual process modeling language: Regulative approach," in *Proceedings of the 9th Undergraduate and Graduate Students eConf. and 14th Business & Government Executive Meeting on Innovative Cross-border eRegion, Univ. of Maribor*, 2007.

[36] J. Allen, "Maintaining knowledge about temporal intervals," in *Communications of ACM*, 26(11)1983, pp. 832–843.

[37] K. Kluza and K. Kaczor, "Overview of BPMN model equivalences: towards normalization of BPMN diagrams," in *8th Workshop on Knowledge Engineering and Software Engineering (KESE2012) at the at the biennial European Conference on Artificial Intelligence (ECAI 2012): August 28, 2012, Montpellier, France*, J. Canadas, G. J. Nalepa, and J. Baumeister, Eds., 2012, pp. 38–45. [Online]. Available: http://ceur-ws.org/Vol-949/

[38] A. Lanz, B. Weber, and M. Reichert, "Time patterns for process-aware information systems," *Requirements Engineering*, vol. 19, no. 2, pp. 113–141, 2012.

[39] D. Gagne and A. Trudel, "Time-bpmn," in *2009 IEEE Conference on Commerce and Enterprise Computing*, July 2009, pp. 361–367.

[40] S. Cheikhrouhou, S. Kallel, N. Guermouche, and M. Jmaiel, "A survey on time-aware business process modeling," in *International Conference on Enterprise Information Systems (ICEIS)*, July 2013, p. 10p.

[41] S. Cheikhrouhou, S. Kallel, N. Guermouche, and M. Jmaiel, "The temporal perspective in business process modeling: a survey and research challenges," *Service Oriented Computing and Applications*, vol. 9, no. 1, pp. 75–85, 2015.

[42] A. Lanz, B. Weber, and M. Reichert, "Workflow time patterns for process-aware information systems," in *Enterprise, Business-Process and Information Systems Modeling: 11th International Workshop, BPMDS 2010, and 15th International Conference, EMMSAD 2010, held at CAiSE 2010, Hammamet, Tunisia, June 7-8, 2010. Proceedings*. Berlin, Heidelberg: Springer, 2010, pp. 94–107.

[43] A. Lanz, U. Kreher, M. Reichert, and P. Dadam, "Enabling process support for advanced applications with the Aristaflow BPM suite," in *Proceedings of the Business Process Management 2010 Demonstration Track*, September 2010, no. 615.

[44] I. Barba, A. Lanz, B. Weber, M. Reichert, and C. del Valle, "Optimized time management for declarative workflows," in *13th BPMDS'12 working conference, Lecture Notes in Business Information Processing*. Berlin, Heidelberg: Springer, 2012, pp. 195–210.

[45] J. Xie, Y. Tang, Q. He, and N. Tang, "Research of temporal workflow process and resource modeling," in *Proceedings of the 9th international conference on computer supported cooperative work in design*, 2005, vol. 1, pp. 530–534.

[46] M. Szpyrka, G. J. Nalepa, A. Ligęza, and K. Kluza, "Proposal of formal verification of selected BPMN models with Alvis modeling language," in *Intelligent Distributed Computing V. Proceedings of the 5th International Symposium on Intelligent Distributed Computing – IDC 2011, Delft, the Netherlands – October 2011*, ser. Studies in Computational Intelligence, F. M. Brazier, K. Nieuwenhuis, G. Pavlin, M. Warnier, and C. Badica, Eds. Springer-Verlag, 2011, vol. 382, pp. 249–255. [Online]. Available: http://www.springerlink.com/content/m181144037q67271/

[47] A. Ligęza, K. Kluza, and T. Potempa, "Ai approach to formal analysis of bpmn models. towards a logical model for bpmn diagrams," in *Proceedings of the Federated Conference on Computer Science and Information Systems – FedCSIS 2012, Wroclaw, Poland, 9-12 September 2012*, M. Ganzha, L. A. Maciaszek, and M. Paprzycki, Eds., 2012, pp. 931–934. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6354394

[48] M. Szpyrka, P. Matyasik, and M. Wypych, "Alvis language with time dependence," in *Computer Science and Information Systems (FedCSIS), 2013 Federated Conference on*. IEEE, 2013, pp. 1565–1570.

[49] Y. Du, P. Xiong, Y. Fan, and X. Li, "Dynamic checking and solution to temporal violations in concurrent workflow processes," *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 41, no. 6, pp. 1166–1181, 2011.

[50] R. Klimek and P. Szwed, "Verification of archimate process specifications based on deductive temporal reasoning," in *Computer Science and Information Systems (FedCSIS), 2013 Federated Conference on*. IEEE, 2013, pp. 1109–1116.

[51] N. Guermouche, "Etude des Interactions Temporisées dans la Composition de Services Web," PhD Thesis, Université Henri Poincaré - Nancy, Jun. 2010.

[52] N. Guermouche and S. D. Zilio, "Towards timed requirement verification for service choreographies," in *2012 8th International Conference on Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom)*, 2012, pp. 117–126.

[53] S. Kallel, A. Charfi, T. Dinkelaker, M. Mezini, and M. Jmaiel, "Specifying and monitoring temporal properties in web services compositions," in *7th IEEE European Conference on Web Services, ECOWS '09*, 2009, pp. 148–157.

[54] K. Watahiki, F. Ishikawa, and K. Hiraishi, "Formal verification of business processes with temporal and resource constraints," in *Systems, Man, and Cybernetics (SMC), 2011 IEEE International Conference on*, 2011, pp. 1173–1180.

[55] M. Makni, S. Tata, M. Yeddes, and N. Ben Hadj-Alouane, "Satisfaction and coherence of deadline constraints in inter-organizational workflows," in *On the Move to Meaningful Internet Systems: OTM 2010: Confederated International Conferences: CoopIS, IS, DOA and ODBASE, Hersonissos, Crete, Greece, October 25-29, 2010, Proceedings, Part I*. Berlin, Heidelberg: Springer, 2010, pp. 523–539.

[56] P. Y. H. Wong and J. Gibbons, "A relative timed semantics for BPMN," *Electronic Notes in Theoretical Computer Science*, vol. 229, no. 2, pp. 59–75, 2009.

[57] K. Jobczyk and A. Ligeza, "Systems of temporal logic for a use of engineering. toward a more practical approach," in *Intelligent Systems for Computer Modelling*. Springer, 2016, pp. 147–157.

[58] K. Jobczyk and A. Ligeza, "Why systems of temporal logic are sometimes (un) useful?" in *International Conference on Artificial Intelligence and Soft Computing*. Springer, 2016, pp. 306–316.

[59] A. Pnueli, "The temporal logic of programs," *Proceedings of the 18th Annual Symposium on Foundation of Computer Science*, 1977:46-57.

[60] J. Halpern and Y. Shoham, "A propositional modal logic of time intervals," *Journal of the ACM*, vol. 38, pp. 935–962, 1991.

[61] M. Mach-Krol and K. Michalik, "Validation and verification of temporal knowledge as an important aspect of implementing a temporal knowledge base system supporting organizational creativity," in *Computer Science and Information Systems (FedCSIS), 2015 Federated Conference on*. IEEE, 2015, pp. 1315–1320.

[62] A. Awad, G. Decker, and M. Weske, "Efficient compliance checking using bpmn-q and temporal logic," in *Business Process Management: 6th International Conference, BPM 2008, Milan, Italy, September 2-4, 2008. Proceedings*, M. Dumas, M. Reichert, and M.-C. Shan, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 326–341.

[63] M. Mach-Król, "Tools for building a temporal knowledge base system supporting organizational creativity," *Procedia Computer Science*, vol. 65, pp. 1031–1037, 2015.

[64] E. Kucharska, K. Grobler-Dębska, J. Gracel, and M. Jagodziński, "Idea of impact of erp-aps-mes systems integration on the effectiveness of decision making process in manufacturing companies," in *International Conference: Beyond Databases, Architectures and Structures*. Springer, 2015, pp. 551–564.

[65] G. J. Nalepa and S. Bobek, "Rule-based solution for context-aware reasoning on mobile devices," *Computer Science and Information Systems*, vol. 11, no. 1, pp. 171–193, 2014.

[66] S. Bobek, M. Slazynski, and G. J. Nalepa, "Capturing dynamics of mobile context-aware systems with rules and statistical analysis of historical data," in *Artificial Intelligence and Soft Computing*, ser. Lecture Notes in Computer Science, L. Rutkowski, M. Korytkowski, R. Scherer, R. Tadeusiewicz, L. A. Zadeh, and J. M. Zurada, Eds., vol. 9120. Springer International Publishing, 2015, pp. 578–590. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-19369-4_51

[67] G. J. Nalepa and A. Ligęza, *Software engineering: evolution and emerging technologies*, ser. Frontiers in Artificial Intelligence and Applications. Amsterdam: IOS Press, 2005, vol. 130, ch. Conceptual modelling and automated implementation of rule-based systems, pp. 330–340.

[68] D. Luengo and M. Sepúlveda, "Applying clustering in process mining to find different versions of a business process that changes over time," in *Business Process Management Workshops: BPM 2011 International Workshops, Clermont-Ferrand, France, August 29, 2011, Revised Selected Papers, Part I*, F. Daniel, K. Barkaoui, and S. Dustdar, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 153–158.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1125–1128
DOI: 10.15439/2016F33
ACSIS, Vol. 8. ISSN 2300-5963

Applying simulations: On the importance of the simulation performance.

Bernd Pfitzinger*, Tommy Baumann‡§, Dragan Macos†, Thomas Jestdt*
*Toll Collect GmbH, Linkstrae 4, 10785 Berlin, Germany. {bernd.pfitzinger|thomas.jestaedt}@toll-collect.de
‡Andato GmbH & Co. KG, Ehrenbergstrae 11, 98693 Ilmenau, Germany. tommy.baumann@andato.com
§Hochschule Aalen – Technik und Wirtschaft, Beethovenstrae 1, D73430 Aalen.
†Beuth Hochschule fr Technik Berlin, Luxemburger Str. 10, 13353 Berlin, Germany. dmacos@beuth-hochschule.de

***Abstract*—Creating new software or software-intensive systems is still a challenge and far removed from a traditional engineering domain. The increasing size of software deployed in typical systems and the emergence of very large highly distributed systems necessitates additional techniques to assure the systems quality. Using the example of the German automatic toll system we briefly outline a simulation driven development approach: Using simulation models starting with the very early design stages to verify and validate the overall dynamic system behavior throughout the whole development process. In practice the approach depends particularly on the performance of the simulation model: Many simulation runs are necessary while exploring the solution space of a proposed change or while calibrating and optimizing parameters of the simulation models. Starting with an existing model of the German automatic toll system we look at two different possibilities for parallelization – parallelized optimization and the partial transformation of the simulation model to a parallelized implementation.**

I. Introduction

MOST of the critical system design problems occur in the early design stages while specialists are specifying the system under a high degree of uncertainty. Many studies show that the probability of critical problems due to poor design decisions is very high in the specification phase [1]. The root cause is that either text-based or non-executable, model-based specifications are utilized: We build models of complex systems because we cannot comprehend any such system in its entirety [2]. Such specifications cannot be validated at the system level where the architecture and performance is determined as an emergent property.

The typical system development process will of course try to mitigate the consequences e.g. by shortening the step from the design to the deployment stage. But in distributed systems – "one in which the failure of a computer you didn't even know existed can render your own computer unusable" [3] – the properties of the whole system emerge only when all subsystems are integrated into a complete system.

Hence we propose to accompany the system development process with an executable specification of the whole system: At any stage during the system development process we implement the current level of detail as a simulation model. In that way the specification becomes executable (i.e. has sufficient detail to be executable).

In this article we briefly touch upon the concept of simulation driven development (SDD) – the starting point of an investigation into the performance of simulation models. When the intention is to base the system development process on simulation models the simulation results should be valid and readily available at any time. At least for complex systems both apsects necessitate a high simulation performance.

In particular the article applies two different ways of parallelizing simulations: The trivial parallelization of a genetic algorithm is of use when the simulation model is used either to explore the solution space in search of an optimal configuration or when parameters need to be fitted to data observed in the real world, e.g. modeling the user interaction. The non-trivial parallelization concerns the simulation model itself where – depending on the specific application – parts might be run independently. Starting with a simple benchmark model we apply the parallelization to an existing simulation model of the German automatic toll system.

The article is split into three parts: The first section explains the concept of simulation driven development in more detail (see section II) whereas the second and third section cover the simulation performance as one prerequisite for basing the software development process on simulation models: Parameter optimization using a genetic algorithm and the partial parallelization of a simulation model (sections III and IV). All parts have the same use case in common (section III): Applying the approach in the context of the German automatic toll system – a large scale liability-critical distributed system and a typical example of an electronic tolling system based on global navigation satellite systems (GNSS) [4].

II. Simulation Driven Development

SDD is characterized by applying modeling and simulation technologies during the whole product lifecycle: An executable model exists at any time encapsulating the current knowledge of the system. The benefit of SDD lies both in the early design stages (when most of the important design decisions are made) and the ability to verify and validate the system at any time through executable models.

A. Design approach

In general a system specification defines the functional and non-functional properties of a system in a formal, consistent, and self-contained manner to enable processing. Functional properties define the tasks of the system including information

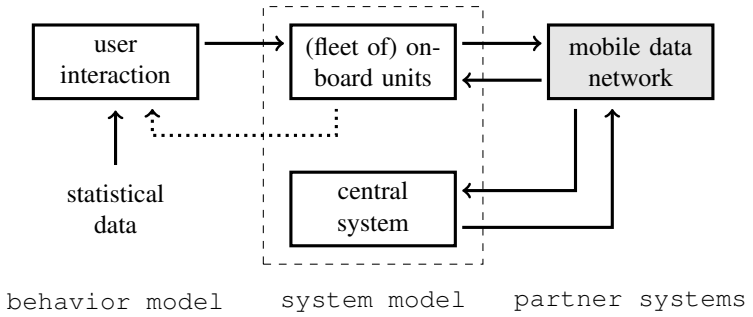


Figure 1. The simulation model of the German automatic toll system encompasses the the user interaction (left), the operators' technical systems (center) and systems under the responsibility of partners (right).

processing in relation to data, operation and the systems behavior.

Non-functional properties are more difficult to pin down there is not even a consensus on the term and its use [5]. They are used to describe the circumstances necessary to render the required functionality, e.g. the performance requirements, quality properties and constraints. In a sense the non-functional requirements become either constraints (to the system development or the system operations) or emergent properties (in the worst case emerging only at the system level when the user interaction is taken into account). SDD allows modeling both the constraints and the whole system including the user interaction and feedback loops.

Unfortunately, the specification phase is far removed from the final acceptance testing and the transition to operations. Many process models exist to close the gap, either by introducing more tests and documentation along the way or by shortening the development cycle (e.g. shifting requirements to the next cycle). To illustrate the system engineering process we take the "V-model" [6], a highly adaptable meta-model commissioned by the German federal government. At the core of this model is the well-known waterfall model dating back to 1970 with the design and development phases placed on the left and the test phases on the right – forming the letter "V". While it proceeds through the typical phases it differentiates the testing phase so that every design or development phase also sets up its own test phase.

Executable system specifications allow the validation, analysis, evaluation, and optimization of complex distributed systems even before the first line of source code is written – including dynamic interactions. Bottlenecks and resource shortages owing to dynamic coupling effects can be captured and resolved without running or implementing the real system. On a meta-level the executable model can include the system development process as well.

B. Applying SDD

In its current implementation we used SDD in the specification phase of two upcoming changes to the application-level protocols governing the interaction between the on-board-unit (OBU) and the central systems of the German automatic toll system in essence switching from an open-loop controller (the OBU decides in large parts independently if and when to connect to the central system) to a closed-loop controller (where the control resides at any time with the central system). The impact of SDD on the specification phase was twofold: Performing simulations of the whole integrated system at a scale of 1:1 allows predicting the system load for the chosen system architecture (and in combination with a cost model the estimation of operational costs) in effect operational risks become visible in the very first step of the system development process.

So far, researching the impact of architectural choices (or the systems parametrization) is a manual task: The simulation model needs to be altered (or at least re-configured) and the simulation run performed and analyzed. Using a fitness-function to grade the simulation results (e.g. the correlation with the real-world update rates observed for a fleet-wide update) is the first step to automate the design process: An additional optimization algorithm is able to change the parametrization or even the systems architecture automatically and to search for the overall optimal solution.

An intangible benefit of SDD in our example is the emphasis on the operational performance of any change already during the specification phase: Using the simulation model the consequences to the day-to-day performance is quantified and easily comparable to todays (or the intended future) performance.

III. Simulation model of the German toll system

In practice we introduced SDD to the system development process of Toll Collect GmbH (e.g. [7] and references therein) the operator of the German toll system for heavy goods vehicles (HGVs).

The automatic toll system consists of four major parts (figure 1): It uses OBUs to automatically collect the toll charges due. Most of the time the OBU is running independently and rarely connects to the central system via a mobile data network connection to upload the tolls collected and possibly download updates to its software, geo and tariff data. Using this architecture the reliability of the system depends heavily on the OBU its hardware, software and operational data as well as the user interaction.

In principle the system behavior could affect the user behavior (dotted line in figure 1): E.g. when the OBU signals 'out-of-order' the user could be tempted to quickly power-cycle the OBU. Particularly in the case of outages, e.g. when the central system cannot be reached by the OBU, the OBU will reach a point where it needs to signal the user a technical problem. It is therefore conceivable that the user behavior changes – either as intended by the system design (the user switches to the manual toll system, e.g. via internet booking) or by power cycling the OBU and thereby potentially triggering additional system load.

Looking at figure 1 it is interesting to note the first indications of a software-intensive system-of-systems: The system depends not only on the user interaction – it also includes partner systems that are operated independently (e.g. the mobile data network) and whose inner working is not accessible.

The model of the toll system depends on the external stimulus of the user interaction. With an emphasis on the fleet-wide propagation of updates we include only the temporal behavior: The points in time when an HGV is powered on (or off) and when a toll event is created.

Regarding the fleet-wide updates we use a genetic algorithm to adapt the simulation results to the data observed in the real-world system (for details and results see [8]). Parameter optimization requires the ability to compute a large number of results for different sets of parameters. In our case we used two sets of 16 probabilities (i.e. the probability for a given HGV of being active N out of 16 weeks, once for German HGVs and again for foreign HGVs). Even using a running at a scale of less than 1:1000 a single simulation run takes almost one minute to complete.

IV. Parallel DES models

For many purposes scaling is the appropriate solution – yet close to the specification limit of a system or once processes become non-linear, scaling is no longer an option. Simulation performance becomes important.

A. Domain independent parallel DES models

Many approaches exist to parallelize existing serial mode simulation models [9]. In our case, the space-parallel domain decomposition accelerates a simulation run by executing parts of the model in parallel. This approach is applicable to any model and shows the greatest potential in offering scalable performance for complex models [10], [11].

An optimistic approach to parallelization is to execute components even if – at a later time – it becomes known, that the execution was unnecessary or violated a causality constraint ("time warp", [12]). As long as excess computing power is available the consequence is only that causality errors need to be rolled back, i.e. each component has to be enhanced by an additional rollback block and from time to time rollbacks will occur. In a real-world application the expected speed-up is limited by those parts of a domain-specific simulation model that are intrinsically serial.

B. Domain specific parallel DES models

"Time warp" or optimistic parallelization has the advantage of being domain-independent – it is a feature offered by the simulation toolset as well as a programming paradigm. However, "artifical" rollback and commit steps need to be implemented and verified by additional testing. This effort could also be spent on an explicit, domain-specific parallelization of the simulation model itself.

In the example of the German automatic toll system (see figure 1) parallelism is inherent in many parts of the model (e.g. in the user behavior, the OBU fleet). Since most of the processing occurring in the simulation model is connected to the OBU fleet a sizable degree of parallelism could be achieved.

Simulation performance quickly becomes a bottleneck when the DES model is coupled to a real-world system, i.e. in a software-in-the-loop scenario. An example could be a server of the central system being subjected to the TCP/IP traffic generated by the OBU fleet.

Here the artificial concept of time – time is expressed in terms of events occurring and jumps immediately to the next event present in the "future event list" – needs to be coupled to the time passing in the real-world (system). Figure 2 depicts both concepts of time: In the real-world system time progresses continuously (indicated by the thick arrow from left to right) – aptly summarized as the wall-clock time. In contrast the DES tool considers time only in the presence of events: Each event generated in the simulation run carries a (future) time of execution and is accordingly put into the future vents list (FEL) for later processing. The current time in the DES model is therefore given by the pointer to the next event up for processing. Whenever multiple events occur at the same time a serial-mode simulation takes the events for processing from the future event list as if it were a queue. I. e. simultaneous events are processed sequentially while the clock is stopped.

When a real-world system is interfaced to the simulation model both possibilities break down: Even during times without events time passes at the same rate and sequential processing of simultaneous events will insert artificial delays, potentially disturbing the simulation results with two noteworthy consequences:

- Any speed-up of a DES simulation is negated once the real-world system is interfaced, the simulation will proceed (at most) at wall-clock time.
- The DES simulation model introduces artificial delays, when taking outgoing events from the FEL or inserting incoming events into the FEL – possibly to the degree of invalidating the simulation resuls by violating the 'real-time' constraint.

V. Summary

As in the test-driven development (TDD) case, testing is not the aim of the SDD rather the "driven [...] focuses on how TDD leads analysis, design, and programming decisions" [13]. In that sense, SDD tries to put the design to

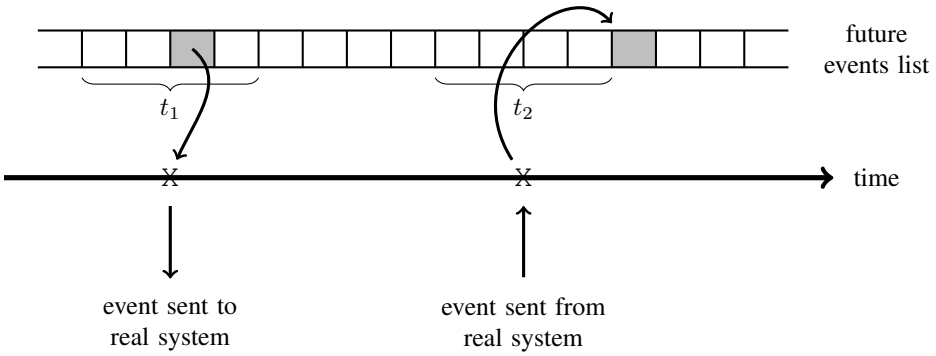


Figure 2. The concepts of time differ between the continuous time line in the real world and the list of (future) events at discrete points in time. Interfacing a real-world system will couple the two concepts.

the ultimate test-case – the real-world operational context. The simulation model – an executable specification of the existing real-world system – is the starting point to focus any software development on the operational consequences. These consequences might be of a purely technical nature, e.g. the system architecture and performance, or include non-functional requirements and business or financial aspects. In particular these challenges dominate environments that are rich in legacy systems, where the on-going development is largely faced with integration issues. SDD addresses integration of systems as a cross-cutting concern by providing the software developer (or requirements engineer) with an executable copy of the real-world system.

A prerequisite for applying SDD is the performance of the simulation model. We summarized two areas where performance matters: Exploring the solution space needs many simulation runs, albeit possibly at a small scale. Interfacing a simulation model with a real-world system is the final challenge – mingling simulated discrete and continuous real time.

REFERENCES

[1] A. Aurum and C. Wohlin, in *Engineering and managing software requirements*, A. Aurum and C. Wohlin, Eds. Berlin: Springer, 2005, ch. Requirements Engineering: Setting the Context, pp. 1–15, ISBN: 978-3-540-28244-0. DOI: 10.1007/3-540-28244-0_1.

[2] ISO, *ISO/IEC 19505-2:2012 Information technology – Object Management Group Unified Modeling Language (OMG UML), Superstructure*. Berlin: Beuth Verlag, 2012.

[3] L. Lamport, *Distribution*, e-mail message, [accessed 16-Nov-2014], May 1987. [Online]. Available: http://research.microsoft.com/en-us/um/people/lamport/pubs/distributed-system.txt.

[4] J. Numrich, S. Ruja, and S. Vo, "Global Navigation Satellite System based tolling: State-of-the-art", *Netnomics: Economic research and electronic networking*, vol. 13, no. 2, pp. 93–123, Jul. 2012. DOI: 10.1007/s11066-013-9073-9.

[5] M. Glinz, "On non-functional requirements", in *15th IEEE international requirements engineering conference*, (Delhi), Oct. 2007, pp. 21–26, ISBN: 978-0-7695-2935-6. DOI: 10.1109/RE.2007.45.

[6] Verein zur Weiterentwicklung des V-Modell XT e.V. (Weit e.V.), *V-Modell XT version 2.0*, [accessed 21-Jan-16], 2006. [Online]. Available: http://www.v-modell-xt.de/.

[7] B. Pfitzinger, T. Baumann, D. Macos, and T. Jestdt, "Modeling regional reliability of 2G, 3G, and 4G mobile data networks and its effect on the German automatic tolling system", in *2015 48th hawaii international conference on system sciences (hicss)*, Jan. 2015, pp. 5439–5445. DOI: 10.1109/HICSS.2015.640.

[8] B. Pfitzinger, T. Baumann, D. Macos, and T. Jestdt, "Using parameter optimization to calibrate a model of user interaction", in *Proceedings of the 2014 federated conference on computer science and information systems*, M. P. M. Ganzha L. Maciaszek, Ed., ser. Annals of Computer Science and Information Systems, vol. 2, IEEE, Sep. 2014, pp. 1111–1116, ISBN: 978-83-60810-58-3. DOI: 10.15439/2014F123.

[9] V.-Y. Vee and W.-J. Hsu, "Parallel discrete event simulation: A survey", Tech. Rep., 1999. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.41.7706.

[10] R. Righter and J. Walrand, "Distributed simulation of discrete event systems", *Proceedings of the ieee*, vol. 77, no. 1, pp. 99–113, Jan. 1989, ISSN: 0018-9219. DOI: 10.1109/5.21073.

[11] A. J. Wing, in *Advances in parallel algorithms*, L. Kronsj and D. Shumsheruddin, Eds. New York: John Wiley & Sons, Inc., 1992, ch. Discrete Event Simulation in Parallel, pp. 179–226, ISBN: 0-470-21907-6.

[12] D. Jefferson and H. Sowizral, *Fast concurrent simulation using the time warp mechanism: Part i, local control*. Santa Monica, CA: Rand Corporation, 1982.

[13] D. Janzen and H. Saiedian, "Test-Driven Development: Concepts, taxonomy, and future direction", *Computer*, vol. 38, no. 9, pp. 43–50, Sep. 2005, ISSN: 0018-9162. DOI: 10.1109/MC.2005.314.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1129–1132
DOI: 10.15439/2016F338
ACSIS, Vol. 8. ISSN 2300-5963

Effects of the transformation of company computer system on cloud computing services – a change in company management

Milena Tvrdíková
VŠB – TUO,
Faculty of Economics
Sokolská 33, 701021 Ostrava
Czech Republic
Email: milena.tvrdikova@vsb.cz

***Abstract*— paper deals with the transition of companies to information systems operated using cloud computing services. The influence on the competitiveness of company and its organizational support is analysed. The developmental trends and their impact on company management are also described. The benefits and drawbacks from the transition to flexible ICT architecture are emphasized. Theoretical framework and literature support the fact that exploitation of cloud computing services is an inevitable component of the information strategy of European countries in the support of companies' competitiveness. Based on the collected information and own experiences with using the cloud computing, the recommendations for the effective transition to the application of cloud computing services in a company are proposed.**

I. Introduction

Managers keep looking for new means to improve management efficiency. Intensive use of modern IT technologies has become an essential part of the management process. Managers are aware of that and want to exploit the potential of new ICTs in the management of companies and their environment.

Thanks to today's modern communication technologies and social networks, each day we create huge amounts of data, both in the corporate environment and beyond.

The impressive boom in information technology, which started in the late 1990s, brought about a rapid acceleration of our lifestyles as well as of the overall approach to entrepreneurship and its support by ICTs.

The current trends in ICT having an impact on the efficiency and the type of organizational structure include, in particular:
the development of dynamic network services,
increase in the performance and capacity of data centres,
business through mobile technologies,
increasing demands on IS security,
processing increasing volumes of data
the development of Cloud Computing (CC). [1]

The economic crisis, development of new technologies and rapid globalization, the growth of the importance of access to relevant information and desire for mobility, act as catalysts for the global population.

II. Transition to the use of ICT as a shared service – Cloud Computing

Managers require a comprehensive information system containing the necessary functions, at different times and in the necessary scope. The aim is to ensure the necessary flexibility in management. This is made possible by Cloud Computing services. The CC services increase company flexibility and have a positive impact on its production and competitiveness.

CC has no uniform definition; each definition depends on the perspective of its author.

„A cloud is defined as the combination of the infrastructure of a data centre with the ability to provision hardware and software.“, says Sosinsky. [2]

Gartner defines cloud computing as "a style of computing in which scalable and elastic IT-enabled capabilities are delivered as a service using Internet technologies".

Another definition of CC from a different perspective says that "*Cloud Computing is essentially a concept that allows you to access applications that are actually located elsewhere than on a local computer or device connected to the Internet, most commonly in a remote data centre.*" [3]

In summary, CC is simply the "*approach to the use of computer technology, which is based on providing shared computing resources and their use in the form of a service.*"

The author of this paper relies on the definition of the US National Institute of Standards and Technology, which defines the CC as a "*model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort or service provider interaction.*" [4]

III. Definition of the problem

This article aims to identify possible pitfalls of the transition of the company IS to a flexible ICT architecture

This work was not supported by any organization

and operation of IS in the form of CC services and to draft procedures and recommendations to successfully manage this transition.

Due to the long experience of the author with CC services, the paper discusses benefits and risks of operating the IS in the form of CC services. In the recommended transformation of the company IS into the CC, the author also draws on her personal experience.

Innovative, so far untested information and communication technologies can significantly improve the performance of a company IS on the one hand, but if incorrectly implemented, they can cause fatal damage in situations beyond the control of the company. [5]

IV. Transformation of a company's information system into Cloud Computing

It needs gradually and conceptually change the overall IT architecture. The investment in building the concept of a flexible IT architecture have no direct financial return, but they lead to a significantly faster and cheaper integration of new components and implementation of changes.

- Demands on the suppliers of CC services:
  Today, managers emphasize the requirements for shortening the delivery times, budget reductions and complexity of mutually integrated systems.
  Customers also increasingly demand the mobility of solutions. Applications can run anywhere; therefore, their mutual integration should be defined by standard interfaces.
  There is an increased pressure on the use of innovative approaches, such as agile development, prototyping or extreme programming.
  Analyses of a new generation of ICT infrastructure. Analytical tools based on big data become critical and their real-time analysis then facilitates the development of applications for business intelligence and control of transmission networks. [6]
  Analytical tools based on big data and their real-time analysis then facilitate the development of applications for business intelligence and control of transmission networks.

The decisions on the transformation of part or the entire company IT to CC is part of the information strategy of a company or institution. The supplier cannot deal with the transition to CC without coordination with company managers, who are informed about the possibilities of current ICT in relation to the entire business strategy of the company, availability of human and financial resources, know-how, etc.

The decision to transform the entire company IT (or its part) to CC is part of the information strategy of a company or institution. The supplier cannot deal with the transition to CC without coordination with company managers. [7]

- Impacts on user companies and organizations:
  As a result of operating applications in CC, companies will reduce the number of technology-oriented specialists (there will be less need for programmers, administrators, and other professionals).
  However, the number of employees ensuring the links between business and ICT services is set to grow. Nevertheless, the total number of workers involved in the use of ICT in the company will not decrease, although their qualification structure will change. [8]
  Companies will need to develop key skills supporting the use of ICT – how to use ICTs to gain a competitive advantage, create new products or services, find new customers, speed up the company's response to external events and reduce the costs of business processes. [9]
- In other words, how to support business processes by an appropriate selection of ICT services:
  - how to determine the content, volume, quality and price of ICT services,
  - how to design the overall architecture of ICT services,
  - how to find and implement the selection of the optimal supplier of ICT services,
  - how to systematically monitor the delivery of ICT services,
  - how to develop rules for the controlling of IT services and measure the impact of ICTs on the quality of company processes. [10]

This will require changes in skill required from managers. Managers who will understand how to use ICT to create a new product or service and how to get new customers will become indispensable members of senior management.

V. Approach to cloud computing services in Europe

Europe pays close attention to CC services because their use can significantly affect performance and competitiveness, particularly of small and medium-sized enterprises, which in many countries have a considerable impact on their economic stability.

A significant role played in these initiatives EuroCloud Europe. EuroCloud agrees that in a world Cloud Computing will become the main tool for collecting, processing, storing and selecting information and knowledge.

EuroCloud Europe is a cloud-based organization with its headquarters in Luxembourg. It builds strong relationships with local governments and the European Commission and supports the environment for the development and growth of cloud computing as a tool to support the growth of business and competitiveness. It sees cloud computing as one of the most important impulses for the creation of a knowledge

society in which physical resources are optimized and shared resources are universally accessible.

Currently, members of EuroCloud include 22 European countries that are very different both in terms of the rate of utilisation and provision of cloud services and the willingness to participate in the observance of standards and legal and security conditions which are crucial for an effective Europe-wide application. The activities of individual members can be found on the EuroCloud Europe website. [11]

VI. How to Maximize the Effect of Shared Services

This chapter provides a summary of the measures that should be implemented to support the strategy of transition to CC services in order to get the greatest possible benefit from its implementation.

Individual steps (phases) in the transition to the CC:

- The initial step when trying to maximize the effects of the transition to CC is the preparation of a schedule of a gradual conceptual transformation of the entire IT architecture. In terms of management, this document is a strategic document (see Chapter 4).
- It is followed by the specification of requirements which will be crucial in the choice of the supplier. Fundamental aspects include:
  - time of delivery,
  - amount of cost,
  - functionality,
  - performance of the information system,
  - complexity of the information system.
- In the third step, it is necessary to specify the progressivity rate of the IS. If the IS is to be progressive, the specification should also include demands on the rate of mobility and alternatives of the required access to its implementation. At this stage, it is also necessary to specify possible requirements for means allowing real-time analyses and the management of computer networks.

  Managers must be trained in these steps so that they can propose how to use ICT to improve the course and management of business processes, thus promoting the development of their own business.

  The possible focus of training:
  - how to use ICT to gain a competitive advantage,
  - create new products or services, find new customers,
  - speed up the company's response to external events,
  - reduce the costs of business processes. [12]
- The next step is adapting the qualification structure of employees as needed. This is due to the increase in the demand for employees ensuring the link between business and ICT services.

  Managers who have been trained and proved their worth among the members of senior management will then participate in all senior discussions on company strategy, changes in marketing and sales.
- Now, attention is paid to the financial aspects of the transition.

  It is necessary to:
  - prepare the budget in consideration of the changes in the cost structure – linearization of costs (elimination of the investment component),
  - prepare the employees for the measurement of consumption – gaining control over their work.

It needs to use of the scalability of shared services. [13]

Also:

Prepare the employees for the measurement of consumption – gain control over their work. The manager gets a good overview of what his/her employees are doing and how they are doing it.

- Thanks to fees for shared services, it is easy to get a detailed overview of the operating costs of individual agendas and identify these agendas.
- The final step is the preparation of a document on contractual relations with the service provider.

  It contains:
  - specific requirements
  - specific responsibilities to be delegated to the provider, under what conditions,
  - with what guarantees and also with what sanctions in case of non-compliance.

Likewise, agreements are to be prepared.

- The specification of the contractual relationship is to be focused on:
  - the provision of services
  - subject, functionality
  - objectives
  - expectations.

  Service scaling:
  - change of scope,
  - quality,
  - time of provision).

  Ensuring connectivity:
  - connectivity provider,
  - connectivity downtime,
  - the protection of personal and other sensitive data,
  - the division of responsibilities between the company and the supplier,
  - the legal status of physical equipment used to provide services,
  - software licensing terms,
  - conditions of system migration,
  - required customization.

Specific impacts will also become evident on the side of interested users of shared services, whether users from among companies or private owners, because they change to shared services so that they could benefit from a profit and be able to operate more efficiently, faster, and with better

planning. They can concentrate more on their core business, their mission or the entrusted tasks and do not have to be distracted by operational or other secondary activities.

VII. Conclusion

Current demands on the pace of life require the appropriate form of management. Management puts demands on the flexibility and information resources to support decision making. In order to maintain the quality of management at the highest level, it is necessary to use all the features of the existing information and communication technologies. The use of CC is one way to achieve this. Based on the findings in the available literature and the author's own experience of working with CC technologies, the paper provides recommendations for the transition to CC in companies and institutions. Recommendations are made with regard to the current situation in each company. These recommendations include a set of activities that need to be done to maximize the benefit from this change. Due to the variability of current conditions, this paper encourages further research into this area to ensure increasing flexibility and quality of the management of companies and institutions.

References

[1] J. Rezek, „key trends for the future", "Klíčové trendy pro budoucnost", in Computerworld, vol. 14, Prague: IDG Czech Republic, 2011, 2011-10-31, http://computerworld.cz/technologie/pet-klicovych-trendu-pro-budoucnost-ict-44055.

[2] B. Sosinsky, Cloud Computing Bible, Indianapolis: Wiley Publishing, 2011, Ch. 1. Pp. 44.

[3] A. T. Velteelte, T. J. Velte, and R. C. Elsenpeter, Cloud Computing—A Practical Approach, Brno: Computer Press, 2011, pp. 23.

[4] NIST National Institute of Standards and Technology's definition of Cloud Computing, NIST, 2011, 2016-1-13, http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf

[5] M. Tvrdikova, "Increase in the Competitiveness of SMEs using Business Intelligence in the Czech-Polish border areas", In Proc. FEDCSI-2013: 2013 Federated Conference on Computer Science and Information Systems, Krakow, IEEE, 2013. pp .981-984. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6644133&tag=1

[6] V. Petrjanoš, „New challenges of system integration", "Nové výzvy systémové integrace", Computerworld, vol. 10, PRAGUE: IDG Czech Republic, 2014, 2014-11-26, http://unicornsystems.eu/cz/novinky/clanek/nove-vyzvy-systemove-integrace.html.

[7] J. Voříšek, "Impacts trends in IS / ICT organizations", "Dopady trendů IS/ICT na organizace", Moderní řízení, vol. 3, Prague: 2006, 2006-3-10, http://modernirizeni.ihned.cz/c1-17978530-dopady-trendu-is-ict-na-organizace-cast-3.

[8] D. Loshin, Business intelligence: the savvy manager's guide, San Francisco: Morgan Kaufmann 2013. http://store.elsevier.com/Business-Intelligence/David-Loshin/isbn-9780123858894/

[9] P. Rozehnal, "Aspects of Approach to Application IS/IT in SMEs", In Proc. IDIMT-2013: Information Technology Human Values, Innovation and Economy, vol. 42, p. 121-128, Prague, 2013, pp. 121-128

[10] J. Ministr, „Modelling and Simulation Support of EMS Processes", In Proc. FIPWG International Symposium on Environmental Software Systems, Berlin, 2013, Springer, pp. 571-578.

[11] EuroCloud Europe, 2016-4-10, http://www.eurocloud.org.

[12] Future ICT profession, Budoucnost ICT profesí, Národní vzdělávací fond o.p.s., 2016-2-26, http://www.budoucnostprofesi.cz/sektorove-studie/ict-profese/spravce-aplikaci-a-it-infrastruktury.html.

[13] E. Ziemba, "The holistic and Systems Approach to the Sustainable Information Society", Journal of Computer Information Systems, vol. 54. pp. 160-116, Katowice: FAL, 2013. http://www.tandfonline.com/doi/abs/10.1080/08874417.2013.11645676

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1133–1136
DOI: 10.15439/2016F149
ACSIS, Vol. 8. ISSN 2300-5963

Knowledge Gained from Twitter Data

Wiesław Wolny
University of Economics in Katowice
ul. 1 Maja 50, 40-287 Katowice, Poland
Email: wieslaw.wolny@ue.katowice.pl

***Abstract*—Social media constitute a challenging new source of information for intelligence gathering and decision making. Twitter is one of the most popular social media sites and often becomes the primary source of information. Twitter messages are short and well suited for knowledge discovery. Twitter provides both researchers and practitioners a free Application Programming Interface (API) which allows them to gather and analyse large data sets of tweets. Twitter data are not only tweet texts, as Twitter's API provides more information to perform interesting research studies. The paper briefly describes process of data gathering and the main areas of data mining, knowledge discovery and data visualisation from Twitter data.**

I. Introduction

TWITTER is a social networking site directed towards short-form, fast communication. Launched in 2006, Twitter rapidly gained global popularity and has become one of the ten most visited websites in the world. As of March 2016, Twitter boasts 302 million active users who collectively produce 500 million tweets per day, and these numbers are continually growing. These characteristics have made it a primary source of real-time information.

Given this enormous volume of data, analysts have come to recognise Twitter as a virtual treasure trove of information for data mining, social network analysis and information for sensing public opinion trends. For companies, Twitter can be used to build business intelligence tools focused on Twitter data acquisition and processing. It can be used in many ways to collect raw Twitter data and to transform it into valuable business intelligence data.

Twitter is an exceptional tool for knowledge discovery due to six key features:

- Twitter's API is clean and well-documented, with rich developer tools.
- Twitter data are rich in information and have a data format that is convenient for analysis
- Twitter's terms of use for data are relatively liberal. Tweets are public and reachable to anyone. Twitter is based on the asymmetric following model that allows access to any account without request for approval.

However, analysis of Twitter messages (tweets) is regarded as a challenging problem due to some difficulties:

- large amounds of data that cannot be easily handled.
- tweets are short,
- over 40% [1] of tweets are of an informal type of discourse that does not cover any functional topics,

Despite these difficulties discovering knowledge can be simple and bring significant value.

II. Twitter data terminology

The Twitter messages are called tweets. Twitter users post messages that show up in the streams of all of the people who have subscribed to them. Unlike traditional blogs, microblogs are typically limited in the amount of text that can be posted. Twitter's limit is 140 characters.

Tweets often contain links to on-line resources, such as web pages, images, or videos, and more often than not, they refer to other users (called mentions). As is the case with most microblogging, when a message is posted, updates are seen by all users who have chosen to "follow" the author who posted the message (submitter).

In Twitter, all the posts are public. Most people may not receive them if they do not follow the submitter, but messages can be searched for with a keyword or topic and found by someone who is talking about a specific subject.

Twitter has its own conventions that renders it distinct from other textual data. Understanding the language and terminology that is used is important for effective knowledge acquisition.

There are some particular features used in Twitter:

1) Tweet — is a message posted on Twitter, consisting of 140 characters or fewer. It can contain text, photos, links and videos.
2) Twitter name — Twitter usernames appear with an at sign "@" before the name.
3) Hashtags — The # symbol, called a hashtag, is used to mark keywords or topics in a Tweet. It was created organically by Twitter users as a way to categorize messages.
4) Mention — Users of Twitter use the "@" symbol to refer to other users. Referring in this manner automatically alerts them.
5) "Reply" — is used to respond to a tweet. Replying to a tweet is a way of building relationships with followers and friends and joining in conversations.
6) A Retweet — is where one chooses to take a tweet from someone else and tweet it to one's own followers. It can either be done directly with the retweet button or by adding one's own message and including the letters "RT" ahead of the content that is being retweeted.

Twitter names, hashtags and mentions provide an easy way of identifying people and topics, and thus allow to search for and filter information on any subject of interest. Twitter messages have also many unique attributes connected with a tweet which are available with the Twitter API or other tools.

978-83-60810-90-3/$25.00

III. Gathering Twitter data

Any social media investigation is only as good as the data used for its analysis. The process of social media analysis involves essentially four steps: data identification, data analysis, data interpretation and, finally, information presentation.

The main problem is how to extract the information that is available on Twitter and how it can be used to draw meaningful insight. To achieve this, first there is a need to build a data analyser for tweets. Tweets are available to researchers and practitioners through public Twitter APIs. Twitter allows developers to collect data via Twitter REST API (https://dev.twitter.com/rest/public/) and the Streaming API (https://dev.twitter.com/streaming/overview).

APIs to access Twitter data can be classified into two types based on their design and access method [4]:

- REST APIs are based on the REST architecture that is now popularly used for designing web APIs. These APIs use the pull strategy for data retrieval, i.e. a user must explicitly request information in order to collect it.
- Streaming APIs provides a continuous stream of public information from Twitter. These APIs use the push strategy for data retrieval. Once a request for information is made, the Streaming APIs provide a continuous stream of updates with no further input from the user.

The response from Twitter APIs is in JavaScript Object Notation (JSON) format. JSON data can be converted to CSV and imported to databases. NoSQL databases, such as MongoDB, allows to store and access JSON data directly, without conversion.

Since gathering of data have particular target, further preprocessing and filtering of collected data can be done using @twitter_names and #hashtags as a search arguments for Twitter API. This method is more precise and provides better results in narrowing data than other text mining approaches.

IV. Storage of data

Explosion in the size of data generated on social media calls for a new data storage paradigm. Commonly used relational databases are insufficiently effective in storing very large datasets. Also the JSON-based data format, used in social media, requires additional conversion to a relational form. At the forefront of this movement is NoSQL, which promises to store large amounts of data in a more accessible way than the traditional, relational model.

There are several NoSQL implementations. One of the most popular ones is MongoDB (https://www.mongodb.org). MongoDB is a proper solution for storing Twitter data due to its adherence to the following principles:

- MongoDB is an open–source document database that provides high performance, high availability and automatic scaling.
- A record in MongoDB is a document, which is a data structure composed of field and value pairs.
- MongoDB documents are similar to JSON objects. This makes it very easy to store raw documents from Twitter's APIs.

In addition to these abilities, it also works well in a single-instance environment.

V. Data structure

Twitter APIs, besides basic information such as the tweet text and the author of the tweet, returns data structure contains additional information which can be used to provide further analysis. For each maximum 140 character tweet, API returns a JSON document containing over 160 items of metadata presented as key and value pairs.

Some the most useful keys for knowledge acquisition are:

- 'coordinates', 'geo' and 'place' — determine the location of the tweet's author;
- 'lang' — allows to easily specify the language of the tweet without text analysis;
- 'created_at' — the date and time of the tweet
- 'entities' — "symbols", "hashtags", "user_mentions" and "urls" included in the tweet;
- 'retweeted_status' — information about retweets;
- 'source' — application or website uses to create a tweet;
- 'text" — text of the tweet;
- "id" — unique identification of the tweet;
- "user" — contains 38 fields about the user. The most useful may be:
  - 'name' — user name;
  - 'time_zone' — time zone of computer or mobile;
  - 'created_at' — date and time of account creation;
  - 'description' — additional information about the user;
  - 'friends_count' — number of friends;
  - 'followers_count' — number of followers;
  - 'location' — actual location.

The above-mentioned keys can make it easier to analyse the text data, but also to perform various specific analyses of users, users nets, time and geolocation information.

VI. Methods of Twitter Data Analysis

Analysing Twitter data is searching through massive amounts of unstructured data. Filtering the data by Twitter names, topic or hashtags may reduce the data size, but it can still be enormous. Also, most Tweets contain no useful information.

Many different types of analysis can be performed with obtained Twitter data. The first can be simple text mining of posts, yet information provided within a tweet's text allows to conduct more Twitter-specific analysis, e.g. user information, connections between users, and localisation at the level of country even any place on the map of the world.

A. Text Mining

Text Mining refers to the process of deriving information from a text. A typical approach in Tweeter analysis is a document-level approach to scan a set of tweets written in a natural language and either to model the document set for predictive classification purposes or to populate a database or search index with the information that is extracted.

Most NLP-based text mining methods perform without particular success in social media. The informal and specialised language that is used in tweets as well as the nature of the microblogging domain make Twitter text mining analysis a very different task. Almost all forms of social media are very noisy and full of all kinds of spelling, grammatical, and punctuation errors. Text mining of tweets can be easier because Twitter API provides information about the language that is used, hashtags and usernames, hence there is no need for its detection.

B. Collecting a User's information

On Twitter, users create profiles to describe themselves to other users on Twitter. When a user creates or reconfigures an account, he/she provides some personal information, such as his/her name, username, password, email address, or phone number. The user may also provide with profile information, such as a short biography, location, website, date of birth, or a picture. On the Twitter Services, the name and username are listed publicly, including on the user profile page and in the search results. A user's profile is a rich source of information about him or her.

The Twitter REST API function users show (https://dev.twitter.com/rest/reference/get/users/show) is an easy way to obtain valuable information about a user, including:

- Real name,
- Description - which typically contains additional information about user,
- Entities such as hashtags, links and media, which can point to further sources of data,
- Followers_count,
- Friends_count,
- Location,
- Language.

An even more valuable function is the followers list (https://dev.twitter.com/rest/reference/get/followers/list). As the name suggest, it allows to access a whole list of followers in one query. The returned data may contain the same information for all followers as the function users/show. Using this function for Twitter's most followed users allows to collect information about millions of users without exceeding twitter API limits.

Information about the friends list is provided by function friends list (https://dev.twitter.com/rest/reference/get/friends/list).

Crawling using the above functions can be used to recognise networks of users.

C. Network Information

A Twitter user network refers to connected user accounts based on various types of relatedness. Structured content, in the form of friends and followers, @replies and @mentions, #hashtags and retweets, makes the Twitter user population networked in multiple ways. Each of these features can be considered as a kind of connection that can exist between two Twitter users. Kumar, Morstatter and Liu [4] categorised two main types of networks:

- Information Flow Networks.
- Friend-Follower networks,

A first kind of network shows who was mentioned or replied-to in the users' Tweets. Second kind of network is based on list of friends and followers of user.

Another type of network is one associated with time-bounded events, such as conferences. Many events like conferences now communicate a common "hashtag" or keyword to identify messages related to the event. Hansen, Smith and Shneiderman [5] [6] created EventGraps. EventGraphs help make sense of the collections of connections that form when people follow, reply or mention one another and a keyword.

To analyse and visualise social media network data from Twitter, the most popular software to use is NodeXL, a free and open add-in for Excel 2007/2010/2013. NodeXL is a project from the Social Media Research Foundation (http://www.smrfoundation.org/). NodeXL is a general purpose network analysis application that supports network overview, discovery and exploration.

D. Sentiment Analysis

The nature of microblogs is that people post real-time messages about their opinions on a variety of topics, discuss current issues, and complain and express either positive or negative sentiment for products they use in daily life. Data from these sources can be used in opinion mining and sentiment analysis tasks, e.g. manufacturing companies may be interested in the following questions:

- What do people think about their product (service, company, etc.)?
- How positive (or negative) are people about our product?
- What would people prefer our product to be like?

All of this information can be obtained from social networks. Opinions and related concepts such as sentiments and emotions are the subjects of study of sentiment analysis and opinion mining.

Sentiment analysis is a growing area of the Natural Language Processing. With the growing population of blogs and social networks, sentiment analysis have become a field of interest for many researches. A very broad overview of the existing studies was presented in [7].

E. Geolocated information

Twitter is also characterised by the diversity of its users, in terms of location. Harvesting this geospatial information provides a unique opportunity to gain valuable insight into information flow and social networking within society.

An important aspect of Twitter data is that some data are geotagged, which means that the posting user has attached a GPS coordinates to the tweet when uploading the information. Such information can be particularly important in order to understand where the user is and what he/she is referring to.

Fischer [8] tracked geo-tagged tweets from Twitter's public API for the last three and a half years. He claimed that there are about 10 million public geotagged tweets every day, which is about 120 per second. This is still a small portion of all tweets, and it is often necessary to use all sources of location information to determine the Tweet's location.

VII. Visualisation of Twitter data

Textual information is generated when users publish on Twitter. Analysing Twitter users and conversations is more than tabulating counts and trends — it is about connections and interactions between people. Twitter enables the collective creation and sharing of digital artifacts. The use of these tools inherently creates network data. These networks represent the connections between content creators as they view, reply, annotate or explicitly link to one another's content.

Twitter provides other embedded information, such as location data. All kinds of gathered data can also be analysed in the time dimension. Visualisation techniques can help to efficiently analyse and understand how and why users interact on Twitter. The display data task requires a remarkable collection of tools and skills.

A. Text visualisation

Text visualisations provide visual representations of documents or small corpora with the primary aim of supporting language analysis. The most popular method of text visualisation are tag or word clouds. Tag clouds are a simple but effective way of representing the distribution of words in a document or corpus, such as a tweet. They are widely employed for both casual use and serious analysis [9]. Clouds give greater prominence to words that appear more frequently in the source text. Clouds can be tweaked with different fonts, layouts, and color schemes.

Another method of text visualisation is the word tree, which is a technique that transforms text into a hierarchical representation based on a selected word or phrase [10].

B. Network visualization

By using network analysis, one can visualise complex sets of relationships such as graphs or sociograms of connected symbols and calculate precise measures of the size, shape, and density of the network as a whole and the positions of each element within it [11]. Structured content, in the form of friends and followers, @replies and @mentions, #hashtags and retweets, makes the Twitter user population networked in multiple ways. Each of these features can be considered as a kind of connection that can exist between two Twitter users.

There are at least as many kind of networks as there are features listed here. All networks can be categorised into network of friends, network of followers and information flow networks — retweet propagation. These and many other kinds of networks are identified and described in depth in [12] and [4].

C. Geolocation visualization

Location information is typically used to gain insight into the prominent locations that are discussing an event. Maps are an obvious choice to visualise location information. A basic method of creating map identifying tweet locations is to simply highlight the individual tweet locations. Each tweet is identified by a dot on the map, and such dots are referred to as markers. Another way is drawing circles of size representing an number of tweets aggregated.

The second kind of map is a trends map. A trends map allows to make real-time mapping of Twitter trends on map. These are displayed as hashtags, @mentions or keywords superimposed over a world map. The map of course can be zoomed in for more detail, and trends from various cities can be selected.

Another kind of map is a heat map of tweets. It allows to quickly identify regions of interest or regions with a high density of Twitter users. A heat map of twitter data can be generated using the https://worldmap.harvard.edu/tweetmap/ website.

References

[1] Pearanalitycs, "Twitter study — august 2009," 2009. [Online]. Available: http://pearanalytics.com/wp-content/uploads/2009/08/Twitter-Study-August-2009.pdf

[2] A. Go, R. Bhayani, and L. Huang, "Twitter sentiment classification using distant supervision," *Processing*, pp. 1–6, 2009. [Online]. Available: http://www.stanford.edu/~alecmgo/papers/TwitterDistantSupervision09.pdf

[3] F. Morstatter, J. Pfeffer, H. Liu, and K. Carley, "Is the sample good enough? comparing data from twitter's streaming api with twitter's firehose," in *International AAAI Conference on Weblogs and Social Media*, 2013. [Online]. Available: http://www.aaai.org/ocs/index.php/ICWSM/ICWSM13/paper/view/6071

[4] S. Kumar, F. Morstatter, and H. Liu, *Twitter Data Analytics*. Springer, Aug. 2013.

[5] *44th Hawaii International International Conference on Systems Science (HICSS-44 2011), Proceedings, 4-7 January 2011, Koloa, Kauai, HI, USA*, IEEE Computer Society. IEEE Computer Society, January 5-8 2011. [Online]. Available: http://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=5716643

[6] D. L. Hansen, M. A. Smith, and B. Shneiderman, "Eventgraphs: Charting collections of conference connections," in *44th Hawaii International International Conference on Systems Science (HICSS-44 2011), Proceedings, 4-7 January 2011, Koloa, Kauai, HI, USA*, IEEE Computer Society. IEEE Computer Society, January 5-8 2011. doi: 10.1109/HICSS.2011.196. ISBN 978-0-7695-4282-9 pp. 1–10. [Online]. Available: http://dx.doi.org/10.1109/HICSS.2011.196

[7] B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Found. Trends Inf. Retr.*, vol. 2, no. 1-2, pp. 1–135, Jan. 2008. doi: 10.1561/1500000011. [Online]. Available: http://dx.doi.org/10.1561/1500000011

[8] E. Fisher, "Making the most detailed tweet map ever," 03 2014. [Online]. Available: https://www.mapbox.com/blog/twitter-map-every-tweet/

[9] F. B. Viegas, M. Wattenberg, and J. Feinberg, "Participatory visualization with wordle," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 15, no. 6, pp. 1137–1144, 2009.

[10] M. Wattenberg and F. B. Viégas, "The word tree, an interactive visual concordance," *IEEE Transactions on Visualization and Computer Graphics*, vol. 14, no. 6, pp. 1221–1228, Nov. 2008. doi: 10.1109/TVCG.2008.172. [Online]. Available: http://dx.doi.org/10.1109/TVCG.2008.172

[11] M. A. Smith, B. Shneiderman, N. Milic-Frayling, E. M. Rodrigues, V. Barash, C. Dunne, T. Capone, A. Perer, and E. Gleave, "Analyzing (social media) networks with nodexl." in *Proceedings of the 7th Conference on Creativity & Cognition, Berkeley, California, USA, October 26-30, 2009*, J. M. Carroll, Ed. ACM, 2009. ISBN 978-1-60558-713-4 pp. 255–264. [Online]. Available: http://dblp.uni-trier.de/db/conf/candt/candt2009.html#SmithSMRBDCPG09

[12] D. Hansen, B. Shneiderman, and M. A. Smith, *Analyzing Social Media Networks with NodeXL: Insights from a Connected World*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2010. ISBN 0123822297, 9780123822291

14^{th} Conference on Advanced Information Technologies for Management

WE are pleased to invite you to participate in the 14^{th} edition of Conference on "Advanced Information Technologies for Management AITM'16". The main purpose of the conference is to provide a forum for researchers and practitioners to present and discuss the current issues of IT in business applications. There will be also the opportunity to demonstrate by the software houses and firms their solutions as well as achievements in management information systems.

TOPICS

- Concepts and methods of business informatics
- Business Process Management and Management Systems (BPM and BPMS)
- Management Information Systems (MIS)
- Enterprise information systems (ERP, CRM, SCM, etc.)
- Business Intelligence methods and tools
- Strategies and methodologies of IT implementation
- IT projects & IT projects management
- IT governance, efficiency and effectiveness
- Decision Support Systems and data mining
- Intelligence and mobile IT
- Cloud computing, SOA, Web services
- Agent-based systems
- Business-oriented ontologies, topic maps
- Knowledge-based and intelligent systems in management

EVENT CHAIRS

- **Dudycz, Helena,** Wrocław University of Economics, Poland
- **Dyczkowski, Mirosław,** Wrocław University of Economics, Poland
- **Korczak, Jerzy,** Wrocław University of Economics, Poland

PROGRAM COMMITTEE

- **Abramowicz, Witold,** Poznan University of Economics, Poland
- **Ahlemann, Frederik,** University of Duisburg-Essen, Germany
- **Andres, Frederic,** National Institute of Informatics, Tokyo, Japan
- **Atemezing, Ghislain,** Mondeca, Paris, France
- **Banaszak, Zbigniew,** Warsaw University of Technology, Poland
- **Bobkowska, Anna,** Gdansk University of Technology, Poland
- **Brown, Kenneth,** Communigram SA, France
- **Bruzda, Jaonna,** Nicolaus Copernicus University, Poland
- **Chmielarz, Witold,** University of Warsaw, Poland
- **Cortesi, Agostino,** Università Ca' Foscari, Venezia, Italy
- **Czarnacka-Chrobot, Beata,** Warsaw School of Economics, Poland
- **De, Suparna,** University of Surrey, Guildford, United Kingdom
- **Dufourd, Jean-François,** University of Strasbourg, France
- **Fosner, Maja,** Faculty of Logistics, University of Maribor, Slovenia
- **Franczyk, Bogdan,** University of Leipzig, Germany
- **Gontar, Beata,** Uniwersity of Lodz, Poland
- **Gontar, Zbigniew,** University of Lodz, Poland
- **Hartványi, Tamás,** Széchenyi István University, Hungary
- **Januszewski, Arkadiusz,** UTP University of Science and Technology in Bydgoszcz, Poland
- **Kannan, Rajkumar,** Bishop Heber College (Autonomous), Tiruchirappalli, India
- **Kersten, Grzegorz,** Concordia University, Montreal, Poland
- **Korczak, Jerzy,** Wrocław University of Economics, Poland
- **Kowalczyk, Ryszard,** Swinburne University of Technology, Melbourne, Victoria, Australia
- **Kozak, Karol,** TUD, Germany
- **Križanová, Anna,** University of Zilina, Slovakia
- **Langviniene, Neringa,** Kaunas University of Technology, Lithuania
- **Leyh, Christian,** Technische Universität Dresden, Chair of Information Systems, esp. IS in Manufacturing and Commerce, Germany
- **Ligęza, Antoni,** AGH University of Science and Technology, Poland
- **Lim, Ming K.,** University of Derby, United Kingdom
- **Ludwig, André,** University of Leipzig, Germany
- **Magoni, Damien,** University of Bordeaux – LaBRI, France
- **Matulewski, Marek,** Poznań School of Logistics, Poland
- **Michalak, Krzysztof,** Wroclaw University of Economics, Poland
- **Montemanni, Roberto,** University of Applied Sciences of Southern Switzerland, Switzerland
- **Owoc, Mieczyslaw,** Wroclaw University of Economics, Poland
- **Pamuła, Anna,** University of Łódź, Poland

- **Pankowska, Malgorzata,** University of Economics in Katowice, Poland
- **Patasiene, Irena,** Kaunas University of Technology, Lithuania
- **Pawełoszek, Ilona,** Częstochowa Univeristy of Technology
- **Quirin, Arnaud,** University of Vigo
- **Rakovska, Eva,** University of Economics in Bratislava, Slovakia
- **Ricci, Stefano,** Sapienza University of Rome, Italy
- **Rot, Artur,** Wroclaw University of Economics, Poland
- **Shinkevich, Aleksej Ivanovich,** Kazan National Research Technological University, Russia
- **Sitek, Pawel,** Kielce University of Technology, Poland
- **Speranza, Grazia,** University of Brescia, Italy
- **Stanek, Stanislaw,** General Tadeusz Kosciuszko Military Academy of Land Forces in Wroclaw, Poland
- **Surma, Jerzy,** Warsaw School of Economics, Poland and University of Massachusetts Lowell, United States
- **Teufel, Stephanie,** University of Fribourg, Switzerland
- **Tsang, Edward,** University of Essex, United Kingdom
- **Wolski, Waldemar,** University of Szczecin, Poland
- **Zanni-Merk, Cecilia,** Universite de Strasbourg, France
- **Ziemba, Ewa,** University of Economics in Katowice, Poland

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1139–1146
DOI: 10.15439/2016F313
ACSIS, Vol. 8. ISSN 2300-5963

Analysis of users of computer games

Witold Chmielarz
University of Warsaw Faculty of Management ul. Szturmowa 3 02-678 Warszawa, Poland
Email: witold.chmielarz@uw.edu.pl

Oskar Szumski
University of Warsaw Faculty of Management ul. Szturmowa 3 02-678 Warszawa, Poland
Email: oskar.szumski@uw.edu.pl

***Abstract*—The main aim of this article is to show the characteristics of individuals playing computer games and their styles of play. In order to present the relevant data, the authors limited the study sample to a selected group of individual users. In the current paper the authors presented the commonalities of gamers, their approach towards participation in games, and the awareness of potential changes and improvements in the area. They also held discussions concerning the obtained solutions and drew conclusions based on the present stage of research.**

I. Introduction

The main aim of this work is to analyze the use of computer games as one of the alternative forms of entertainment in the selected group of users under the circumstances of a dynamic development of devices and mobile applications running on them. The aim of this article is to analyze the situation where computer games are used by people who treat them not only as a form of entertainment but also as a kind of sport. The popularity and specific universal nature of the access to computer games facilitates a fast development of information technologies. A broadly defined concept of mobility also impacts the use of computer games, moving the focus from using PCs to the use of smartphones and tablets.

According to the statistics of Newzoo [12] service, in Poland in 2013 the number of gamers amounted to 13.4 million, out of which 98% used their PCs to play computer games (together with other platforms). We take the second position in Europe among the examined countries. The market of computer games in Poland is growing every year – in the end of 2014 it was worth about 280 million dollars and it will be growing by 3.8% a year, thus increasing the value of the entire market to 437 million dollars at the end of 2016 [9]. Hence, undoubtedly the subject matter is worthy of attention.

Unfortunately, the phenomenon itself is difficult to define and examine taking into account the formalized scientific analyses. Firstly, there is no clear definition of computer games [8, 10, 11, 13, 15, 20, 24]. In its narrow sense, this concept is treated literally as games in the form of software running only on traditional hardware such as (desktop, microcomputers, laptops or palmtops). In its broad, historical approach, the group encompasses also games running on devices such as a console, TV, gaming machines, smartphones and tablets (which are in fact communication and application computers). As the games running on all kinds of devices were being developed in parallel, and, in fact, there are PC equivalents of all kinds of games, we sometimes use this term in its broad meaning. Thus, for the needs of this study, the authors assumed that computer games are a generic term (hypernym) encapsulating the whole class of all kinds of games presented as a homogenous phenomenon. Secondly, there is no one generally accepted definition of a person playing computer games (e-gamer). Thus, in the narrow sense of the word, an e-gamer is a person who plays computer games every day or a few times a week, individually or taking part in a multi-player game. Sometimes, the scope of this term is limited to include only those players who treat MMO (Massively Multiplayer Online games) class games as a sport, and they try to play them professionally. However, we observe a more and more common tendency to expand the term to include also any individuals who play any kind of game from time to time, perceiving it as just one more alternative kind of entertainment. This article treats the concept of e-gamers in such a way. Thirdly, there is no (specific or clear) classification of computer games: there are a number of typologies based on various criteria, most frequently taking into account the type of activity required from the e-gamer playing games (e.g. logic, strategic, arcade, RPG (role-playing games), MMO (Massively Multiplayer Online games) etc., with a number of varying kinds and versions.

The phenomenon of computer games has been examined in numerous studies, in numerous countries and social groups [e.q. 4, 6, 7, 22, 25], including large-scale studies [e.g. 5, 21, 26]; nevertheless, they were carried out before the recent period of extreme popularity and growth in the number of applications running on smartphones and tablets. And the second point is – that they are concentrated on statistics of the players (with their features) or social field of problem rather then on IT development. The authors hoped to establish certain implications of the new phenomena with regard to the direction of computer games development. Therefore, the authors have undertaken the studies whose main aim is to analyze the use of such applications among users. The findings presented in this article constitute a brief report on the first stage of the research conducted among the gamers in Poland in 2015.

II. The assumptions of research methodology and population sample

Due to limited and fragmentary research concerning the area of internet computer games and e-gamers, both from the point of view of an individual client and a group of customers, in Polish and foreign literature, the studies have been based on the authors' own approach [1], quite different from surveys in Poland [25,26] and some different from research in the other countries [18, 19], consisting of the following steps:

- analysis of a selected group of players on the basis of a quantitative and qualitative survey, divided into the following parts:
  - characteristics of a computer player and identifying his or her preferences in computer games,
  - identification of potential effects and consequences of playing computer games for e-gamers.
- placing an internet version of a survey on the servers of the Faculty of Management of the University of Warsaw, conducting functionality test and its verification,
- carrying out the survey among the users, analysis and discussion of the findings,
- drawing conclusions from the obtained results concerning the current situation and possible directions of the future development of internet computer games on the basis of the users' opinions.

The article presents the results of the analysis of the first part of the completed survey. It allowed for identifying a particular group of people who play various kinds of games, using different kind of hardware and software, with a varying level of skills and expectations concerning the organizational and technical aspects of playing games. Only after the selecting the group of best, "professional" players, we may proceed to specify the implications and psychophysical effects of their involvement in individual and multi-player games. The latter aspect was examined in the second, sequentially conducted, stage of the survey, whose results and conclusions will be presented in subsequent publications [3].

The questionnaire surveys were conducted near the end of December 2015. The selection of the study sample was not accidental: it belonged to the category of convenience sampling, the respondents were mainly students of selected universities in Warsaw (University of Warsaw and Vistula University (Akademia Finansów i Biznesu Vistula)), of full-time and part-time BA, BSc and MA studies. The survey was also completed by two members of university staff who declared playing computer games. The surveys were circulated electronically, and the response rate did not exceed 70%. Students are particularly open to all kinds of innovation, especially if it concerns their private life or entertainment [23].

A specific limitation concerning this particular sample was an anticipated high percentage of smartphone, tablet, laptop and mobile phone users, devices of lower quality but with a longer durability. An additional argument for conducting research in this social group was the demand from company cooperating with us on the design and construction of specific game platforms. The company depended on the wide market recognition of students as the main customer of such a platform.

The survey was completed by 274 people, out of which 254 participants submitted correctly completed questionnaires (which constitutes 92.70% of the sample). Among the respondents there were 59.45% of women and 40.16% of men; 0.39% respondents did not answer this question. An average age of the respondent was 20.62 years, and the medium value was 19 years. The age is typical of students of the first years of BA and BSc students and the first years of the studies of the second cycle – the group asked to complete the questionnaires. The oldest person taking part in the survey (member of the university staff) was 37. Among the survey participants there were 63.39% of students, 35.83% working students and 0.79% employees. 70.87% indicated secondary level education and 20.08% post-secondary education – the survey was primarily conducted among the students of BA studies. 8.66% declared holding a BA degree or a certificate of completion of studies, only one person indicated having a PhD degree.

Over 45% of survey participants indicated that they are inhabitants of cities with over 500,000 residents, over 14% came from cities with 100,000-500,000 of inhabitants, over 21% from towns with 10,000-100,000 residents, almost 5% from towns up to 10,000 residents, and 12.6% declared that they come from rural areas. The simplicity of the survey did not cause many distortions during its completion; few respondents (17) completed also additional sections of the survey.

III. Analysis of the findings and discussion

Respondents provided answers to forty-one substantive questions, out of which responses to first twenty-one questions concerned the issues which are the aim of this article. The first group of questions concerned the characteristics of e-gamers and their use of computer games.

Nearly 40% of respondents provided positive answers to the question concerning frequent use of computer games, i.e. every day (20%) and a few times a week (over 19%). This is the score which is 10 percentage points lower than rare use of e-games, which amounts to more than 49%. After preliminary interviews with respondents it seemed that the interest in computer games will be higher. The high score of a reasonable way of playing computer games (a few times a month) - 22% showed that the games are just one of many alternative kinds of entertainment available today. Fig. 1 illustrates the findings of the research.

Taking into account the technical aspects concerning platforms which e-gamers use, in the last 12 months we observe a specific shift towards mobile devices, smartphones in particular. Thus, over 35 % of e-gamers (80.75% including other platforms) used mobile platforms (mainly Android) last year. The second place was taken by PC platform – 28.31% (65.24% including other devices), and the third position was occupied by the console (e.g. Xbox, PS) with the

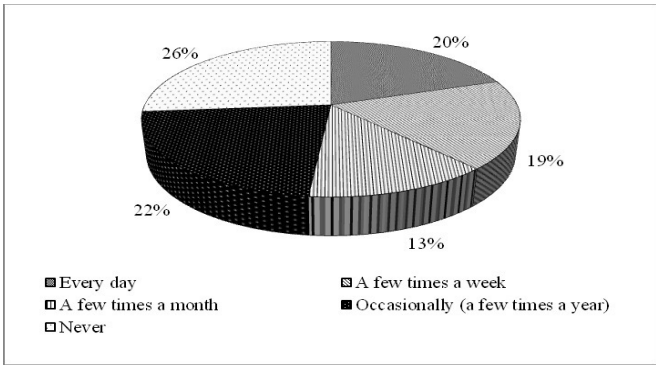


Fig 1. The frequency of playing computer games

score of 27.38 (63.10% - respectively). Smart TV platform received the lowest scores in this ranking – 2.09% (4.81).

In the perception of particular platforms among the e-gamers, we notice considerable discrepancies, amounting to 33 percentage points. The greatest number of respondents simultaneously use smartphones and PCs as platforms for games. Here, the dispersion of the results reaches almost 76 percentage points. The observed tendencies are presented in Fig. 2.

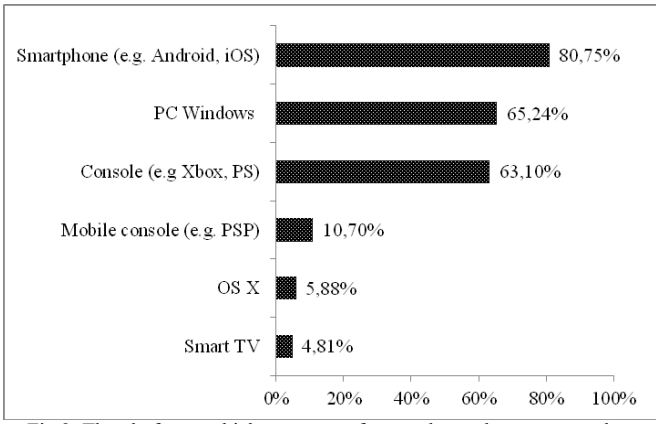


Fig 2. The platforms which were most frequently used as e-games platforms in the last year

On the other hand, it probably stems from the fact that the majority of e-gamers 59.87% (including other kinds of game access – 97.33%) use the games installed on a device (a PC or a smartphone). The second position of Steam, Origin etc. platforms amounting to 26.32% (respectively 42.78%) is a rather interesting phenomenon. The two main sources of games together constitute over 86% of "places" where e-gamers used the possibility of playing games in the last year. The remaining places where games were downloaded e.g. Facebook (6.25%) and low score of browsers (e.g. Quake Live) -7.57% seem to be of marginal importance in this relation. The Fig. 3 illustrates the scores.

The responses to the question concerning the age of e-gamers at the moment when they started to play games brought about very interesting results. The age which was most frequently indicated by respondents (almost 50% of responses) was within the range of 6-9 years (median of 6-7 years). If we add a group of people aged 10-11, we have

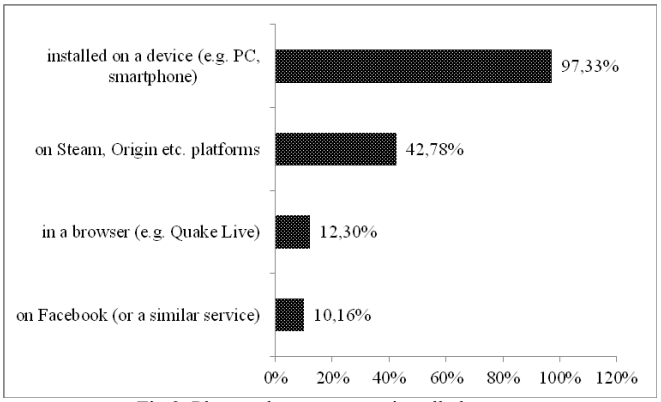


Fig 3. Places where e-gamers installed games

more than two thirds of all gamers! It is also significant that 17.11% of e-gamers declare that they started being interested in games at the age of 5. A marginal number (1.07%) admits starting playing games at the age of 20-25 (and the group that indicated the age of 16-25 amounted to 2.76%). This indicates the very early age when people become interested in computer games and treating the games as an alternative kind of entertainment in relation to films, TV, games or outdoor activities. Unfortunately, the limitation of the research was the fact that the authors did not examine children and young people from this age group. Nevertheless, the obtained results explain where – among others – the interest in computer games later in life comes from. The responses of survey participants were presented in Fig. 4.

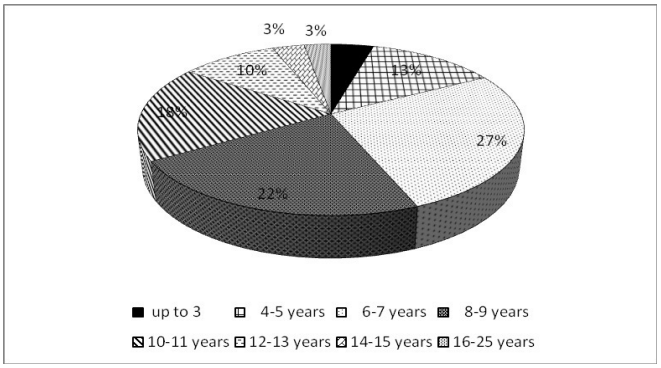


Fig 4. The age when respondents started playing computer games

Subsequently, the authors examined also the amounts of money which e-gamers spend monthly at playing computer games. The vast majority of them – 79.14% - use free applications installed on smartphones or free (or, as some people claim, illegally downloaded from the Internet) PC games. The remaining 18.18% of respondents are willing to pay up to PLN 80 monthly, and only 2.67% from PLN 81-300. From the commercial point of view, the last group (in particular 2.14% of survey participants who are willing to pay between PLN 151-300) is most interesting to examine because it includes mainly hobbyists, enthusiasts and fanatics – as it seems – professional e-gamers. The representatives of this group are interested in sport, which in this case is realized by means of various electronic tools (PC, smartphone or

tablet, console, etc.). The study results are presented in Fig.5.

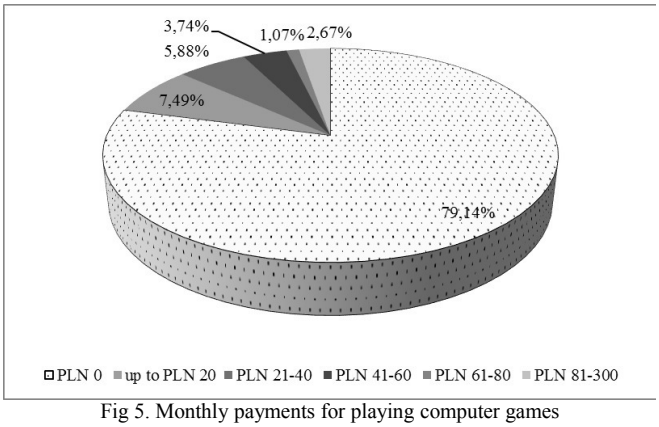


Fig 5. Monthly payments for playing computer games

The subsequent questions were used to evaluate the situation. Their goal was to indicate what kind of games the e-gamers played most frequently in the last year. The games were divided according to the typology indicated by the most frequent e-gamers:

- arcade games (shooting, fighting) (e.g. Counter Strike, Tom Clancy's Rainbow Six, Super Mario),
- action-adventure games (e.g. Assassin's Creed, Half-Life),
- adventure games (e.g. The Walking Dead, Wallace & Gromit),
- RPG (role-playing games) games (e.g. Diablo, Fallout),
- simulation games (e.g. The Sims, FIFA 16, Need for Speed),
- strategic games (e.g. StarCraft II, Civilisation, Warhammer, Heroes of Might and Magic),
- survival horror games – (e.g. Resident Evil),
- Massively Multiplayer Online games –MMO and their variants (e.g. World of Warcraft, Lord of the Rings Online).

Subsequently, the respondents answered the questions related to whether they played a particular kind of game in the period of last year. The questions formulated in such a way seemed to allow for more accurate responses that the ones which referred to the type of games they played most frequently. The authors worried that the responses concerning the last few months would dominate in the survey. They did not examine the recent trends in the market, the influence of newly published books and films related to particular themes, etc. The greatest number of positive responses, 80.75%, was indicated in the case of the simplest type of games – e.g. simulation games. The group of simple games also includes arcade games (57.75%) and action-adventure games (50.27%), where the number of positive responses exceeded 50%. In general, the greater complexity, the more complex relations, or the duration and additional limitations, the smaller the percentage of e-gamers admitting that they play a particular kind of game. The external factors, such as the history (the game was on the market "since I remember"), the popularity of a hero or a heroine or a plot constructed and popularized in films, books, board games, etc. contribute to the preeminence of the game. The games where the gamer needs to be more involved and stay in one place are less popular. The results of this part of the study are presented in Fig. 6.

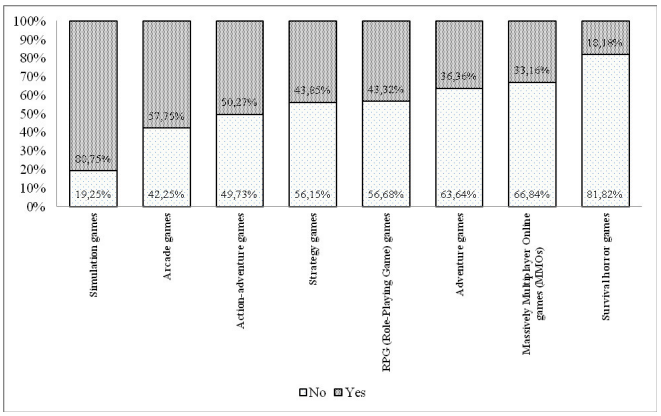


Fig 6. The most popular and most frequently played kinds of games

A good indication of the level of the engagement of the player is his or her willingness to choose to spend time in a game over other kinds of entertainment. The respondents were asked two questions if they 1) ever or 2) within the period of last year have chosen to spend time in a game over other, alternative forms of entertainment, such as:

- going to the cinema,
- meeting friends,
- going on a date,
- going for a trip with friends,
- going to a party,
- no such case.

It turned out that computer games are not enjoyable enough for players to give up anything in the past (61.78%) or in the last year (77.40%). If the respondents are willing to resign from something, it is mainly a meeting with friends 15.11% and a party 9.78%. In case of giving up anything in the last year in favor of a computer game, the results were similar. The respondents indicated a social meeting - 8.17% and a party – 6.73%. In reality, the difference indicated in the percentage of people who are willing to give up other forms of entertainment amounts to 17.32 percentage points, and decreases the actual numbers of indications in particular categories – the greatest with regard to social meetings – nearly 7 percentage points and parties – over 3 percentage points. The detailed scores are illustrated in Fig. 7.

In the respondents' views, the quality of computer games meets all or most expectations of players in 70% (Fig 8). The response that the game fulfills e-gamers' expectations to a moderate and limited degree is indicated only by 28% of respondents. A fraction of the sample evaluated the games as not enjoyable enough to consider giving up other activities

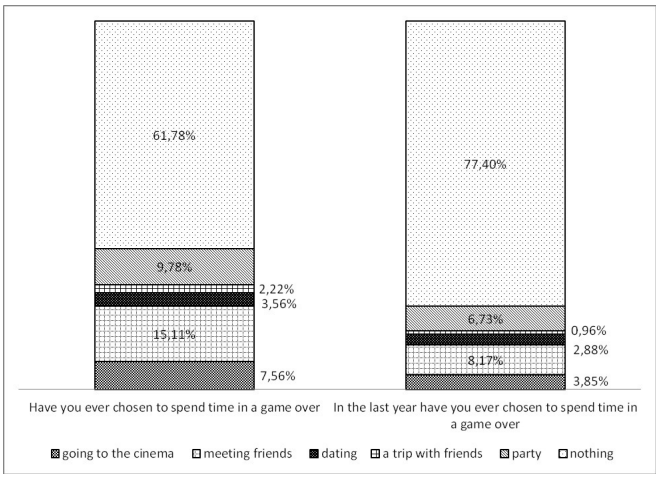


Fig 7. The willingness to give up other forms of entertainment among e-gamers

in favor of spending time in the game. Probably, it is one of the reasons why games are still so popular.

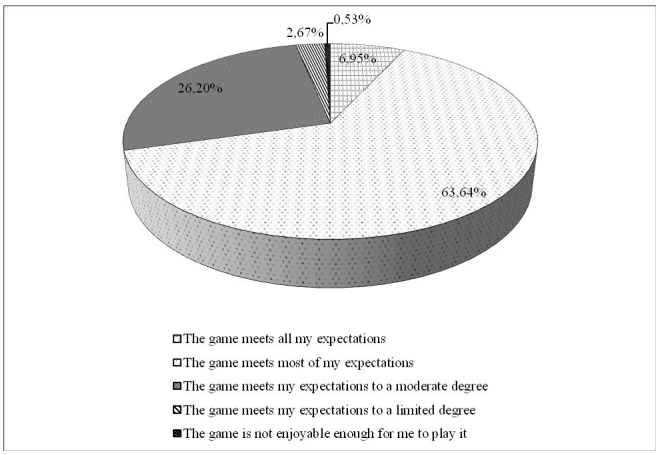


Fig 8. The quality of computer games in e-gamers' opinions

The vast majority of interviewed e-gamers (64%) are not interested in being leaders in games (provided that games offer such an opportunity). The remaining options are rather evenly distributed: 11% - clan leader, 7% - officer, 6% - advisor, 4% - higher-ranked officer and 8% - playing other roles (Fig. 9).

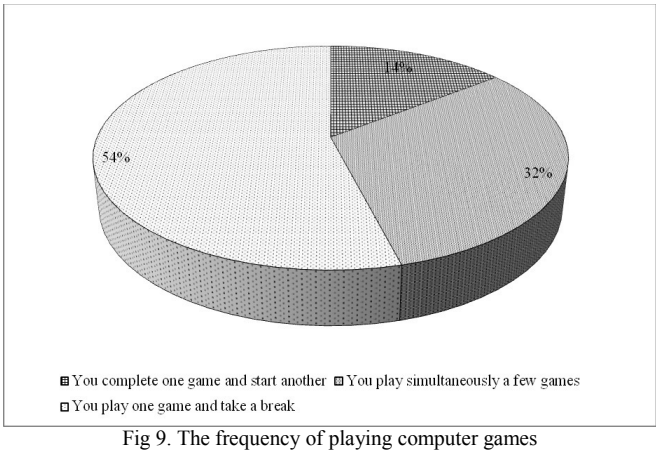


Fig 9. The frequency of playing computer games

Most e-gamers (54%) complete one game, take a break and only later start to play another computer game. Nearly 32% play a few games simultaneously. Only 14% finish playing one game and immediately start playing another. The frequency of playing games among respondents is shown in Fig. 10.

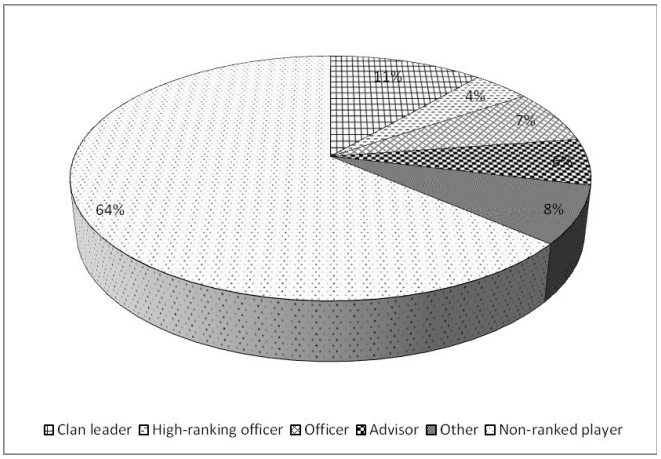


Fig 10. The willingness to be a leader among e-gamers

e-Gamers have a high opinion of their own skills as players (over 60%): they describe their skills as master level (18.72%) or advanced (42.72%). The number of gamers who see their skills as intermediate amounts to 33%, and less than 6% claim that they possess gaming skills at beginner level. Of course, due to the fact that gamers play games for a number of years, and general rules stay the same, e-gamers usually perceive themselves as specialists at using such opportunities, even if, thanks to new technologies these possibilities are constantly being developed. The structure of the e-gamers' skills is presented in Fig. 11.

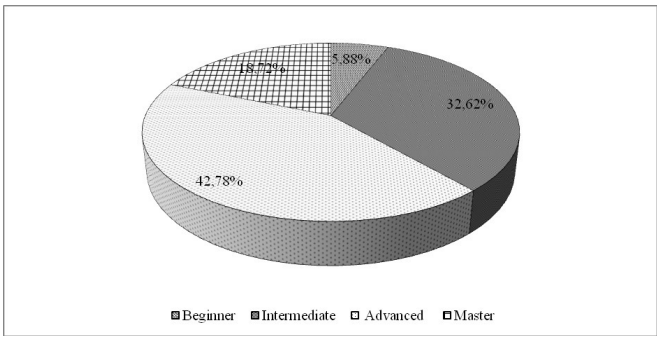


Fig 11. The structure of e-gamers' skills

The two remaining questions concerned the possible hardware conveniences and software advantages. In the first case, the respondents were given the following options to select from:

- obtaining mentor's help,
- using video and text tutorials (from game publishers),
- using in-game help,
- getting help from other gamers (e.g. forum),

- getting virtual or real payment,
- other advantages,
- no other advantages.

Almost 30% of gamers do not expect any advantages in this regard. They focus on the game they are currently playing, and they are satisfied with the game itself (passive players). Undoubtedly, the other e-gamers would be more satisfied if they could get help from other game users e.g. forum (22.14%), use text and video tutorials (12.55%) or in-game help (12.18%). Their satisfaction (18.08% of respondents) would increase if they received bonuses (additional options, game paths, etc.) or even actual reward (payment); yet, they have unrealistic or vague expectations concerning the latter. They do not pay attention to other conveniences or advantages of such kind. The results of this query are presented in Fig. 12.

With regard to the technical conveniences, e-gamers were

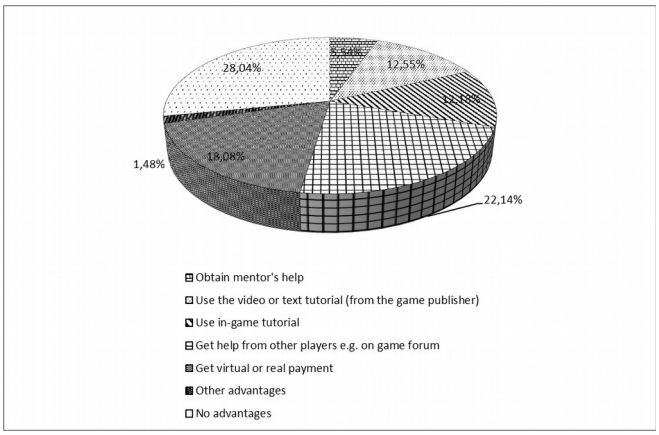


Fig 12. Non-technical conveniences for e-gamers

asked about the following, potential possibilities of changes concerning:

- computer hardware (e.g. graphic card) or a better tablet,
- armchair/seat,
- accessories (e.g. professional mouse, keyboard, earphones),
- better monitor/ VR goggles,
- other,
- I don't want to change anything.

In this case the responses were completely different than in the previous rankings. First of all, the structure of their responses was not evenly distributed. Nevertheless, almost one fourth (23.53%) of respondents are not satisfied with the hardware they own that they use to play a game, and they would like to change it. The distribution of the potential changes or lack thereof, was actually similar in relation to the remaining elements: better monitor or goggles – 18.53%, better armchair/seat – 21.76%, better accessories – 16.76% or no change at all – 16.47%.

Similarly to the previous case, basically e-gamers do not notice any potential for changes – less than 3% provided positive responses to this question, and there were no significant indications which we could relate to (e.g. additional lighting, additional monitors, etc.). The results are presented in Fig. 13.

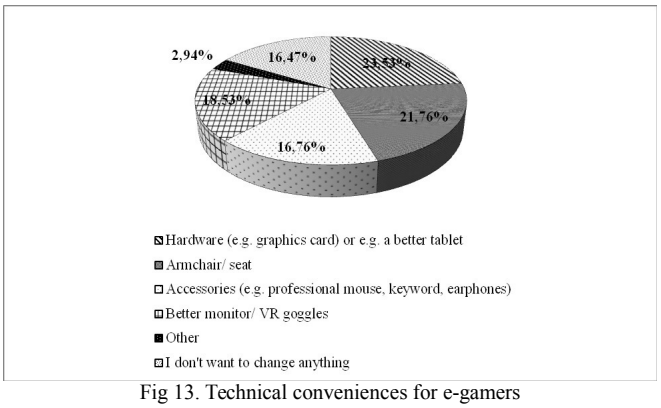


Fig 13. Technical conveniences for e-gamers

IV. Conclusions

The research conducted and presented so far points to the following conclusions:

- almost all respondents (over 99% of the sample) in the current study were students, which was reflected in the obtained scores. The older the students, the weaker interest in completing the questionnaire or its findings. It is caused by the increasing number of tasks connected with studies as well as the heavy workload connected with regular or temporary work (nearly 36% of working students). The latter is confirmed in the scores of other surveys [14, 16, 17, 26], despite the fact that, in total, fewer than 25-16% students participated in the study (even though it was always the largest group of players),
- among people who completed questionnaires there were markedly more women (almost 60%) than in other survey studies (around 43-48%) [25], conducted two or three years ago. Thus, we may conclude that there occurs a specific change with regard to the number of women playing computer games. Naturally, we should also be aware of the fact that the present study examined mainly the responses of students of economic faculties, and in this case the general number of female students in these faculties is greater than men. Still, the survey included also the option I don't play computer games, which the women could indicate,
- the frequency of playing the game (every day, up to a few times a week) in the examined sample was 20 percentage points smaller than in the case of other studies (39%, as compared to 62-63%). We should also consider the fact that in the other studies we took into consideration also another large group of potential gamers - pupils the group which ranks second with regard to the number of individuals spending time in the game –. The pupils have more free time than students, especially senior students.

All in all, majority of players – 54% of the interviewees, after completing one game, take a break before they start playing another game, and only 14% immediately start to play another game,

- the vast majority of players use their smartphone to play computer games (over 80%, mainly Android system – a large number of free games), which does not exclude also simultaneous use of other devices, mainly PC (over 65%) and a console, a regular (63%) or mobile one (11%). Smartphones and tablets started to take a role of a PC. Two or three years ago the proportions were more or less reversed; approximately 90% of respondents [9, 16, 17] used mainly personal computer, and only half of them a smartphone or a tablet. The devices allow for occasional use of many kinds of generally simple games at any place or time (not just during a break at work using your PC), killing the time while waiting for something else to take place,
- 97% of gamers use the games installed on a smartphone or a PC, with a surprisingly low percentage (10%) of people using Facebook games,
- due to the dynamic development of the use of smartphones and tablets in the last two years it occurred that the greatest number of people play simple simulation games (over 80%) and arcade games (58%) and action-adventure games (50%) which are becoming popular again. When we compare the present research with the earlier studies related to this area [9, 16, 17], the RPG games lost its popularity due to the increasing importance of mobile devices use (here: 44%, 65% - in other studies),
- notably, however, the early age when children start playing computer games, a shift towards younger and younger children (3-4 years younger since 2013) contributes to further development of computer games. More and more frequently it is caused by the fact that the first device with access to games is a smartphone, not a PC, and the fact that smartphones offer a greater number of free game applications,
- in general, the respondents (almost 80%) are not willing to pay for this kind of entertainment, and, as a vast majority, they use free smartphone applications and computer games which they received for their PC free of charge. It is reflected in the studies concerning the use of smartphones [1, 2] and a low tendency among students to spend their earnings on this purpose,
- it also explains the unwillingness to give up other kinds of entertainment, social life or rest to spend one's time in a game: almost 62% are not interested in choosing a game over any other kind of entertainment, and over 77% declared that they did not give up any activity in favor of a game last year,
- it appears that the fact that over 70% respondents claim that the level and quality of computer games fulfill all or almost all their expectations does not impact the situation,
- they have no expectations concerning taking leadership in a game (64%), they treat the games as a simple, not overly complex, form of entertainment. In general, they play games individually, and they are not interested – at least to a considerable degree – in multiplayer games,
- e-gamers have high opinions about their gaming skills – over 60% of participants claim that their skills are at least at an advanced level, and only 6% that they are beginners. On the one hand, it may be caused by the length of time of playing computer games (experience); on the other, it may result from the simplicity of most games that they play,
- the above said phenomenon is the reason why they do not expect too many advantages (none - 28%). If they were to choose, they would get help from other users (22%) or they would try harder to succeed in the game (18%) if they had a chance to obtain a reward or virtual bonus for winning the game,
- the case of technical conveniences is somewhat different. 23% would like to change their hardware hoping that this way they would have better chances to participate in the existing games and a greater possibility to participate in games of higher technical requirements. They pay attention to better accessories. Less than 3% would not improve anything as far as technical conditions of gaming are concerned.

The conclusions from the first stage of the research constitute good basis for further studies and expanding their offer, their consequences and impact of using games from the point of view of players. However, the present results already show interesting implications for the development of mobile information technologies towards new development trends of the use of this kind of software as a source of entertainment.

The further research – after preparing discussion and conclusions about sociological and psychological aspects of the gaming (include discussion of perceived positive and negative aspects of being a gamer or attempt to identify of the subcultures of players (the first attempt see [3]) - will focus on the market for suppliers of computer games and video games, in particular delivered to for mobile devices.

Results of a survey may be used not only by researchers in the field of computer games but by computer firms which want to make one step ahead in the development of this phenomenon.

References

[1] Chmielarz W.: Study of Smartphones Usage from the Customer's Point of View, Procedia Computer Science, Elsevier, Vol. 65, 2015, pp. 1085-1094; DOI:10.1016/j.procs.2015.09.045;

[2] Chmielarz W.: Porównanie wykorzystania sklepów internetowych z aplikacjami mobilnymi w Polsce z punktu widzenia klienta

indywidualnego (Comparison of the Use of Mobile Applications Websites in Poland from the Point of View of Individual Client) in: Innowacje w zarządzaniu i inżynierii produkcji edited by R. Knosala, in: Vol. II, Part IX Inżynieria jakości produkcji i usług, Oficyna Wydawnicza Polskiego Towarzystwa Zarządzania Produkcją, Opole, 2015, pp. 234-245,

[3] Chmielarz W., Szumski O. (2016), Analiza wykorzystania gier komputerowych (Computer Games Application Analysis), in: Mobilne aspekty technologii informacyjnych (Mobile aspects of IT), red. W. Chmielarz, Wydawnictwo Naukowe WZ UW, Warszawa, Dom Wydawniczy Elipsa, pp. 81-106; DOI: 10.7172/978-83-65402-25-7.2016.wwz.7;

[4] Duggan M. (2015): Gaming and Gamers, at: http://www.pewinternet.org/2015/12/15/gaming-and-gamers/, access, January 2016;

[5] Essential Facts about Computer and Video Games industry, ESA Entertainment Software Association, 2015 at: http://www.theesa.com/wp-content/uploads/2015/04/ESA-Essential-Facts-2015.pdf, access: January, 2016;

[6] Fang X., S. Chan, C. Nair (2009): An Online Survey System on Computer Game Enjoyment and Personality, in: J. A. Jacko (ed.), Human Computer Interactions, Part IV, HCII 2009, LNCS 5613, pp. 304-314, Springer Verlag Berlin Heidelberg; DOI:10.1007/978-3-642-02583-9_34;

[7] Fromme J., Computer Games as a Part of Children's Culture (2003), The International of Computer Game Research, volume 3, issue 1, 2003, at: http://www.gamestudies.org/0301/fromme/, access, January 2016;

[8] Homo Ludens 1/(2) (2010), Polskie Towarzystwo Badania Gier, access January 2016;

[9] http://akcjonariatobywatelski.pl/pl/centrum-edukacyjne/gospodarka/1033,Polski-rynek-gier-komputerowych-na-tle-rynku-swiatowego.html, access, January 2016;

[10] http://it-pomoc.pl/komputer/gra-komputerowa; access, January 2016;

[11] http://wiedzaiedukacja.eu/archives/tag/analiza-gier, access, January 2016;

[12] http://www.gry-online.pl/S013.asp?ID=82806; access, January 2016;

[13] http://www.gry-online.pl/S018.asp?ID=208&STR=2, access, January 2016;

[14] http://www.jestemgraczem.com/wyniki, access, January 2016;

[15] http://www.kipa.pl/index.php/promocja-filmu/gry-komputerowe/definicje-gier-komputerowych, access, January 2016;

[16] http://www.marketing-news.pl/message.php?art=43734, access, January 2016;

[17] http://www.newzoo.com/product/global-games-market-report-premium/, access, January 2016;

[18] https://www.surveymonkey.com/r/2WCW3K9, SurveyMonkey Inc. (US), access, January, 2016;

[19] http://www.survio.com/survey/d/D8Q9F2M7N4E0W5F0P, access, January 2016;

[20] https://pl.wikipedia.org/wiki/Gra_komputerowa, access January 2016;

[21] Lofgren K. (2015): 2015 Video Game Statistics & Trends; Who's Playing What & Why?; at: http://www.bigfishgames.com/blog/2015-global-video-game-stats-whos-playing-what-and-why/, access, December 2016;

[22] Mijal M., Szumski O., Zastosowania gier FPS w organizacji, in: Chmielarz W., Kisielnicki J., Parys T. eds), Informatyka @ przyszłości, Wydawnictwo Naukowe WZ UW, Warsaw 2013, pp. 165-176;

[23] Świerczyńska-Kaczor U., J. Wachowicz (2013), Student Response to Educational Games – An Empirical Study, Proceedings of the 2013 FedCSIS, pp. 1293-1299 at: https://fedcsis.org/proceedings/2013/plics/55.pdf,, access, January 2016;

[24] Zając J.: Jestem graczem w social media, at: http://blog.sotrender.com/pl/2014/12/jestem-graczem-w-social-media/, access, January 2016;

[25] Żywiczyńska E.: Co tak naprawdę wiemy o graczach, 2014, at http://zgranarodzina.edu.pl/2014/10/12/co-tak-naprawde-wiemy-o-graczach/, access, January 2016;

[26] Żywiczyńska E.: Optymizm czy myślenie życzeniowe. Zaskakujące wyniki badania #jestemgraczem, at: http://zgranarodzina.edu.pl/2014/12/20/optymizm-czy-myslenie-zyczeniowe-zaskakujace-wyniki-badania-jestemgraczem/, access, January 2016;

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1147–1152
DOI: 10.15439/2016F310
ACSIS, Vol. 8. ISSN 2300-5963

An Intelligent Context-aware System for Logistics Asset Supervision Service

Fan Feng, Yusong Pang and Gabriel Lodewijks
Section of Transport Engineering and Logistics
Delft University of Technology
The Netherlands
Email: {f.feng, y.pang, g.lodewijks}@tudelft.nl

***Abstract*—The use of Information and Communication Technology (ICT) has touched various aspects in the domain of transport engineering and logistics (TEL). As the development of TEL tends to be more complex in operation and large in scale, recent practices start to pay more attentions on improving system robustness and reliability. In addition, current ICT innovations (such as WSN and IOT) could record and deliver system descriptors (physical measurements, virtual resources, operational configurations) in real time. Such large-stream and heterogeneous data requires an integrated framework to process and management. To address such challenges, in this paper, a novel concept of context-aware supervision is proposed. An intelligent system with integration of semantic web and agent technology is proposed to support the concept realization, which aims at providing condition-monitoring and maintenance service to relevant user. A generic ontology-agent based framework will be illustrated. Finally, it will be applied for the supervision of a large-scale material handling system- belt conveying system as a proof-of-concept.**

***Index Terms*—Context-awareness, ontology-agent integration, system supervision, material handling system**

I. Introduction

THE USE of information communication technology (ICT) in TEL domain can be traced back to 1960s. It is chosen as a primary enabler to deal with increasing complexity of TEL development and enhance its competitive position with cost reduction and service promotion. Several conceptual ideas have been proposed that support the process of logistics and ICT technologies integration. The concept of *Integrated Logistics* is proposed to integrate IT with logistics management system to achieve synergy [1]. Afterwards, the concept of *E-Logistics* emerges that integrates Internet and mobile technology with logistics for providing one-stop value-added services to end-users [2], [3]. Recently, the concept of *Prognostics Logistics* has been put forward which utilizes wireless sensors (particularly RFID) together with decision making tools to enhance traceability and reliability of the logistics system [4].

Since the scale and complexity of TEL system has tremendously expanded, the attention of researchers and engineers have been shifted from enhancing operational efficiency towards improving system reliability and sustainability. Fact has been revealed by [5] who give a statement that the ultimate goal for a manufacturing system is guaranteeing an efficient production while providing functions needed by society in a sustainable and reliable way. It provides a new perspective regards the future development of TEL system. However, challenges still remain which can be categorized as follows:

- Heterogeneity: The heterogeneity is defined as system entities have different types of data model, properties, operation mechanisms and even different hardware and operating system [6]. As for TEL management especially for asset management and supervision, different data resources, operational information, past experiences and knowledges are characterized as heterogeneous resource and thus impose difficulties in integration.
- Interoperability: When it comes to interoperability, three perspectives can be identified [7], (1) organizational level: generic approaches and shared understanding of concepts, process, beliefs and terms [8]. (2) system level: interconnection between independent systems. (3) data level: consider the data properties include data format, data availability, data representation and semantic meanings. With respect to asset supervision, the interoperability challenges are inevitably presented at all three levels.
- Integrated decision making: Logistics asset is considered as large-scale and complex equipment. If a malfunction of single component or process has not been detected and corrected timely, it could lead to an expensive downtime and furthermore impose a great impact on the entire logistic activities. Consequently, a system with decision support becomes an essential element to provide relevant users a consistent understanding regards the system status and enabling an effective planning and execution of maintenance, such functionality could be referred as integrated decision making.

To cope with above mentioned challenges, in this paper, a context-aware supervision system is proposed which is used for information integration and supervision of large-scale asset service in TEL domain. The key ICT enablers are semantic web and autonomous agent. The ontology is used to model the semantic connections for heterogeneous data and various entities in supervision domain , thus enable information integration, data filtering and problem decomposition for specific supervision tasks. The usage of agent system intends to provide intelligent diagnosis and decision making functionalities through agent intelligence and cooperation. The integration of ontology-MAS delivers a context-aware intelligent system

978-83-60810-90-3/$25.00

and its practical usage will be considered with a case study of intelligent belt conveying system supervision.

The remaining part of the paper is organized as follows: Section II will introduce the concept of context-aware supervision system (CASS) and the motivations behind it. Section III will provide key technological enablers that could support the implementation of a CASS system. Section IV will first present the system design from an abstract structure perspective and the design of each functional block is discussed. Section V will presents a case study of applying CASS for intelligent supervising of a large-scale belt conveying system. The conclusion and future work will be addressed in section VI.

II. Context-aware supervision system

A widely recognized definition of context is given by [9] as *Context is any information that can be used to characterize the situation of an entity. An entity is a person, place or object that is considered relevant to the interaction between a user and an application, including the user and the application themselves.* And a system can be termed as context-aware if *it uses context to provide relevant information and/or services to the user, where relevancy depends on the user's task* [9]. A context-aware system (CAS) adapts and provides relevant information and the most appropriate service to users in an active and autonomous manner while requires little interactions [10].

In this paper, we focused on investigating the potential of applying CAS for asset supervision service. In essence, the supervision service includes system monitoring, failure/abnormality diagnosis/prognosis and maintenance planning. Apart of being context-aware of delivering meaningful information to user, it also requires a transparent flow from data to supervision method. To achieve that, the objective of a supervision system is to deliver accurate and timely information regards system conditions and propose effective maintenance actions to ensure reliability and availability of the system. Its success relies on integrating different diagnosis/prognosis methods. However, such methods often have certain scope of applicability and input context. As such, it put additional requirements on CAS to systematically integrate and manage system supervision processes. To integrate CAS for system supervision and fulfill additional requirements, a novel concept of *Context-aware supervision* is defined:

A context-aware supervision system (CASS) should include a series of functionalities include monitoring, supporting and advising in relation to system events. It not only focuses on diagnosing and prognosis failures, but also responsible for managing and organizing system knowledge, reasoning facts, integrating resources and analyzing problems. As such, the failure context can be given in a more meaningful manner that delivers information include: the specification of fault condition, recorded data linked to it, maintenance or operating actions linked to it, users that responsible for it and method that been used to determine it.

The characteristics of *context-awareness* are presented at two levels in CASS: (1) the supervision method should be aware of the context information it operates upon; (2) the end user should comprehend the created supervision context. In literature, several works have been established which attempt to consolidate the concept of context-aware with asset management such as context-awareness predictive maintenance [11], context-aware e-maintenance [12] and context-aware condition monitoring [10]. Two limitations are drawn from previous works: (1) the scope of applicability of proposed concept is limited given the fact that it only concerns partial aspect of the supervision process, For instance, the work of [11] concerns predictive maintenance, it lacks details regards how context been modeled and processed. (2) most works stay on a conceptual level which lacks sufficient technical details on how to put the concept into practice. Our work contributes to the literature by first introducing the concept of CASS, then we will discuss the technology been selected for putting such concept into action. Finally, a case study would demonstrate how the system works.

III. Key technological enabler for CASS

A. Context-modeling methodology

Intuitively, large amounts of context information are either acquired or derived from sensor devices. Normally, there exist gaps between raw data and the level of information which is useful to applications [13]. The context-modeling is used to bridge this gap by processing and transforming raw data before passed to context-aware services. Krummenacher et.al [14] have proposed several criteria for context model selection which include applicability, comparability, traceability, quality and so on. Meanwhile Hoareau et.al [13] conducted an extensive review for existing modeling choices such as key value models, makeup scheme, logic based models, object oriented models, ontology and so on. According to their in-depth discussion, the use of ontology is proposed to be the most expressive models to fulfill our requirements [15]. A formal definition of ontology is given by [16] as *an explicit specification of a conceptualization* which was used to describe a specific domain knowledge where concepts and relationships are unambiguously defined and checked. Recent works extensively applied ontology to facilitate the context modeling, its applicability covers domain include risk management of cold chain logistics [17], enterprise application [18], process supervision [19] and so on.

B. System supervision

A system supervision process considers providing users with decision support before/during/after the occurrence of system failure or abnormal situations. A typical supervision process consists of data acquisition, condition diagnosing/prognosis and maintenance planning. In this paper, rather than considering specific method or algorithm, we focus on how to provide a generic and adaptive environment to incorporate and integrate different methodologies and mechanisms operate together with flexibility and scalability.

Agent, as a tool in artificial intelligence domain, provides a way of dealing with complex engineering problem and

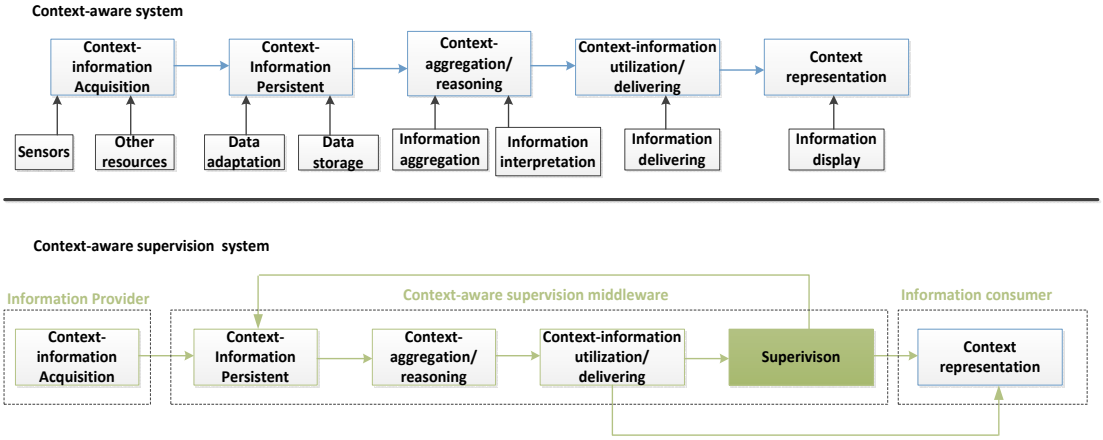


Fig. 1. System abstract architecture: a comparison between CAS and CASS

establishing adaptive system for decision making and information management through agent intelligence and collaborations [20]. State-of-arts demonstrate that agent system are largely applied to support system supervision functions, which include condition monitoring [21], risk management [22], e-maintenance [23] and so on. As such, agent technology is chosen as the key enabler for supervision system design, reasons are given as: (1) agent could cooperate and deploy on top of existing software. (2) In a multi-agent-system, agents could collaborate with each other to communicate and exchange information. (3) agent system could be deployed in distributed environment where new agent could easily join the system or leave the system as needed.

C. Ontology-agent integration

Key technological enablers have been chosen in previous section. We select ontology as the context-modeling method and agent system as the environment to support system supervision integration. In order to implement CASS, a next step is to consider the integration issues. In literature, several works have been established that concerns the integration of ontology and agent system. Dibley et.al [24] presented a work of building monitoring system where three ontologies are developed to capture the major semantics of a building environment and agent system is deployed to facilitate the monitoring tasks. For most of existing works, attentions are paid on using ontology to assist agent communication and knowledge retrieving. To implement CASS, potentials include information analysis, problem decomposition, agent status control are needed. To achieve this, a novel ontology-agent integrated framework is proposed, it will be elaborated in next section.

IV. System Framework

Fig 1 presents a comparison between a context-aware system and the proposed context-aware supervision system from an abstract structure perspective. A classical context-aware system follows five key processes [10]: (1) context information acquisition: gather information from virtual resources and physical sensors; (2) context-information persistent: data filtering and storing ;(3) context-aggregation/reasoning: interpretation and transfer low-order data to high-level applicable information via aggregation and reasoning; (4) context information utilization/delivering: apply context information to implement application-specific service; (5) context representation.

All of the functional blocks from CAS are inherited and implemented in CASS. The major difference is distinguished from two aspects. **The supervision block**: The key objective of CASS is to assist system supervision with aspects include equipment monitoring, condition diagnosis and maintenance planning. Such tasks are unable to perform well by only using context-modeling(e.g. ontology reasoning). In most cases, it requires advanced platform/engine for decision making. As such, the supervision block is introduced. **The information flow**: The information flow is also adjusted. In CAS, the information flow follows an open loop style where data is gathered from ground layer, processed through each functional block and becomes context-aware. For CASS, a partial closed-loop is formed. In essence, the aggregated and processed context information will be the input source for supervision module. The output of supervision module will be feedback for further processing. It will be first stored in data base and then processed by context model. By doing so, not only the measured data from ground layer would be context-aware but also the supervised result will be aggregated with other relevant information together to make result meaningful to end user. Moreover, it would be helpful to use the returned supervision information to infer new knowledge and propose further actions.

A. Context Model Design

We design an ontology termed *ontoSupervision* to capture major concepts and relationships in the domain of system supervision. The schematic of ontoSupervision is given in Fig 2 and explanation of each taxonomy is given below:

(1) *System taxonomy* is the core concept that presents a description of the system. It includes notions of system

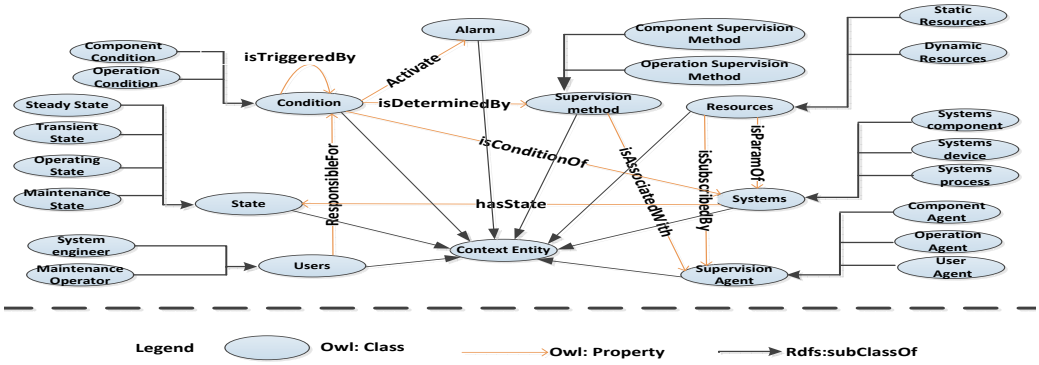


Fig. 2. Upper ontology taxonomy: definition of classes and object properties

fundamental components and operations that needs to be supervised. In addition, the subclass *system devices* contains peripheral devices. Other taxonomies either directly or indirectly connect with system through well defined relationships. (2) *Condition taxonomy* incorporates the notion of condition in the system. Two subclasses are included, namely operation condition and component condition. The former one concerns the system abnormal condition during operations and the latter one addresses the physical condition of system at different levels (component,instrument and equipment). (3) *Resources taxonomy* represents all relevant information resources that need to be accessed by supervision method. It is composed of two subclasses, namely static resources and dynamic resources. The former one can be thought as the resource that does not change over time such as system specification, historical information and system configurations. The latter one represents the notion of resources that change in real time, such as data acquired from sensor devices and any updated supervision results. (4) *Supervision agent & method* represents all available agents been deployed in current system and its associated supervision method. In essence, it serves as a bridge that connects context model with agent system. (5) *State taxonomy* represents possible state of the system. In current model, four states are considered namely maintenance state, transient state and operation state and steady state. (6) *Alarm taxonomy* represents the notion of possible alarm level that been activated by supervised conditions in the system. (7) *User* represents the information consumers in the system. It specifies the responsibilities and point of interests of respective user.

B. Agent system design

Regardless of the models, scopes and design tools, all supervision methods require system measurements as input and generate results which transform system conditions (operation and component condition) as supervision result. Such common structure allows the supervision methods be represented as supervision agent. The multi agent system can be perceived as a wrapper which provides environment for different supervision intelligence to perform supervision tasks. It also enable integrated decision making by taking the advantages of agent communication and collaboration. In the proposed framework, three kinds of agent are developed:

(1) **Supervision Agent**: Two categories of agent groups are considered as supervision agent. The first one is termed healthiness agent (HA) which is responsible for fault diagnosis at different level of system granularity. Single HA could be used to assess the condition of piece of equipment while multi HAs could work together for evaluating the overall healthiness of the whole system by consolidating different conditions. Another one is termed operation agent (OA) which is used to capture the abnormality during system operation. Typically, it is used to identify the abnormal deviation from normal operations or improper configurations. The agent intelligence, scope of interest, input information and responsibility are determined by its associated methods. (2) **Information Mediator Agent**: The information mediator is used to manage and control the agent execution and interactions. Its necessities are given as: It serves as an information portal for supervision agents; It keeps an active connection between agent system and ontology knowledge model. (3) **User Agent**: It contains the information consumer of the system. Any on-going supervision conditions will be relayed to it via IMA. Sophisticated GUI will connect with it to provide end user a friendly interface.

C. Agent-Ontology integration

The key of agent-ontology integration is achieved via the interaction between information mediator agent and ontology knowledge base. Such interactions aim at manipulating ontology to acquire information and knowledges where actions include create, read, update and delete entities in ontology. We identified three major processes:

- *Information acquisition*: In this case, ontology is treated as a hybrid database which is used to locate and retrieve information. A typical scenario can be that when a new sensor measurement is available in ontology, the IMA could retrieve it by executing well defined query template. An example of a query template is shown in Fig 3.
- *Knowledge acquisition*: It fully utilizes the reasoning capability of an ontology model. When certain information is available, the ontology could infer new knowledge

```
SELECT ?par ?agent ?system ?method ?timeStamp
?state ?measurement
WHERE {
        ?par onto:isSubscribedBy ?agent.
        ?par onto:isParametersOf ?system.
        ?agent rdf:type onto:AgentName.
        ?method onto:isAssociatedWith ?agent.
        ?system onto:hasState ?state.
        ?par onto:hasTimeStamp ?timeStamp.
        ?par onto:isMeasuredBy ?measurement.}
```

Fig. 3. Example of information acquisition template

by executing context-rules. For instance, if a belt idler temperature is over 70 degree, the ontology could use rules to determine that such context indicates a idler is in fault condition.

- *Knowledge reasoning and agent control*: As discussed previously, a partial closed loop is formed in the structure of CASS. The key motivation is given that any returned supervised information could be further processed by ontology. And the inferred knowledge could be useful in coordinating agent activities. For instance, a misalignment condition often occurs during the running of a belt conveyor and such condition could be induced by multiple reasons(improper power supply, overloading and so on) and such casualty relationships could be predefined in the ontology via proper object properties. By doing so, when a misalignment condition is supervised and returned, the ontology could running context rules to find the relevant condition relate with it. Consequently, the associated agent will be activated to allow a depth investigation of the root cause.

V. Implement CASS for large scale material handling system

A. case demonstration

Belt conveying system is widely accepted as a major equipment in continues material handling domain. Its usages are well developed in various logistics domains, such as container/dry bulk terminals, airport and mine industry. Normally, the BCS is deployed in an open and harsh environment, as such, major components could suffer severe damage as system ages. Consequently, a monitoring and supervision system with decision support is essential to help users(from operations, maintenance, reliability and other departments) gain a consistent understanding about the system status and enabling effective planning and execution of maintenance. Due to the limit space of the paper, we demonstrate a typical fault supervision process- belt tear condition supervision which is made up for 85% among all system component damages for a BCS [25].

B. Scenarios

This scenario demonstrates the CASS capability of the system. Specifically, when inspection of tear shape is available, the system should analyze the damage level and propagation pace of the damage by intelligent supervision method, and create decisions in the form of possible maintenance activities and/or warning/alarm message if needed. We identify three key processes of implementing the context-aware supervision service for belt tear condition:

- Context modeling: It concerns extending the upper ontology (ontoSupervision) with definition of new entities for application purpose. Such extension is termed domain-specific ontology and partial illustration for BCS supervision (ontoBeltCon) is depicted in Fig 4. The semantic meaning is given as: a tear shape (TS) *isMeasuredBy* a human inspection tool(HIT), which *isParameterOf* belt (belt section01). To supervise the tear condition, a belt tear supervision agent (BTSA) is designed. The BTSA *hasAssociatedMethod* belt tear supervision method (BTSM1). For supervision purpose, TS and a belt tear condition log (BTCL) *isSubscribedBy* BTSA. Upon successful decision making, a belt tear condition (BTC) is supervised which *isDeterminedBy* BTSM1 and *activate* alarm (alarm level 1). Finally the BTC *isResponisbleFor* user (maintenance operator 01).
  Besides newly added entities and its individuals, the data properties for a belt tear shape and belt tear condition is also available in Fig 4.
- Context supervision: After context information are collected and pre-processed by ontology, the agent intelligence should be invoked. For a belt tear condition supervision, a fuzzy logic based approach is applied [25]. Decisions are made based on the current tear shape measurement and history inspection log for the same shape. Two indicators (belt wear index and inspection frequency index) are provided to deliver a consistent understanding and interpretation of the supervision result and give straight forward suggestions for possible maintenance actions.
- Response actions: The supervised condition will be send back to ontology model for further processing before finally delivered to end users. In essence, it will use the supervision indicators to quantify the alarm level by running defined rules. For the given scenario, the rules can be given as:
  BeltTearCondition(?condition), greaterThan(?level, 0), hasWearIndex(?condition, ?level), lessThanOrEqual (?level, 0.7) $->$ AntiHealthCondition(?condition)

VI. Conclusion and future work

In this paper, a novel concept of *context-aware supervision* and its associated implementation techniques are proposed. The motivation behind the concept is to enable an efficient and transparent information flow for asset supervision tasks. We implement such system for supervision of a large-scale material handling system to demonstrate its major functionalities and potential usage in the domain of logistics. Future works include further extending the ontology model to incorporate more generic entities and concept in the system supervision domain. Moreover, to cope with more complex diagnosis/prognosis problem and enable more sophisticated decision making engine, the agent intelligence should be future investigated.

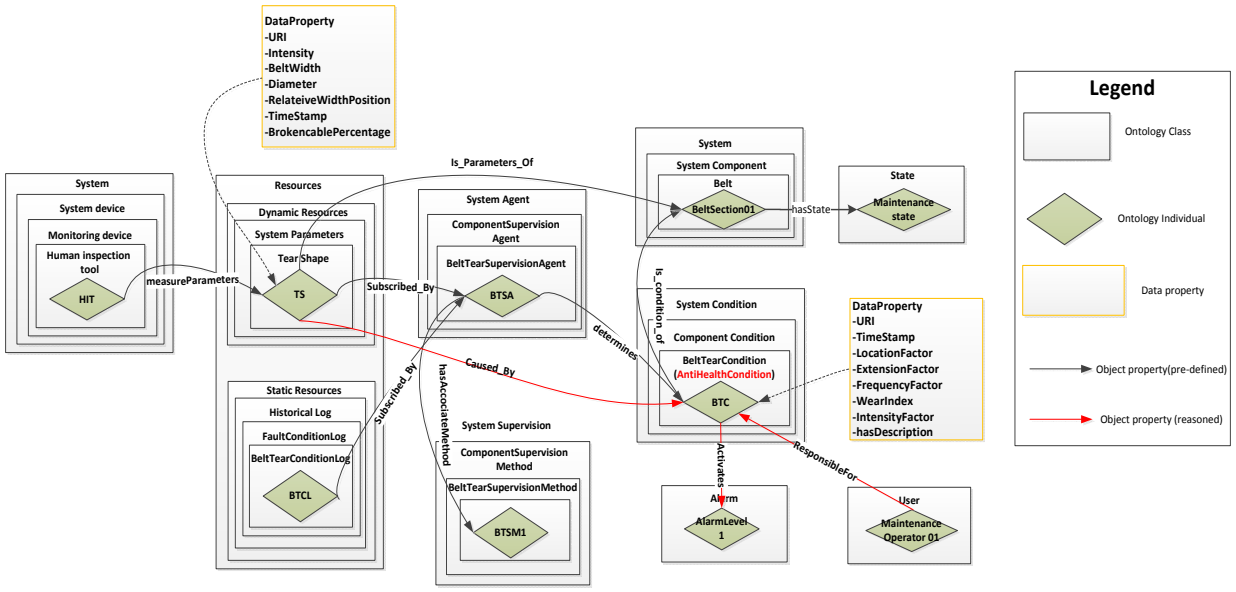


Fig. 4. OntoBeltCon configuration for belt tear condition supervision: include class, subclasses, data properties and object properties

REFERENCES

[1] H. N. Chiu, "The integrated logistics management system: a framework and case study," *International Journal of Physical Distribution & Logistics Management*, vol. 25, no. 6, pp. 4–22, 1995. doi: 10.1108/09600039510093249

[2] A. Gunasekaran, E. W. T. Ngai, and T. C. E. Cheng, "Developing an e-logistics system: a case study," *International Journal of Logistics Research and Applications: A Leading Journal of Supply Chain Management*, vol. 10, no. 4, pp. 333–349, 2007. doi: 10.1080/13675560701195307

[3] U. Arnold, J. OberlÃd'nder, and B. Schwarzbach, "Advancements in cloud computing for logistics," in *Proceedings of the 2013 Federated Conference on Computer Science and Information Systems*, M. P. M. Ganzha, L. Maciaszek, Ed., 2013, pp. pages 1055–1062.

[4] D. La Cruz, A. López, H. Veeke, and G. Lodewijks, "Prognostics in the control of logistics systems," in *IEEE International Conference on Service Operations and Logistics, and Informatics, 2006. SOLI'06.* IEEE, 2006, pp. 1–5.

[5] S. Takata, F. Kirnura, F. van Houten, E. Westkamper, M. Shpitalni, and etc, "Maintenance : changing role in life cycle management," vol. 1, no. 1, 2004.

[6] G. Thomas, G. R. Thompson, C.-W. Chung, and et.al, "Heterogeneous Distributed Database Systems for Production Use," *ACM Computing Surveys (CSUR) - Special issue on heterogeneous databases*, vol. 22, no. 3, pp. 237–266, 1990.

[7] L. S. Winters, M. M. Gorman, and A. Tolk, "Next generation data interoperability: It's all about the metadata," in *IEEE Fall Simulation Interoperability Workshop*, 2006.

[8] T. Clark and R. Jones, "Organisational interoperability maturity model for c2," in *Proceedings of the 1999 Command and Control Research and Technology Symposium*, 1999.

[9] A. K. Dey, "Understanding and using context," *Personal and ubiquitous computing*, vol. 5, no. 1, pp. 4–7, 2001.

[10] D. Galar, A. Thaduri, M. Catelani, and L. Ciani, "Context awareness for maintenance decision making: A diagnosis and prognosis approach," *Measurement*, vol. 67, pp. 137–150, 2015.

[11] B. Schmidt, D. Galar, and L. Wang, "Current Trends in Reliability, Availability, Maintainability and Safety," 2016.

[12] P. Pistofidis and C. Emmanouilidis, "Profiling context awareness in mobile and cloud based engineering asset management," in *Advances in Production Management Systems. Competitive Manufacturing for Innovative Products and Services*. Springer, 2012, pp. 17–24.

[13] C. Hoareau and I. Satoh, "Modeling and processing information for context-aware computing: A survey," *New Generation Computing*, vol. 27, no. 3, pp. 177–196, 2009.

[14] R. Krummenacher and T. Strang, "Ontology-Based Context Modeling," *Ieice Transactions On Information And Systems*, vol. E90-D, no. 8, pp. 1262–1270, 2007.

[15] R. Schmohl, U. Baumgarten, and D.-G. M, "A Generalized Context-aware Architecture in Heterogeneous Mobile Computing Environments A Generic Context-aware Architecture," *Wireless and Mobile Communications*, pp. 118–124, 2008.

[16] N. Guarino, D. Oberle, and S. Staab, "What is an ontology?" in *Handbook on ontologies*. Springer, 2009, pp. 1–17.

[17] K. Kim, H. Kim, S.-K. Kim, and J.-Y. Jung, "i-RM: An intelligent risk management framework for context-aware ubiquitous cold chain logistics," *Expert Systems with Applications*, vol. 46, pp. 463–473, 2015.

[18] D. Nadoveza and D. Kiritsis, "Ontology-based approach for context modeling in enterprise applications," *Computers in Industry*, vol. 65, no. 9, pp. 1218–1231, 2014.

[19] S. Natarajan and R. Srinivasan, "Implementation of multi agents based system for process supervision in large-scale chemical plants," *Computers and Chemical Engineering*, vol. 60, pp. 182–196, 2014.

[20] G. KovÃącs and K. Grzybowska, "Supply chain coordination between autonomous agents: A game-theory approach," in *Proceedings of the 2015 Federated Conference on Computer Science and Information Systems*, ser. Annals of Computer Science and Information Systems, vol. 5. IEEE, 2015, pp. 1623–1630.

[21] I. Mahdavi, B. Shirazi, N. Ghorbani, and N. Sahebjamnia, "IMAQCS: Design and implementation of an intelligent multi-agent system for monitoring and controlling quality of cement production processes," *Computers in Industry*, vol. 64, no. 3, pp. 290–298, 2013.

[22] R. J. Dawson, R. Peppe, and M. Wang, "An agent-based model for risk-based flood incident management," *Natural Hazards*, vol. 59, no. 1, pp. 167–189, 2011.

[23] R. Yu, B. Iung, and H. Panetto, "A multi-agents based e-maintenance system with case-based reasoning decision support," *Engineering applications of artificial intelligence*, vol. 16, no. 4, pp. 321–333, 2003.

[24] M. Dibley, H. Li, Y. Rezgui, and J. Miles, "An ontology framework for intelligent sensor-based building monitoring," *Automation in Construction*, vol. 28, pp. 1–14, 2012.

[25] G. Lodewijks and J. Ottjes, "Application of Fuzzy Logic in belt conveyor monitoring and control," *International Materials Handling Conference (Beltcon) 13*, pp. 1–13, 2005.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1153–1158
DOI: 10.15439/2016F145
ACSIS, Vol. 8. ISSN 2300-5963

Towards Paired Transactions Modeling

Frantisek Hunka
University of Ostrava, Faculty of Science
Dvorakova 7, 701 03 Ostrava
Czech Republic
Email: frantisek.hunka@osu.cz

Jiri Matula
University of Ostrava, Faculty of Science
Dvorakova 7, 701 03 Ostrava
Czech Republic
Email: jiri.matula@osu.cz

***Abstract*—Paired transactions or paired transfers have their origin in accountancy systems. The Resource-Event-Agent (REA) ontology uses paired transactions as a basic building block for business process modeling. A business process (REA model) is composed of two sets of paired transactions. REA itself originates from accountancy systems and has gradually developed into a full-fledged information system framework. REA model used to be depicted by ER diagrams and later by UML class diagrams. However, both these diagrams were not designed to capture conceptual models which are more precise, comprehensible for domain experts and easily to modify. ORM (Object Role Modeling) represents approach to conceptual modeling which fulfils above mentioned requirements. The main goal of the paper is to derive and describe the ORM model of paired transactions which corresponds to REA exchange model and assess this approach.**

I. Introduction

The REA ontology can be classified as a domain specific ontology that is focused on value modeling of business processes. The three core REA concepts are *resource*, *event* and *agent* from which the name of the modeling approach was derived. The aim of the REA modeling approach is to record any changes in property rights to resources, register resource usage, resource consumption or resource production, see [1,4,7].

Resource entities can be exchanged for other resource entities in REA exchange processes and can be converted to other resources in REA conversion processes as well. A REA application keeps track of which resources were exchanged for which ones or which resources were converted for which others.

The REA model records information based on the coherence between data of one or more business events. The REA process is defined by related REA events and has at least two composite economic events: a *decrement event* that outflows, consumes or uses the outgoing resource(s) and an *increment event* that inflows or produces the incoming resource(s). The REA process is called the REA model and represents the notion of a business process.

The main benefit of the REA modeling approach is the possibility of keeping track of primary and raw data about economic resources, by [5]. All accounting artifacts such as debit, credit, journals, ledgers, receivables and account balances are derived from the data describing exchange and conversion REA processes [4,8]. For example, the data describing the sale event is used in the warehouse management, payroll, distribution, finance and other application areas, without transformations or adjustments.

The quality of a database application depends crucially on its design, see [13]. To ensure correctness, clarity, adaptability and productivity, information systems should be specified at the conceptual level first, using concepts and the language that both designers and customers can easily understand [6]. Object-Role-Modeling (ORM) is a fact oriented approach for modeling information at the conceptual level. A fact is a particular arrangement of one or more objects. Depending on the number of objects that are involved in a fact, we speak about unary, binary, ternary, etc., facts. An example of unary fact is that *Vendor is a Person*. Another example of binary fact is that a *Customer receives a Pizza*. Unlike traditional approaches, ORM make no use of attributes as a base constructs, instead expressing all facts types as relationships [6]. This attributes free-approach leads to greater semantic stability in conceptual models and enables ORM fact structure to be directly verbalized and populated using natural language sentences.

The ORM method provides a more precise way to capture and validate data concepts and business rules with domain experts. ORM diagrams simply capture the world in terms of objects (entities or values) that play roles (parts in relationships) which forms a fact. ER notation as well as UML notation allows relationships to be modeled as attributes. ORM models the world in terms of objects and roles, and hence has only one data structure – the relationship type. As a consequence, ORM diagrams take up more room than corresponding UML or ER diagrams. The aim of the paper is to find out a "semantic" connection between REA business process model and fact-based model utilizing ORM approach.

The structure of the paper is as follows. Section Two describes REA modeling approach. Concise possibilities of the ORM modeling method are mentioned in Section Three.

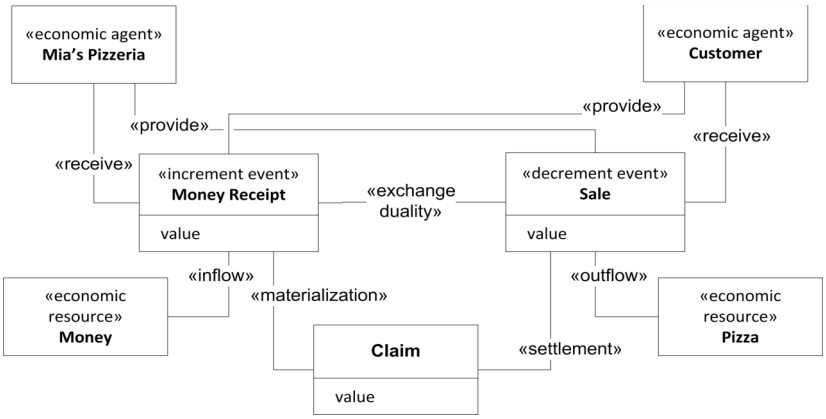


Fig. 1 REA core pattern example

ORM model of paired transactions is described and illustrated in Section Four. Discussion of the results is mentioned in Section Five and conclusion is contained in Section Six.

II. REA MODELING APPROACH

The REA ontology is based on the REA core pattern that expresses the basic principle [2,4]. The fundamental entities of the pattern are different events that are involved in various transactions which have something in common; there has always been a *decrement economic event* (one in which something is provided) and an *increment economic event* (one in which something is received). Apart from economic events, the economic agents that represent human beings partake in the exchange process. Resources are entities which are kept track because their property rights can be exchanged or they can be converted to create new resources. As the mutually bound events cannot happen at the same time, the claim entity is utilized for deferred revenue, prepaid expenses, accounts payable and so on. The REA core pattern is illustrated as a Pizzeria shop in Fig. 1. The economic events in REA models usually encapsulate properties for *date, time* and *location* in space.

The REA model is an extension of the REA core pattern. The principal feature of the REA modeling approach is that it explicitly distinguishes between past and current events and events performed in the future for which it introduces the commitment entity. The relationships of *committed provide* and *committed receive* mean that some agreement about the future exchange has to be achieved between economic agents. The commitment entity addresses the issue of modeling promises of future economic events and the issue of reservation of resources. Commitment entities and their relationships with other entities are shown in Fig. 2. To a considerable extent, the commitment entity copies the structure of the event entity, by which we mean the existence of an increment and decrement commitment and the exchange reciprocity relationship. The exchange reciprocity relationship between the increment and decrement commitments identifies which resources are promised to be exchanged for which other resources.

Each commitment is related to an economic resource by a reservation relationship which specifies which resources will be needed or expected by future economic events. The reservation relationship between the resource and commitment represents obligation of economic agents to provide or receive rights to economic resources in exchange processes and represents scheduled usage, consumption or production of economic resources in conversion processes.

The most important relationships of the REA model are the *exchange reciprocity* and *exchange duality* relationships, by [5,9,10,11]. The exchange reciprocity relates a pair of an increment and decrement commitment entities. The exchange reciprocity relationship identifies which resources are promised to be exchanged for which others.

The exchange duality relationship which relates corresponding increment and decrement economic events keeps track of which resources were exchanged for which ones.

III. ORM CONCEPTUAL MODELING METHOD

Object-Role Modeling is a conceptual modeling method that views the world as a set of objects that play roles (parts in relationships) according to [6]. For example, you may play a role of walking in the country (a unary relationship involving just you) or you may play a role reading this paper (a binary relationship between you and the paper). Thus a role in ORM corresponds to an association-end in UML, except that ORM also allows unary relationships. Object-Role Modeling is a conceptual modeling method that views the world as a set of objects that play roles (parts in relationships) according to [6].

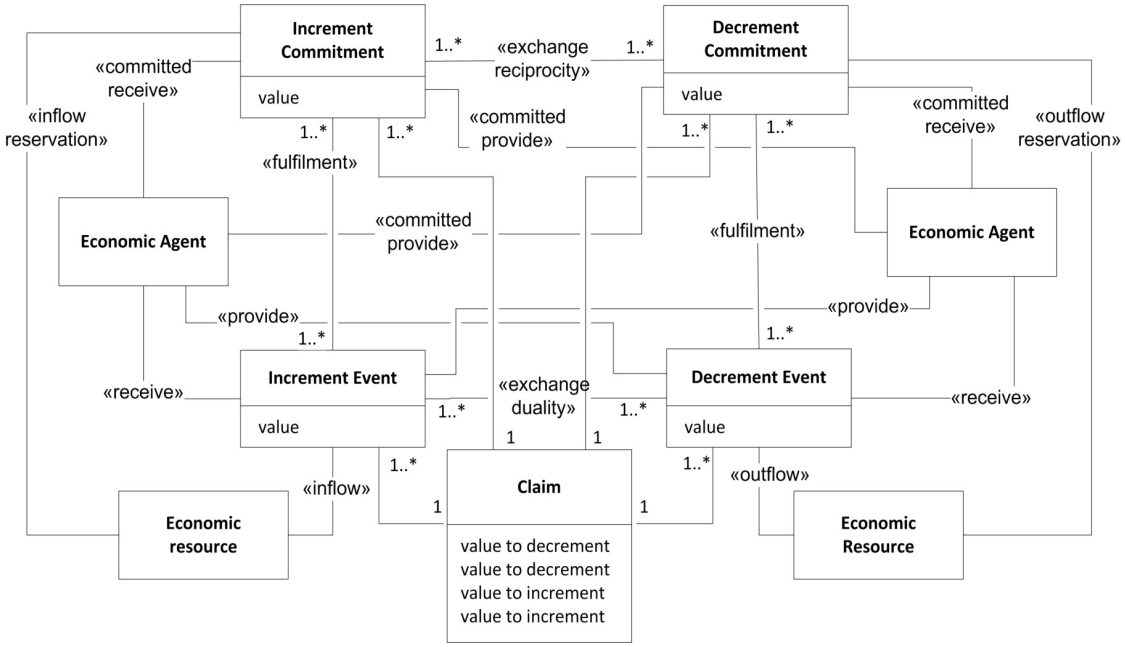


Fig. 2 REA model. Adapted from [1]

The main structural difference between ORM and UML is that ORM excludes attributes as a base construct and treats them instead as a derived concept. The conceptual schema using ORM specifies the information structure of the application in the forms of: *fact types* that are of interest; *constraints* on these; and *derivation rules* for deriving some other facts.

A fact is a proposition that is taken to be true by the relevant business community. A fact type is a kind of fact that may be represented in the database [3]. The constraints represent constraints or restrictions on populations of the fact types. The derivation rules include rules that may be used to derive new facts from other facts, see [6,12].

The ORM model (left part of Fig. 3) indicates that employees are identified by their employee numbers. The top three roles (*EmpName*, *Title* and *Sex*) are mandatory roles. This is indicated by the black dots at the *Employee* box. The other black dot where two roles are connected (at the bottom of *Employee*) is a disjunctive mandatory role constraint indicates that an employee must have a social security number or a passport number or both. The uniqueness of constraints (cardinalities in UML) indicates vertical lines over roles. In Fig. 3 it means that *empNr*, *EmpName*, *Sex*,

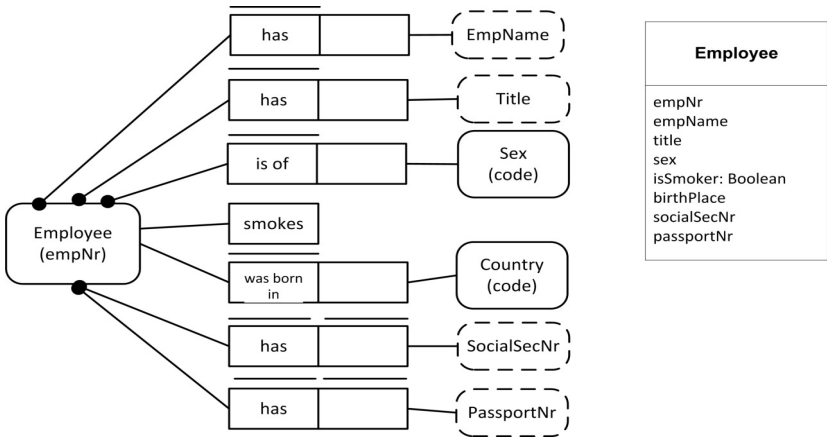


Fig 3. ORM and UML models of Employee

and *Country* is unique for each employee. Two vertical lines over each roles (*SocialSecNr*, *PassportNr*) indicating that each employee number, social security number and passport number refers to the one employee at most. The dashed line over e.g. *PassportNr* indicates that this is a value not an object.

Graphically, object types are depicted as named boxes (solid for entity types, and dotted for value types). As in logic, a predicate is a proposition with object-holes in it. In ORM, a predicate is treated as an ordered set of one or more roles, each of which is depicted as a box, which may optionally be named. A fact type is formed by applying a predicate to the object types that play its roles.

IV. ORM Model of Paired Transactions

The ORM model of paired transactions is composed of two kinds of transactions (left hand side, right hand side). These paired transactions actually represent result of an exchange process in which some resources were exchanged for other resources. The cardinality between exchanged resources is in general many-to-many as was stated in the REA model, see Fig. 2. Despite the REA model, the ORM model of paired transactions contains only one relationship that joins both kinds of transactions which is called the duality relationship. The name *duality* expresses the final state of an exchange process. The duality has mandatory relationships to both kinds of transactions. The left-hand side transactions represent a transfer of goods and the right-hand transactions stand for a transfer of money. We consider the common case of transfer in which resources or services are exchanged for money. These two transfers are depicted by object classes with the same name. The object class *Goods Transfer* is related to two kinds of actor roles the *vendor* and the *customer* who both belong to the object class *Person*. There is a relationship *exclusive or* between two actor's roles which means that these actor roles have to be different persons. The *customer* is an actor role who receives goods in *Goods Transfer*. The *vendor* is an actor role who provides the goods for the transfer. The corresponding object class *Money Transfer* is related to the *payer* and the *cashier* actor's roles. Between both actor roles there is a relationship *exclusive or* with the same meaning as in the previous case. Both *payer* and *cashier* belong to the object class *Person*. It is important to use a proper actor role. The *customer* can be e.g. wife and the *payer* can be her husband.

The object class *Goods Transfer Contracted* is a subclass of the object class *Goods Transfer*. This relationship between these object classes means that there must exists at first *Goods Transfer* object class and subsequently it may happen that the object class *Goods Transfer Contracted* becomes existent. Conversely, *Goods Transfer Contracted* cannot exist without the object class *Good Transfer*. This point is very important and differs from the REA model. The object class *Goods Transfer Contracted* means that the *customer* promised that he would receive the goods and the *vendor* promised that he would deliver the goods. But as we are talking about paired transactions there must be the other transaction(s) because paired transactions mean that some resources are transferred in consideration of the other resource transfers.

The other transaction(s) of the paired transactions is represented by the object class *Money Transfer Contracted*. In this case, the *payer* promised that he would pay for the goods and the *cashier* promised that he would accept the amount of money (resource). At this point it is essential that the state *promise* is reached on both object classes *Goods Transfer Contracted* and *Money Transfer Contracted*.

Each object class (*Goods Transfer Contracted*, *Money Transfer Contracted*) contains property types of contracted goods kind and contracted location. The object class *Money Transfer Contracted* has the same kinds of property attributes and scale attributes. These property attributes and scale attributes represent facts that were contracted. The construction of this model enables that the resource kind can occur more times. For instance a *customer* would like to buy the given number of a specific pizza types, a certain number of cola kinds and a certain number of chocolate kinds. In general, *Money Transfer Contracted* may represent some kind of payment before the purchase, payment after delivery and possibly payment in installments.

Goods Transfer Completed is the next object class that is a subclass of *Goods Transfer Contracted* object class. The meaning of the subclass relationship is as follows: a transfer have to be contracted at first and then it can be completed. The property attributes and the scale attributes are the same as in *Goods Transfer Contracted* but in this case the property attributes express the real values. In the detail insight, the proposed solution enables differences between the individual number of *Goods Transfer Completed* and the number of *Goods Transfer Contracted*.

Goods Transfer Contracted represents planned transactions. The delivery, which is performed in *Goods Transfer Completed* is usually performed in several shipments. The corresponding object class to *Goods Transfer Completed* is the object class *Money Transfer Completed*. This object class property attributes deal with real payment transactions which means that they deal with the real installments. The business rules are stated in the contract which comes into existence when the transactions are contracted. The structure of the property attributes is the same as the structure of *Money Transfer Contracted* but the real values may be different. All this recorded information is required in database solutions.

V. Discussion

Designing an information system involves building a formal model of the application domain which requires a good understanding of the application domain and utiliziation of the proper tools for modeling specifications in a clear and unambiguous way. ORM simplifies the design

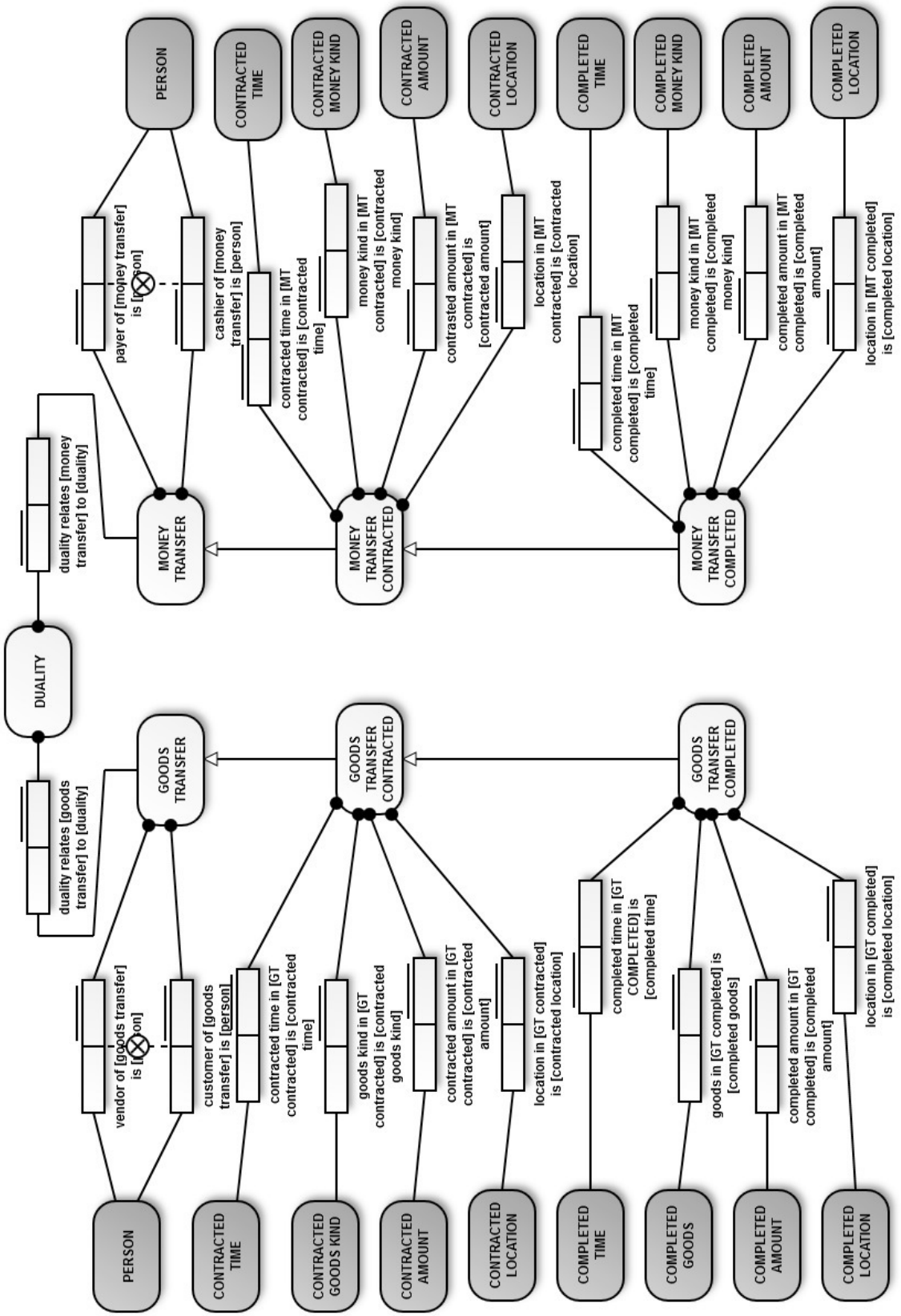


Fig. 4 ORM model of paired transactions

process by using natural language, and by examining the information in terms of simple and elementary facts. By expressing the model in terms of natural concepts, like objects and roles, it provides a conceptual approach to modeling.

The REA modeling approach addresses the paired transactions model in the REA core pattern and in the REA model. The REA core pattern corresponds to accounting model and captures events which were completed. The REA model is more general than the REA core pattern and provides possibilities to address future events. However, connection between the commitment entity and the event entity in the REA model is insufficiently consistent. This is done by the fact that both commitment and event entities represent production and that the REA model doesn't have a proper state machine, see [9]. For these reasons it is difficult to distinguish the phases in which the paired transactions were *contracted* (promised) and *completed*. When comparing the REA core pattern with the REA model it is evident that the REA model covers all operations of the REA core pattern despite the fact that the commitment actions are only formal.

The ORM model of the paired transactions enables clear distinguishing between the *contracted phase* and the *completed phase*. This is done by utilization a sub-classing mechanism of the ORM modeling approach. The corresponding object classes *Goods Transfer Contracted* and *Money Transfer Contracted* express the *contracted* (planned) property types and attributes types which are essential to be stored in the database solution. In the same way, *Goods Transfer Completed* and *Money Transfer Completed* capture the real value of property types and attribute types of finished paired transactions.

The object classes *Goods Transfer* and *Money Transfer* form the beginning of the paired transactions process which means that they identify partaking actor roles which will be involved in the process. They have to be created first. After that, the object classes *Good Transfer Contracted* and *Money Transfer Contracted* can be created. Similarly, existence of *Goods Transfer Completed* and *Money Transfer Completed* is dependent on the existence of contracted transfers.

VI. Conclusion

The paper deals with paired transactions modeling utilizing the ORM conceptual modeling method. The benefits of this mathod can be sumarized as follows. This method and modeling approach enables to explicitly distinguish the contracted phase from the completed phase of the paired transactions model. It also ensures unified transaction processing which means utilizing only contracted and completed phases. The proposed solution also eliminates usage of the claim temporal entity. Future research will cover implementation, verification and validation of the proposed ORM model.

Acknowledgement

The paper was supported by the grant provided by Ministry of Education, Youth and Sport Czech Republic, reference no. SGS15/PRF2016.

References

[1] G. L. Geerts, and W. E. McCarthy, “The Ontological Foundation of REA Enterprise Information Systems”. Paper presented at the Annual Meeting of the American Accounting Association, Philadelphia, PA., 2000.

[2] McCarthy, W.E. “The REA Accounting Model: A Generalized Framework for Accounting Systems in A Shared Data Environment.” The Accounting Review (July) 1982, pp. 554-578.

[3] J. L. G. Dietz, “The Essence of Organization. An Introduction to Enterprise Engineering”. Sapio bv, 2012.

[4] Ch. L. Dunn, O. J. Cherrington, and A. S. Hollander, Enterprise Information Systems: A Pattern Based Approach. New York: McGraw-Hill/Irwin, 2004.

[5] Hruby P., Model-Driven Design Using Business Patterns. Springer-Verlag Berlin Heidelberg, 2006.

[6] Halpin, T. A., Morgan, T. Information Modeling and Relational Databases. San Francisco: MorganKaufmann, 2nd edt 2008.

[7] Dudycz, H., Korczak, J. “Conceptual design of financial ontology”, Proceedings of the 2015 Federated Conference on Computer Science and Information Systems. pp. 1505-1511. DOI: 10.15439/2015F162

[8] F. Hunka, and J. Zacek, “Detailed Analysis of REA Ontology”, Lecture Notes in Busines Information Processing, Vol. 174, 2014, pp. 61-75. DOI: 10.1007/978-3-319-06505-2

[9] Hunka, F., Zacek, J. A new view of REA state machine (2015) in Applied Ontology, 2015, vol. 10 (1), pp. 25-39.

[10] R. Klimek and P. Szwed, “Verification of ArchiMate Process Specification Based on Deductive Temporal Reasoning”. Proceedings of the 2013 Federated Conference an Computer Science and Information Systems. pp. 1103-1110.

[11] Korczak, J., Dudycz, H., Dyczkowski, M. “Design of Financial Knowledge in Dashboard for SME Managers”. Proceedings of the 2013 Federated Conference an Computer Science and Information Systems. pp. 1111-1118.

[12] Kersten, G., Wachowicz, T. “On Winners and Losers in Procurement Auctions”. Proceedings of the 2014 Federated Conference an Computer Science and Information Systems. pp. 1163-1170. DOI: 10.15439/2014F271.

[13] Paweloszek, I. “Approach to Analysis and Assessment of ERP System. A Software Vendor's Perspective”. Proceedings of the 2015 Federated Conference an Computer Science and Information Systems. pp. 1415-1426. DOI: 10.15439/2015F251.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1159–1168
DOI: 10.15439/2016F118
ACSIS, Vol. 8. ISSN 2300-5963

Comprehensive Methods of Evaluation and Project Efficiency Account

Anna Kaczorowska
University of Lodz
Faculty of Management,
Department of Computer Science
ul. Matejki 22/26, 90-237 Lodz,
Poland
Email: annak@wzmail.uni.lodz.pl

Jolanta Słoniec
Lublin University of Technology
Faculty of Management,
Department of Enterprise
Organization, Nadbystrzycka Str.,
38, 20-618 Lublin, Poland
Email: j.sloniec@pollub.pl

Sabina Motyka
Cracow University of Technology
Faculty of Mechanical, Department
of Manufacturing Processes,
M6 Institute, al. Jana Pawła II 37,
31-864 Cracow, Poland
Email: motyka@mech.pk.edu.pl

***Abstract*—Project management is one of the main areas of contemporary organization management. This article is aimed at analysis of existing tools of project assessment – mainly comprehensive methods, but also partial techniques - and referring them to undertakings efficiency calculation which comprises cost-effectiveness evaluation, risk analysis, and investment decision taking. Business case (BC), as one of the comprehensive methods of project assessment, is simultaneously the main subject in PRINCE2 (PRoject IN Controlled Environments). It is for this reason that the case study for BC is based on this recognised method of project management. This article contains also the proposals of solutions of discerned problems within project assessment.**

I. Introduction

SITUATIONS of assessment occurring in project management are largely diversified and require the use of appropriate methods adjusted to each of them. A correctly and reliably carried out assessment will support the manager in taking conscious decisions as to the projects which should be invested in.

Speaking about assessment we usually mean definition, estimation, and valuation of something. Assessment in relation to projects means the „functional value, i.e. the whole of the project features determining its ability to meet specific needs" [9].

Choosing any tool for project assessment we expect that it will first ascribe to them specific parameters of assessment, second – classify projects, third – enable a comparison of projects, and fourth – enable following the progress of project works.

Each project should cause positive effects consisting in generation of profits and negative effects related to incurring of inputs to obtain new values. Therefore, the main components of project assessment are inputs and profits. Such perspective combines the most important aspects of project assessment, i.e. assessment of its result and assessment of its course [2].

The final results of the projects are always an outcome of inputs and profits. The inputs are treated comprehensively, i.e. as the consumption of all sorts of means – both countable and such which cannot be expressed in monetary units. Benefits are understood as various positive effects of projects and may be nominal, tangible and intangible.

A summary evaluation independent of the type of the project is the total result which is a difference of benefits and inputs. If the benefits and inputs may be expressed monetarily, the total result of assessment is profit. Many techniques have been worked out for this case. They are called the investment account techniques, because they are used in assessment of investment undertakings. The following techniques of investment undertakings assessment may be singled out:

- simple techniques of absolute assessment – e.g. payback period (PP), accounting rate of return (ARR),
- discount techniques of absolute evaluation – e.g. gross present value (GPV), internal rate of return (IRR), profitability ratio (PR), techniques of discounted period of return and their modifications [14],
- techniques of relative efficiency account [14].

Commercial projects are assessed using the monetary techniques of assessment:

- simple – such as : simple rate of return (return on equity - ROE, return on investment - ROI), PP, ARR,
- comprehensive (discount) – discounted payback period (DPP), net present value (NPV), IRR, modified internal rate of return (MIRR) [14]-[15]-[16].

The project in which only inputs may be expressed monetarily requires different techniques of assessment, such as: cost benefit analysis and cost effectiveness analysis. The advantage of those techniques is shown in points as a result of multi-dimensional assessment in points. When in the cost benefit analysis the benefit estimated in points includes the probability of its occurrence, it becomes a cost effectiveness analysis.

When neither benefits nor inputs may be expressed monetarily, the total result of assessment is profitability. In this case it consists of their determination by multi-criterial assessment in points and comparison of the obtained values, e.g. as the quotient of the value of profits in points and the point-wise value of input.

If the profit may be expressed monetarily, and the input is not considered (little importance, problems with estimation), the projects are assessed according to the account of incomes. Otherwise, when benefits are not considered and the input may be estimated in monetary units, assessment is based on the costs account.

When inputs can be determined neither monetarily nor non-monetarily, assessment is made as an analysis of the project's functional value or as analysis of effectiveness.

Insignificance of benefits during project assessment and possibility of only non-monetary expression of input cause the need of the point assessment of the input value.

Taking an investment decision within the project efficiency account may refer to [1]: a single project (absolute or time-related decision), variants of the project or projects competing for common limited resources, including the capital (relative or portfolio decision).

II. Classifications of Project Assessment Methods

The project assessment methods constitute an extensive collection where various groups may be singled out due to: application area, scope of problems, details of recommendations.

Table I presents classifications of project assessment methods, including the mentioned items as the division criteria.

TABLE I.
CLASSIFICATIONS OF PROJECT ASSESSMENT METHODS

Classification criterion of assessment methods	
APPLICATION AREA	
Universal	Special
DETAILS of RECOMMENDATIONS	
General	Detailed
SCOPE of PROBLEMS	
Comprehensive	Partial

In project assessment both universal methods (usable in assessments of any type) and special methods (usable in project management assessments, e.g. the Earned Value method – assessment of project implementation advancement) may be applied.

Methods containing general recommendations are considered as general methods of project assessment, whereas those which present recommendations precisely and accurately are determined as detailed ones.

Partial methods refer to partial processes and problems within the project assessment. They are called techniques and may be used in various stages of project management and in different phases of its life cycle.

Comprehensive methods comprise with their recommendations the whole process of project assessment. These methods of project assessment comprise: feasibility studies (hereinafter referred to as FS), business plans of the project, business cases (BC), and cost benefit analysis. The characteristics which should describe the methods to make them considered as comprehensive were presented in Table II.

TABLE II.
SPECIFICITY OF COMPREHENSIVE METHODS OF PROJECT EVALUATION

Characteristics	**Description of characteristic**
Completeness	Detailed description of all issues important for project estimation
Perspectiveness	Including into the description the whole period of preparation, performance and use of the project
Accuracy of assumptions and reliability of data	Data are derived from trustworthy sources and enabling the preparation of reliable evaluation results
Internal compliance	Subordination of individual components to assumptions and their non-contradiction
Reality	Reality-consistent presentation of circumstances affecting implementation of the project and its solutions
Variants	Analysis of possible solutions in several variants
Flexibility	Predispositions to introduce amendments, changes and supplements connected with inflow of new information
Operations	Possibility to translate the method description into concrete decisions
Comprehensibility	Adjustment of the contents, volume and form of method description to recipients' needs and requirements
Communicativeness	Explicit and comprehensible transmission of the contents of the method to all potential users
Extensiveness of the stage of determining the assessment objectives	In case of FS and BC this stage consists even in examining the circumstances of future functioning of assessment objects and designing them according to these conditions

All comprehensive methods of the projects are in many elements similar to each other but due to the differences occurring between them they were discussed in separate parts of the article.

III. Feasibility Study

A. Description of the method

Feasibility study is defined as:

- "A short, preliminary study undertaken to assess the validity of a full-scale project" [17],
- formal study to determine the probability of success of a particular project or achieve a particular result [18].

"The purpose of the feasibility study tool is to identify whether the concept of a project is viable." The study contents seven parts: "executive summary, background information, description of current situation/problem, description of proposed idea, project timelines, feasibility review board, go/no-go decision. Dow mentions "types of

feasibility studies used today: schedule: analyze how long it will take a project or process to complete and what go/no-go decision point will be; organizational: analyze what the impacts would be on costs and resources if the company decides to reorganize their current resources base; legal: analyze whether it is worth pursuing a particular litigation; technical: analyze if a technical concept will work within the current environment; cultural: analyze whether offshore teams and onshore teams can communicate and effectively execute a project; construction: analyze if it is cost effective to construct a building, based on the height and location; environmental: analyze whether the environmental conditions at the particular location are suitable" [19].

The methodology of creating feasibility study is similar to the methodology of problem solving. It has also much in common with the methods of scientific work, and it consists of the following parts: precise problem definition, project limits indication, identify the characteristics and functions of a good solution to the problem, description of alternative solutions, ranking of alternative solutions, conclusions from the analysis of alternative solutions and recommendation the best solution, determine the timing and expected costs of the project [20].

FS is one of the complex methods of project evaluation and it is used to check to what extent, by what means and at what time the project can be realized. The method consists of five parts: technology and system feasibility, economic feasibility, legal feasibility, operational feasibility, schedule feasibility. It should be also taken into account: market and real estate feasibility, recourse feasibility, cultural feasibility [21].

The number of parts of the method is different depending on the author (from five to eight), and each of them could be used in some types of projects.

By Newton "feasibility is the review of the project in more detail" [22]. This method may include activities such as technical tests, market research and assessment of the impact of the project on the organization. After the completion of the feasibility study the cost, duration and the outcome of the project should be known.

"Feasibility studies (…) reduce the risk of incorrectly accepting or rejecting a project. Although it can be considered as a stage in a project's lifecycle. (…) Feasibility studies have a cost and absorb resources, so in turn they must be reviewed and prioritized. (…) The decision whether to undertake a feasibility study is a trade-off between the cost and the value of the information determined [22].

The feasibility study should be carried out to provide the information needed to determine whether a project should proceed. It includes: technical feasibility to determine whether a project will successfully create the expected deliverables, commercial feasibility to determine whether a project will achieve its business case, market feasibility (usually in the case of new products) to determine whether the business's customers will buy the product, organizational feasibility to determine the operational impact of making the change resulting from a project successfully, exploring requirements and designs to produce more accurate plans, costs and resource profiles for a project, exploring project options.

Feasibility studies allow you to determine beyond any doubt whether the problem can be solved at all, or whether it is possible to use the opportunity. Usually they are created for the management board. Feasibility study consists of eight parts: abstract, defined business problem or business opportunity, requirements and purpose of the study, description of evaluated options, assumptions made in the study, users affected by introductions changes, financial commitments, recommended procedures [23].

As you can see there are a lot of definitions of feasibility studies, but the most important characteristics of this study is that it is a short, complex, preliminary study, made for the management board, undertaken to assess the validity of a project.

B. Reference to project efficiency

The efficiency of the projects in relation to the projects is defined as a relationship between total expenditure and the effects and the evaluation of the results concerning their utility, which answers the question of whether the information needs of users of information system have been met. Three main streams of IT projects evaluation have been classified [24]:

- technical and functional trend, which assumes that the effects of investment in information systems are short-term and have no connection with the business strategy; assumptions of this trend are correct in relation to simple automation systems,
- economic and financial trend, which treats IT investments as aimed at increasing business efficiency or extend it; the evaluation moves here from the project treated in isolation to the quality of its products or services provided to internal and external customers; techniques assessments of projects derived from the management of value may be used herein among others,
- trend of possible interpretations, resulting from the specificity of investments; it takes into account the entire life cycle of the project, including the expenditure and benefits in its course, and the emphasis is on decision-making context in which the project exists.

Table III contains a comparison of these trends [25].

TABLE III.
COMPARISON TRENDS OF EVALUATION IT PROJECTS

Dimension	Trend		
	Technical and functional	**Economic and financial**	**Possible interpretations**
Objective	Technical efficiency, IT resource control, cost of maintaining the system	Quality and rate of utilization the system and the effects of its introduction	Solutions sensitive to context, learning of organization

Object of the evaluation, used criteria	IT system, automation, cost reduction	Product of IT system: productivity, enterprise value, user satisfaction	Wallets of IT systems, measurement of indirect effects
Time horizon	Ex-ante and ex-post investment, system lifecycle	Ex-ante and ex-post in relation to the life cycle of the system	Perpetual effects management
The role of people in the evaluation process	IT experts	IT experts, financial managers	Participants in the evaluation process, internal and external customer of IT services
Used methodology	Related to the quality and costs	Orientation to economics, financial and behavioural	Development of meta-methodology
Assumptions	Cost efficiency	System efficiency	Understanding the problem

Currently multi-criteria projects evaluation methods are often used. They include [26]: scoring method, AHP method, PROMETHEE-II method.

Scoring method is the most frequently used multi-criterial method. It allows you to list the ranking of decision variants based on the score, which projects received in each category. There are different versions of the method, simple and complex. Simple version assumes that each criterion has the same maximum number of points. The final evaluation of variant is the sum of all the points awarded in all criteria. The complex method involves the granting of weight to each criterion due to its validity. Criteria can be also described as linear functions. An important assumption of the method is that the criteria are independent in terms of preferences, that means the one criterion does not depend on the value which takes the other one. A similar to scoring method is the SMART method (Simple Multi-Attribute Ranking Technique). It assumes the existence of an additive utility function, which can be described as non-linear function.

AHP method (Analytic Hierarchy Process), like the scoring method, involves weighing of criteria. Rating variant is the sum of ratings of individual criteria. In the method, decision maker deliver his/her opinion on relations between variants. The opinion is expressed verbally using a nine-point scale for comparison. Ranking decision variants in relation to the criteria is calculated as a weighted average of the ratings which options have obtained due to the individual criteria. A limitation of the use of method is that the project must have a small number of variants and variants cannot be dependent on each other. Development of AHP method taking into account the linkages between criteria and feedback relationships between the variants and the criteria is the ANP method (Analytic Network Process).

In the PROMETHEE-II method (Preference Ranking Organization METHod for Enrichment of Evaluation) it is built variants ranking decision as in the AHP method. But the options are compared automatically based on the information provided by the decision maker. The decision maker determines the value of the difference between the variants that one variant was the preferred relative to the other.

Each method is applicable to multi-criteria evaluation of projects in certain specific cases.

C. Case study

In the literature it is described the problem of selecting IT project using AHP method [27]. The company is to be implemented management support system and owners have to choose contractor for the implementation of IT system among three alternative projects. Evaluation of projects is based on the methodology proposed by Parker [28, 29].

The criteria consist of three factors and they make up to maintain the company's competitiveness: the financial contribution (the value to achieve, acceleration, restructuring, innovation, efficiency, productivity, NPV, IRR), support projects providing management information (the reaction of competitors, compliance with the strategy and structure, organizational risk), technological requirements (technological risk, providing innovation, compatibility with existing IT, uncertainty of the construction project).

Assessment of the criteria for each of the remaining criteria is based on a scale proposed by Saaty [30].

TABLE IV.
GRADING SCALE IN THE AHP METHOD

Definition		Factor
Equally important	Activities contribute identically to aim	1
Slightly more important or preferred	Experience and judgment slightly favours one activity over the other	3
More important or strongly preferred	Experience and judgment strongly favours one activity over the other	5
More important or very strongly preferred	Activity is strongly favourably and its dominance is demonstrated in practice	7
Extremely important or more preferred	Evidence favouring one activity over the other is the greatest possible in order affirmation	9
Intermediate values	Expressive identification between two basic values of the scale	2,4,6,8

Table V shows the matrix of relationships between different criteria in the presented case study of multi-criteria selection of the project implementation the enterprise management system (relationships are assessed by the decision-maker). Vector the matrix of domination of criteria allows the estimation of the relative importance of each criterion in relation to the overall objective, which is to maintain the competitiveness of the company.

TABLE V.
MATRIX EVALUATION CRITERIA

Criteria	**Efficiency**	**Business support**	**Future importance**	**Technological risk**
Efficiency	1	1/5	1/7	3
Business support	5	1	1/3	5

Future importance	7	3	1	7
Technological risk	1/3	1/5	1/7	1
Inconsistency = 0.09				
Scale vector	0.092	0.282	0.574	0.052

Next the matrices of evaluations for alternative projects were built (Table VI).

TABLE VI.
MATRICES OF EVALUATION ALTERNATIVE PROJECTS (P1, P2, P3)

Efficiency	**P1**	**P2**	**P3**
P1	1	3	5
P2	1/3	1	3
P3	1/5	1/3	1
Inconsistency = 0.04			

Business support	**P1**	**P2**	**P3**
P1	1	1/5	3
P2	5	1	7
P3	1/3	1/7	1
Inconsistency = 0.06			

Future importance	**P1**	**P2**	**P3**
P1	1	1/9	1/7
P2	9	1	3
P3	7	1/3	1
Inconsistency = 0.08			

Technological risk	**P1**	**P2**	**P3**
P1	1	3	7
P2	1	1	5
P3	1/7	1/5	1
Inconsistency = 0.06			

Finally the evaluation of each project according to each criteria was summarized.

TABLE VII.
EVALUATION OF PROJECTS

Project	Criteria				**Evaluation**
	Efficiency	Business support	Future importance	Technological risk	
P1	0.637	0.188	0.055	0.649	0.178
P2	0.258	0.731	0.655	0.279	0.616
P3	0.105	0.081	0.290	0.072	0.206
Inconsistency = 0.08					

The overall rating of projects allows to establish the ranking of solutions P2, P3, P1 (Table VII). Project P2 has been rated highest, it has a significant advantage over alternative projects P3 and P1. According to the assessment using multi-criteria AHP method, the company should implement a management IT system using the offer of organization which has presented project P2.

The final decision about project execution can be based on some complex method of project evaluation or on project efficiency. Previously discussed feasibility study can include elements specified in chapter IV, at least as followed: technology and system feasibility, economic feasibility, legal feasibility, operational feasibility, schedule feasibility.

Considering which method of projects assessment to choose when we need to choose one IT project among several projects, or deciding to start a new IT project it should be noted that a complex assessment methods of projects evaluation helps us choose the best project among several alternative projects. While the feasibility study of the project is a detailed and comprehensive analysis of a specific project. Developing such a document is expensive and usually it is performed for one project only. In this case study AHP method allows you to choose the best variant of the project. The final decision on the project can be taken on the basis of this analysis. However, you can perform a feasibility study for project P2. Although the feasibility study of the project is costly, it gives us more certainty that the implementation of the project will bring the company the desired results and will be successful.

IV. BUSINESS PLAN

A. Description of the method

Business plan is defined as:

- a plan of launching a business or as an action plan and development of the company; it is a special case of economic plan [31];
- a set of document developed by entrepreneur during preparation process of launching of the new project, showing its strategy and structure [32];
- a description method of a business and evaluation of the prospect of execution [4].

Business plan of the project is a study of the planned economic project which contains: purpose of analysis; circumstances of implementation; assessment of the advisability, feasibility and effectiveness evaluation. Business plan can be used for comprehensive evaluation of projects of small range and complexity [4].

Business plan refers to both results and costs and it shows project as a product market intended for potential customers. Business plan contains: objectives, impacts and process of the project.

Business plan is characterized by [4]:

- specificity – business plan must include: detail information of the: market, enterprise, project and planned actions,
- comprehensiveness - business plan must include all aspects of the planned activities (market, technical, personnel, organizational and financial activities),
- long-term effect – business plan is mostly a long-term plan, rarely annual.

Structure of the business plan depends on its purpose. General scheme of business plan includes four elements: general information – abstract, market conditioning of the project, operation conditioning of the project, assessment of the project's consequences (Table VIII; [33]).

TABLE VIII.
GENERAL SCHEME OF BUSINESS PLAN

No.	**Description of the item**
1	**General information** - front page and table of contents - subject of the business plan - basic information - performer of business plan - current events and approvals **Abstract**
2	**Market conditioning of the project** - users and sponsors of the project - stakeholders of the project

	- external and internal conditions of the project - expected benefits of the project
3	**Operation conditioning of the project** - tasks and process of the project - employment, equipment and resources of the project - subcontractors and suppliers of the project - organization and management of the project - costs of the project (budget)
4	**Assessment of the project's consequences** - benefits - expenses and costs - risk analysis - evaluation (economical) - evaluation indicators - recommendations

An exemplary structure of business plan of venture investment is presented in Table IX.

TABLE IX.
AN EXEMPLARY STRUCTURE OF BUSINESS PLAN OF VENTURE INVESTMENT EVALUATION CRITERIA

No.	Chapter
1	Abstract
2	General information about project
3	Planned location of the investment
4	Market analysis
5	Competetive analysis
6	Marketing strategy
7	Technology of production
8	Costs of production
9	Organizational plan
10	Project schedule
11	Project budget
12	The financial part
13	Summary

Business plan of the project starts with an abstract in which the most important conclusions and a summary of entire plan should be included. This part of the plan should be designed very carefully as it is always read in detail during pre-selections of the projects.

Second chapter should contain basic information about the project: name of the project, information about participants, localization and information about lead organization.

Actual state supply, state demand and actual prices of product or services on the market are shown in the market analysis. This part of analysis is, in most cases, based on forecasts as investment can be realized in a new sector.

The fifth chapter gives information about main market competitors by analyzing their strengths and weaknesses. The marketing strategy includes main components of market activity such as: product, prices, distribution and promotions and can be found in chapter number six.

In a description of technology of production main steps of technological process and evaluation of its modernity should be characterized. Technology of production can be original investor's solution or a typical solution. This chapter should also include environmental legislation and information about certificates and licenses acquisition.

Determination of equipment's life time, development of modernization plan and repair plan are also necessary. Production is seasonal in many types of business therefore, the business plan should answer the question on how seasonality affects financial results (incomes and expenses). Chapter number eight covers estimation of production's costs or service delivery.

Organizational plan determines how the investment will be managed in the implementation phase and in the operational phase. Organizational scheme is presented in the chapter which shows work division and presents organization of key functions, like: supply, production and distribution.

Investment implementation schedule is described in chapter number ten. It should include all main investment tasks with the assume time of their implementation.

Detailed budget project should be developed in 11th chapter.

Economic and financial analysis is based on the financial plan and contains financial evaluation, profitability evaluation and risk assessment.

The most important data and conclusions are included in the last chapter.

Responsibility problem is associated with evaluation of the business plan. Business plan is used mainly for certain decision-making which results in specific measurable amount of expenses. Comprehensiveness is the most important advantage of business plan. Significant disadvantages include statistics and descriptive form of business plan.

B. Reference to project efficiency

If a long term calculation is assumed, every project can be consider as effective. Therefore, more precise definition of efficiency is the ratio investment results in a specific period of restoration capital time to investment's expenditures.

The investment efficiency account is a very important part of business plan which can be determine in a pre-investment phase (ex ante effectiveness account) and in operational phase (ex post effectiveness account). An ex ante effectiveness account is more important, from the point of view of business plan preparation, as it is a basis for investment decisions.

Assessment of investment effectiveness is carried out with the use of several tools which include:

- statistical methods for assessing the effectiveness of investment: payback investment rate, payback/return period,
- dynamical methods for assessing the effectiveness of investment: updated net value, discounted payback investment rate, discounted payback/return period, internal rate of interest.

Simpler statistical methods do not include variable time value of money which is why these methods are accounted as less useful.

The last element of the economic and financial analysis of the project is financial risk assessment of investment. Projects that are financially viable may be financially too risky to be selected for implementation. Risk covers many

types of risks, including: the risk of contract, the risk of supply, investment risk, financial risk, credit risk, operational risk, transportation risk, insurance risk, price risk, currency risk and inflation risk.

Each business plan contains many assumptions about investment, costs, prices, demand, inflation etc. Effectiveness of planned investment is determined under assumptions which do not have to be confirm.

Investor who develops business plan should identify factors of financial risk and should also attempt to measure the risk. Financial risk can be measure by following methods: the scenario method, sensitivity analysis, analysis of the breakeven point.

Final evaluation of business plan comprises four elements: assessment of formal correctness, assessment of methodological correctness, assessment of accepted data, formulated assumptions and accounting evaluation.

Common errors and data manipulation in business plans mostly rely on [31]: overpricing, costs reducing, skipping some expenditure and costs, the discount rate reducing, inflate production capacity, the need of current assets reducing, accepting unrealistic growth rates of production and sales.

Business plan usually consists of four elements (Fig. 1).

I **Informational part**	**II** **Market's part**
III **Operational part -** **Expenses and costs**	**IV** **Evaluation – comparison of benefits and expenditures**

Fig. 1 Business plan

First part, informational part presents subject of business plan, performers and basic data of the project. The second - market's part contains internal and external benefits of the project. The third part determines tasks, employment and management styles of the project. The result of the third part is to determine expenditures and costs developed as a project budget. The last part of business plan is a comparison of benefits and expenditures of the project with the use of: cash flow statement, balance sheet, or income statement. Business plan assessment methods depend on needs.

V. BUSINESS CASE AND COST BENEFIT ANALYSIS

A. Description of business case

Business case is a document prepared usually in organization which performs the project for its own needs. It may be prepared for every project, to support its planning and decision-taking. It is a method of assessing the business benefits of analysed project. During the project performance this method is used to analyse the impact of obtained partial results on expected business benefits. Analyses used in business cases are mainly of quantitative nature.

Based on estimated costs, benefits and savings as well as risks, BC involves all changes in business area which the project affects and description of the causes of the undertaking.

In practice it is mostly used in the ICT sector (where all benefits from implementation of information systems are necessary) and sporadically for social projects, although e.g. in Great Britain this document is obligatory for all projects undertaken by public entities.

The following types of business case are singled out [4]: strategic business case – general document indicating the relations between the project and organization strategy, full business case – document containing detailed plans of project performance and cost benefit analyses, ongoing business case – document precisely determining the inputs on works (usually within the nearest stage of the project).

BC does not have any strictly defined and mandatory structure, as it depends on the project specificity, i.e. its duration, budget, branch in which the employer is functioning etc. M. Trocki [2] repeats (after Taschner) that BC should consist of eight parts presented in Table X.

BC is one of the fundamental terms in PRINCE2. It plays a strategic role in the project assessment process, because it is assumed that the undertaking may last as long as this document gives affirmative answers to questions about the need, feasibility, profitability and cost-effectiveness of investing into the project.

It is an element of documentation which initiated the project. It gives rise to the decision to start, continue, interrupt or completely withdraw from the project. BC is being updated throughout the project life cycle [3]-[11].

BC consists of two documents:

1. Outline of the business case (introduction and basic information about the project: title of the project and possibly subtitle, author of the document and its recipient, date of the document preparation and submission, generally determined BC area, defined goals of the project, purpose of BC preparation).
2. Detailed business case (DBC) – its most important element is assessment of the project cost-effectiveness.

A very important element of DBC is definition of the pattern according to which the project implementation variants will be described, because BC should contain a description of several possible methods of achieving the same goal. The indicated methods should enable a comparison of the variants of this undertaking [5].

The basic element of DBC is presentation of possible variants of the project performance. Each of the variants requires a detailed feasibility study [6].

According to the PRINCE2 (2009) method the project requires first of all focussing on business aspects starting from the reasons for which it was launched till it was closed [10].

TABLE X.
PARTS OF BUSINESS CASE

Part name	Description
Macroeconomic part	It comprises the most important macroeconomic information influencing the project, subjecting it to analysis and formulating conclusions
Tax-related part	It contains tax analyses necessary for assessment of the project
Financial part	It involves determination of the project financial needs, indicating and choice of financing sources and preparing the cash flow plans
Market part	Market analyses connected with the result and conditions of its market performance
Marketing part	Market analyses are the starting point for marketing surveys determining market effects of project implementation
Operational part	It describes technical conditions of project implementation and its effects
Investment part	Information from all parts is focused in investment part, where the project's financial assessment is made, using the investment efficiency account method
BC results	Each BC should refer to free and comprehensive decisions consisting of the choice from at least two alternatives; the decision consequences should wholly or largely be expressed in monetary units; importance of the decision consequences should determine the inputs on BC; The last part of BC should contain: a short summing up of the project goals, expected benefits, necessary inputs, the most important risks and list of recommendations related to the project implementation formula

The main and obligatory element of business case is a list of expected benefits and optional components of the cost benefit analysis and assessment of investment.

B. Description of cost benefit analysis

Cost benefit analysis (CBA) consists in determining and comparing the expected costs and benefits from variants of the project to choose the best and most profitable variant.

The cost benefit analysis usually consists of the following parts [7]-[12]-[13]: identification of the project, defining the goals, feasibility study, economic analysis, multi-criterial analysis, other criteria of evaluation, sensitivity analysis and risk evaluation.

The results of the cost benefit analysis are often presented in the so called cost-benefit matrix [12] (Fig. 2).

The projects placed - in result of analysis - in area I usually are not referred to performance. The possibility to perform the area II projects (they lead to significant benefits but simultaneously require quite high inputs) is limited, because they require special conditions to be performed, e.g. foundation of consortia, special companies etc. The area III projects are considered to be dangerous because of the temptation to implement them which resulted from low costs. However, they do not provide sufficient benefits, therefore other performance variants should be taken into account. Most demanded is the performance of area IV projects because these are most advantageous.

C. Reference to projects efficiency account

I **LOW BENEFIT** **HIGH COST**	**II** **HIGH BENEFIT** **HIGH COST**
III **LOW BENEFIT** **LOW COST**	**IV** **HIGH BENEFIT** **LOW COST**

Fig. 2 Cost-Benefit Matrix

Business case usually supports taking of investment decisions i.e. the situations where decisions are made about alternative possibilities of using financial means in the project. Owing to BC at least the financial consequences of those decisions may be analysed, but also their non-financial effects may be considered. So BC is a practical use of the investment efficiency account in project management.

Assessment of the project cost-effectiveness as the most important component of DBC in PRINCE2 comprises analysis of joint benefits and adverse effects with the project performance costs and future maintenance of its final products. For the financial part PRINCE2 recommends here: analyses of complete costs and benefits, net costs and products. Recommended for the financial part of PRINCE2 are: analyses of total costs and benefits, net costs and benefits, return on investment, simple payback period, discounted methods, current net value or analysis of project sensitivity to disturbances.

The CBA part involving the financial analysis should show the project profitability and its financial sustainability by solving the following problems: choice of the time horizon, determination of total inputs and incomes, calculation of residual value at the end of the year, determination of the inflation factor, choice of an appropriate discount rate, financial calculation or economic rate of return and the method of using those indices in project assessment. The project effects prognosis should be made for its performance time and the period allowing to show its effects in close and further perspective.

During economic analysis carried out at CBA the costs and notable socio-economic benefits of the project are determined. If it is possible, the benefits for external environment should be expressed in monetary units, Otherwise, ascribed to those benefits should be an appropriate numerical value enabling to bring them down to rational values. Both all the costs and social benefits expected for various years should be discounted to the basic year value or using a harmonized discount rate for a given sector of economy or region. An alternative approach is calculation of internal economic rate of return or economic current net value.

TABLE XI.
SIMPLIFIED BUSINESS CASE FOR THE PROJECT A WITH COST BENEFIT ANALYSIS AND INVESTMENT APPRAISAL

YEAR	0	1	2	3	4	5	**VALUES**
Costs (in thousands of PLN)	-100	-20	-20	-20	-40	-30	
Benefits (in thousands of PLN)	0	40	80	100	100	100	
Cash flow (in thousands of PLN)	-100	20	60	80	60	70	**58**
Cumulative cash flow (in thousands of PLN)	-100	-80	-20	60	120	190	
Discount ratio	1.00	0.94	0.88	0.83	0.78	0.74	
Cash flow after discounting (rounded up to thousands of PLN)	-100	19	53	66	47	52	$\sum CF_t = 237$
Cumulative cash flow after discounting (in thousands of PLN)	-100	-81	-28	38	85	137	**NPV = 137**

D. Case study

Let us consider two investment projects: project A and project B. Both have identical performance time reaching 5 years and identical investment inputs amounting to 100 000 PLN. In the case of project A the following fluxes of financial surpluses (CF) are forecast: at the end of the first year – 20 000 PLN, the second year – 60 000, the third year – 80 000, the fourth year – 60 000 and the fifth year – 70 000 PLN (line 3 in Table IV). In project B the forecast financial surpluses at the end of each of the periods are identical and amount to 58 000 PLN.

Simplified BC including the cost benefit analysis and investment assessment is presented for project A in Table XI, whereas for project B in Table XII.

Actually, introduced to pattern BC are all costs (the first lines in Table XI and Table XII) with tangible benefits (the second lines in these tables), which gives cash flows for the projects (the third lines), equal to benefits (incomes) decreased by costs.

The cost benefit analysis is aimed at discerning the moment when the first incomes from the project appeared. They are then assessed using a discount rate. In the fifth line for each period the discount ratio is calculated. The assumed (in these examples) rate of return within a year is 6.3% for every project.

The discount ratio for a given period is treated similarly as weight while counting the weighted average, except that in the NPV case it is the „weighted total". Pursuant to this premise, a further stage is discounting of cash flows (the sixth lines in Table XI and Table XII) by multiplying the value of cash flows from a given period (the third line) by the value of the discount ratio (the fifth line). Subsequently, at the intersection of the sixth line and the seventh column the total of discounted cash flows is calculated ($\sum CF_t$).

On the other hand, the aim of the whole investment is to determine the NPV value, otherwise called the updated net value or the present net value. NPV is a method of assessment of the tangible investment economic efficiency but also an indicator determined according to this method.

In the present situation, NPV is considered as an indicator constituting a difference between the sum of discounted cash flows and initial inputs.

It is assumed that the investment should pay for itself in the period not longer than 2 years. The period of return on inputs in the case of project A amounts to 2.4 years, and in the case of project B it reaches 1,9 years. The calculations used the formula in which inputs are compared with cumulated positive cash flows and then we should observe when the sum is zero. Project B is better than project A, because it provides a faster return on inputs and does not

TABLE XII.
SIMPLIFIED BUSINESS CASE FOR THE PROJECT B WITH COST BENEFIT ANALYSIS AND INVESTMENT APPRAISAL

YEAR	0	1	2	3	4	5	VALUES
Costs (in thousands of PLN)	-100	-20	-20	-20	-40	-30	
Benefits (in thousands of PLN)	0	78	78	78	98	88	
Cash flow (in thousands of PLN)	-100	58	58	58	58	58	**58**
Cumulative cash flow (in thousands of PLN)	-100	-42	16	74	132	190	
Discount ratio	1	0.94	0.88	0.83	0.78	0.74	
Cash flow after discounting (rounded up to thousands of PLN)	-100	55	51	48	45	43	$\sum CF_t = 242$
Cumulative cash flow after discounting (in thousands of PLN)	-100	-45	6	54	99	142	**NPV = 142**

exceed the assumed two-years' period of the investment pay-back.

NPV for project A amounts to 137 000 PLN, and for project B – 142 000 PLN (intersection of the seventh line and the seventh column in Table XI and Table XII). In NPV we accept only projects with NPV higher than zero, and we reject the other ones. The decision about the choice of a specific project is correct for the specified interest rate.

Of the two projects excluding each other we choose the project with a higher NPV. In the given case it is project B which not only provides a faster return of inputs but it also provides an additional surplus of 142 000 PLN.

Implementation of business case in organization changes the way of thinking about project initiatives, because it requires explicit defining of profits from performance of the undertaking. Problems with indicating the profits or inability to measure the profits demonstrate that the suggested project has significant drawbacks.

An important advantage of the use of BC is the necessity to prepare some variants of the project performance, which gives the organization's managers more decision-making possibilities e.g. as to alternative investment methods. On the other hand, the need to include many complex aspects constitutes the greatest drawback of this document.

REFERENCES

[1] W. Rogowski, "Współczesny rachunek efektywności projektów – ujęcie problemowe", in Ocena projektów – koncepcje i metody, M. Trocki, M. Juchniewicz, Ed. Warsaw: Szkoła Głowna Handlowa w Warszawie - Oficyna Wydawnicza, 2013, pp. 43-76.

[2] M. Trocki, "Kompleksowe metody oceny projektów", in Nowoczesne zarządzanie projektami, M. Trocki, Ed. Warsaw: Polskie Wydawnictwo Ekonomiczne, 2012, pp. 280-295.

[3] Managing Successful Projects with PRINCE2, TSO (The Stationery Office), London: 2009.

[4] M. Juchniewicz, M. Trocki, "Metody oceny projektów", in Ocena projektów – koncepcje i metody, M. Trocki, M. Juchniewicz, Ed. Warsaw: Szkoła Głowna Handlowa w Warszawie – Oficyna Wydawnicza, 2013, pp. 207-211, pp. 201-202, pp. 196-202.

[5] J. Ward, E. Daniel, J. Peppard, "Building Better Business Case for IT Investments", MIS Quarterly Executive, vol. 7, no. 1, 2008, pp.1-15.

[6] M. J. Schmidt, The IT Business Case: Key to Accuracy and Credibility, Solution Matrix Ltd., Boston: 2008.

[7] H. Smith, J. McKeen; C. Cranston; M. Benson, "Investment Spend Optimization: A New Approach to IT Investment at BMO Financial Group", MIS Quarterly Executive, vol. 9, no. 2, 2010.

[8] E. Sońta-Drączkowska, Zarządzanie wieloma projektami. Warsaw: Polskie Wydawnictwo Ekonomiczne, 2012, pp. 7–87, 104-141.

[9] M. Trocki, "Wprowadzenie do problematyki oceny projektów", in Ocena projektów – koncepcje i metody, M. Trocki, M. Juchniewicz, Ed. Warsaw: Szkoła Głowna Handlowa w Warszawie – Oficyna Wydawnicza, 2013, pp. 9-16.

[10] P. Wyrozębski, "Metodyka PRINCE2®", in Metodyki zarządzania projektami, M. Trocki, Ed. Warszawa: Bizarre, 2011, pp. 95-122.

[11] A. Kaczorowska, E-usługi administracji publicznej w warunkach zarządzania projektami. Łódź: Wydawnictwo Uniwersytetu Łódzkiego, 2013.

[12] W. W. Sofko, The Role of Cost-Benefit Analysis in Achieving Results nin Special Education, Retrieved April, 05,2016 from http://www.wildwoodinstitute.org/knowledge/results.html

[13] T. Melton, P. Iles-Smith, J. Yates, Project Benefits Management. Linking Projects to the Business. Elsevier, 2008.

[14] W. Rogowski, Rachunek efektywności przedsięwzięć inwestycyjnych. Kraków: Oficyna Ekonomiczna, 2004.

[15] T. J. Kloppenborg, Project Management: A Contemporary Approach. CA: South-Western Cengage Learning, 2009, pp. 34-50.

[16] O. Nadskakuła, Ewaluacja projektów. Warsaw: Bizarre, 2009.

[17] P. Harper-Smith, S. Derby, Fast Track to Success Project Management. Edinburgh: Pearson Education Limited, 2009, p. 205.

[18] S. E. Portny, Zarządzanie projektami dla bystrzaków. Gliwice: Helion, 2013, p. 46.

[19] W. Dow, PMP, B. Taylor, Project Management Communication Bible. Indianapolis: Wiley Publishing, 2008, pp. 132-136, 491-494.

[20] R. K. Wysocki, R. Mcgary, Efektywne zarządzanie projektami. Gliwice: Helion, 2005, p. 112.

[21] M. Trocki, E. Bukłaha, B. Grucza, P. Wyrozębski, M. Juchniewicz, and W. Metelski, Nowoczesne zarządzanie projektami. M. Trocki, Ed. Warsaw: Polskie Wydawnictwo Ekonomiczne, 2012, pp. 280-284.

[22] R. Newton, The Practice and Theory of Project Management. New York: Palgrave Macmillan, 2009, pp. 21, 93-94.

[23] J. Phillips, Zarządzanie projektami IT. Helion, Gliwice, 2011, pp. 64-71.

[24] T. Kasprzak, Biznes i technologie informacyjne. Perspektywa integracji strategicznej. Warsaw: Katedra Informatyki Gospodarczej I Analiz Ekonomicznych UW, 2003.

[25] G. Serefeimidis, "A Review of Research Issues in Evaluation of Information Systems", in Information Technology Evaluation Methods and Management, W. V. Grembergen, Ed. London: Idea Group Publishing, 2001, DOI: 10.4018/978-1-878289-90-2.ch004.

[26] M. Nowak, T. Błaszczyk, B. Nowak, K. Targiel, Wspomaganie decyzji w planowaniu projektów. Warsaw: Difin, 2014, pp.82-136.

[27] C. Pineiro Sanchez, "Aplicaciones de la teoria de la decision multicriterio", in Revista Gallega de Economia, vol. 12, no. 1, 2003, pp. 105-122.

[28] M. M. Parker, H. E. Trainor, R. J. Benson, Information Economics. Linking Business Performance to Information Technology. New Jersey: Prentice Hall, 1988, pp. 36-37.

[29] M. M. Parker, H. E. Trainor, R. J. Benson, Information Strategy and Economics. New Jersey: Prentice Hall, 1989.

[30] T. L. Saaty, "Priority Setting in Complex Problems", in IEEE Transaction of Engineering Management, vol. EM 30, no. 3, 1983, pp. 140-155, DOI: 10.1109/TEM.1983.6448606.

[31] Z. Pawlak, Biznesplan zastosowania i przykłady. Warsaw: Oficyna Wydawnicza WSEiZ w Warszawie, 1999, pp. 196-200.

[32] R. W. Griffin, Podstawy zarządzania organizacjami. Warsaw: Wydawnictwo Naukowe PWN, 2004.

[33] M. Trocki, E. Bukłaha, B. Grucza, M. Juchniewicz, W. Metelski, and P. Wyrozębski, Nowoczesne zarządzanie projektami. M. Trocki, Ed. Warsaw: Polskie Wydawnictwo Ekonomiczne, 2012, p. 285.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1169–1174
DOI: 10.15439/2016F238
ACSIS, Vol. 8. ISSN 2300-5963

Fundamental analysis in the multi-agent trading system

Jerzy Korczak, Marcin Hernes, Maciej Bac
Wrocław University of Economics, ul. Komandorska 118/120, 53-345 Wrocław, Poland

e-mail: {jerzy.korczak, marcin.hernes, maciej.bac}at ue.wroc.pl

Abstract—The paper presents issues related to developing methods for fundamental analysis used to expand capabilities of multi-agent trading system, to better predict the financial market. The fundamental analysis indicators can be used as confirmation of decisions generated by other strategies of the system. The first part of the article discusses briefly the fundamental analysis issues in relation to the online trading on FOREX market. The statistical analysis of correlations of the different time series indicators and algorithms of fundamental analysis agents are examined. The final part discusses the results of the performance evaluation of selected investment strategies, including fundamental–based agents.

I. Introduction

In trading support systems, the advices might be computed by one or many algorithms, or by one or many software agents using one or many information sources. An overview of of the agents operating on financial markets has already been given by B. LeBaron [1]. Currently, most trading systems are based on one or only a few algorithms. For example, the solutions described by L. Mendes, P. Godinho and J. Dias [2] use genetic algorithms to perform analysis on the basis of historical quotations. The system described by J.R. Thompson, J.R. Wilson and E. P. Fitts [3] is based on the multifractal time series. In the system described in [4, 5] technical analysis indicators are used. There are many solutions based on multi-agent approach. H. C. Aladag, U. Yolco and E. Egrioglu [6] present an evaluation of the portfolio optimisation strategies by three agents: rational agent, interference agent, and technical analysis agent. P. Singh and B. Borah [7] apply a multi-agent system where the agents' intelligence is based on fuzzy expert system. O. Badawy and A. Almotwaly [8] developed the neural networks and neuro-fuzzy computing for taking into account the geometrical patterns of the financial data. M. Aloud, E.P.K. Tsang and R. Olsen [9] introduce an agent-based model for simple prediction of financial markets, where each agent predicts the development of selected subsets of the assets pairs in real time by separately examining the similarities between ask and bid assets histories. P. Kaltwasser [10] describes an agent that uses multiple behavioral techniques to make bidding decisions in the face of market uncertainty. J. Glattfelder, A. Dupuis and R. Olsen [11] develop a system that supports multiple strategies, but they use only moving-average crossover strategies in the current stage of development. The paper by R. Barbosa and O. Belo [12] describes a multi-agent system that consists of a set of trading models such as an ensemble of classifiers, regression models, case-based reasoning, and an expert system. F. H. Westerhoff [13] presents a system where two groups of agents applying the methods of fundamental and technical analysis try to shape market dynamics.

Summing up, more often the trading advice is provide by multiple software agents that use mainly technical analysis. However, many papers related to economic basics [e.g. 14, 15, 16] state that using a fundamental analysis is also necessary for supporting trading decisions. A few of solutions combine fundamental analysis and behavioral sentiments.

Our platform, called A-Trader [17, 18], allows for the implementation of various algorithms or trading decision support methods [e.g. 19, 20]. The A-Trader is aimed at supporting trading decisions on the FOREX market (Foreign Exchange Market). On FOREX currencies are traded in pairs, for example USD/PLN, EUR/GBP. In general, trader on FOREX can open/close long/short positions. A long position relies on "*buying low and selling high*" in order to achieve a profit.. A short position, instead, relies on "*buying high and selling low*". On FOREX, when one currency in a pair is rising in value, the other currency is declining, and vice versa [18]. The A-Trader receives tick data which are aggregated to minute (M1, M5, M15, M30), hour (H1, H4), day (D1), week (W1) and month (MN1). The A-Trader mainly supports High Frequency Trading (HFT) [17], and puts strong emphasis on price formation processes, short-term positions, fast computing, and efficient and robust indicators [19].

High frequency traders seek profits from the market's liquidity imbalances and short-term pricing inefficiencies. Access to quote data must be near real time. Therefore systems supporting trading must provide as soon as possible advice as to which position should be taken: open, close or no entry (do nothing).

The architecture of A-Trader and the description of the different groups of agents have already been detailed [17, 18]. In general, the agents applied technical and behavioral analysis in order to support trading decisions.

The aim of this paper is to present methods for fundamental analysis used to improve the trading efficiency of A-Trader. These methods take into consideration both HFT and also different time resolutions (multiresolution is an integral part

of A-Trader: from M1 to MN1 periods). Multiresolution is very important issue due to different time periods needed for fundamental analysis on FOREX market. The macro-economic indicators (e.g. inflation, Gross Domestic Product) are characterized by low fluctuation. (analysis of these indicators is usually performed monthly).

The first part of the article briefly discusses the fundamental analysis issues in relation to the FOREX market. Next, the statistical analysis of correlations of the different time series indicators and algorithms of fundamental analysis agents are examined. The final part discusses the results of the performance evaluation of selected investment strategies, including fundamental– based strategies.

II. Fundamental analysis on FOREX market

The main assumption of fundamental analysis is the notion of equilibrium. At any considered time, a currency pair should trade at a particular rate that balances trade and investment flows. Fluctuations in market risk affect a rate that may be well above or below what financial-economic conditions justify [15]. Unlike technical analysis, fundamental analysis doesn't become less profitable when competition increases. If market dynamics raise interest rates, in consequence a currency will rise proportionately. This cause will not be undone if too many people are aware of it. On the contrary, it will become even stronger [15].

Putting it briefly, FOREX fundamental analysis concerns the following:

1. Driving supply and demand in the market currencies.
2. Economic indicators and asset markets.
3. Indexes.
4. Political and social powers.

There are two major factors affecting *the supply and demand balance*: interest rates and the state of international trade [22]. Interest rates can have either a strengthening or weakening effect on a particular currency. The high interest rates attract foreign investment, which will strengthen the local currency. Stock market investors often react to interest rate increases by selling off their holdings. How changes in central bank interest rates impact exchange rates are described by following overarching forces: interest rate parity and the carry trade. Emerging (growth) currencies tend to trade in direct proportion to relative interest rate levels, since higher rates attract speculative investors. Recall that investors in the carry trade seek to profit from positive interest rate differentials; hence, the higher the interest rate, the more attractive the corresponding currency [17]. An international trade balance arises if the economy shows a deficit (more imports than exports) which mean that money is flowing out of the country to purchase foreign-made goods, and this may have a devaluing effect on the currency [27]. A decent fundamental analysis comprises the examination of *macroeconomic indicators* and *asset markets*, when evaluating a country's currency. Macroeconomic indicators include figures such as [10, 15, 16, 23]:

1) Growth rates, measured by Gross Domestic Product.
2) Inflation.
3) Unemployment.
4) Balance of payments.
5) Market correlations, such as [16, 24]:
   - gold prices ratio: when gold goes up, the USD often goes down (and vice versa); therefore, an inverse relationship appears between gold and USD ,
   - oil prices ratio: the economies of oil-dependent countries weaken as oil prices rise; in such situation a trader can consider buying currencies of commodity-based economies like Australia or Canada or selling oil-dependent currencies.
6) Productivity.
7) Purchasing Managers' Index (PMI), based on new orders, inventory levels, production, supplier deliveries and the employment environment.

Asset markets comprise stocks, bonds and real estate. Other indicators that may be considered are the Consumer Price Index (CPI), Durable Goods Orders, Producer Price Index (PPI), and retail sales.

Index correlations also have influence on currency quotations [25]. Examples of indexes are:

- S&P 500 is an index which includes the 500 companies with the largest capitalization companies listed on the New York Stock Exchange and NASDAQ,
- FTSE 100 is a share index of the 100 companies listed on the London Stock Exchange with the highest market capitalization,
- WIG index is listed on the Warsaw Stock Exchange; it includes shares listed on the Polish market.

Political considerations impact the level of confidence in a nation's government, the climate of stability and level of certainty [25].

The trading advices generated by agents based on fundamental analysis are used as confirmations of a buy/sell decision suggested by technical analysis–based agents or behavioral–based agents.

III. Comparison of correlation of selected fundamental analysis factors and FOREX quotations

As was stated in the previous section, the fundamental analysis factors (macro-economic indicators) of a given country are often correlated with its currency quotation.. Fig. 1 presents the example of correlation inflation, oil price, and S&P 500 index with USD/GBP quotations in 2015 (monthly). On the basis of visual analysis of this chart, we can draw the conclusion that in particular periods most of the macro-economic indicators are correlated with USD/GBP quotations. For example, taking into consideration the period May-June, the USD/GBP quotations rise and inflation and S&P 500 also rise (while the oil quotation falls); taking into

consideration the period November – December, the USD/GBP quotation rises and inflation, S&P500, and oil also rise.

In order to determine the correlation between these factors, the Mann-Kendall test was used. The non-parametric Mann-Kendall test is commonly employed to detect monotonic trends in series of environmental data, climate data or hydrological data [24]. For two trends correlation a tau Kendall ratio was calculated that counts the number of pairwise disagreements between two time series. The larger the distance, the more dissimilar the two series are [24]. If tau Kendall value is near 1 or -1, then two time series are correlated, if it is near 0, then time series are independent.

For example, for the time series presented on Fig 1 (period - 2015 year) tau Kendall correlation values are the following:

- USD/GBP vs. inflation: 0,1,
- USD/GBP vs. oil: – 0.04,
- USD/GBP vs. S&P500: – 0.2.

According to this metric, the macroeconomic values are not strongly correlated with the USD/GBP quotation, whereas if we look at the chart, they seem to be correlated. This may result from tau Kendall specifics and also from time-shift of the trend change considered quotation and indicators. For example, the USD/GBP quotation starts rising in April, the S&P500 also in April, but inflation starts rising in May or oil starts rising in March. Therefore, macro-economic indicators can be used mainly for confirmation of long time trends of FOREX quotations.

IV. Description and Evaluation of the Fundamental Analysis Agents

Input signals of fundamental analysis agents appear with different frequency. One of them is available in high time periods, for example quarterly, monthly, weekly, like the Inflation PPI M/M, Net Capital Inflows, Change of employment M/M, and Industrial production.

These signals are very important in establishing and confirming long time trends. According to the rule „*Trend is your friend*", investing in pursuance of the trend should gain more profits. The Elliott wave theory says that movements in line with the trend are longer and more dynamic. This indicator may be a good advisor in the investment strategy. But it is not enough in the high frequency trading strategy where trades are frequent and positions are opened only for a short time. This indicator can therefore be taken only as advice, not as a signal to open position. The actual macro-economic situation of the country is the real important indicator. Namely, the main stock exchange indicators such as S&P500 and FTSE 100 are the main advising factor and trigger for making a trade. Principal goods such as gold and oil are also included, and they usually function as short time indicators.

Neural networks have already demonstrated great potential for discovering non-linear relationships in time-series. The published results of forecasting financial data are particularly good [21, 26, 27]. Therefore, we took the decision to make use of the neural network model as a predictor.

In general, our fundamental analysis agents were built on the Multi-Layer Perceptron model. The diagram of the agent operation is schematically presented in Fig 2. It uses the sigmoid activation function and the back-propagation learning algorithm. Long and short term fundamental indicators were taken into account. The input vector contained the long term indicators and the last sequences of S&P500, FTSE 100, oil and gold tics. As output, changes in USD/GBP rates were expected. In the learning process, the output values of the neural network were shifted by T_n units in time.

The trading strategy of the fundamental analysis agent can be specified as follows:

```
Input: q_FTSE100 =<q_ftse1, q_ftse2, .... q_ftseM>
    // FTSE100 quotations,
   q_S&P500=<q_s&p1, q_s&p2, .... q_s&pM>
   // S&P500 quotations
   q_GOLD=<q_gold1, q_gold2, .... q_goldM>
   // GOLD quotations
   q_OIL=<q_oil1, q_oil2, .... q_oilM>
   // OIL quotations
   i_Inflation // M/M inflation change
   i_NCI // Net Capital Inflows change
   i_COE // Change of employment M/M
   i_IP // Industrial Production M/M change
```

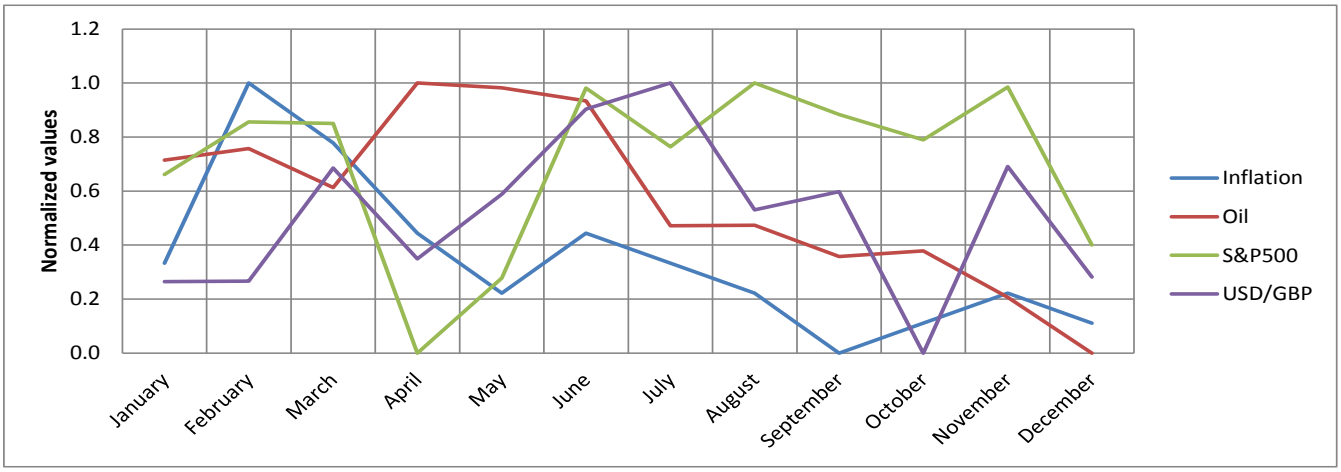


Fig. 1. The example of correlation between inflation, oil prices and S&P 500 index with USD/GBP quotation in 2015 (monthly).

```
    thresholdopen   //threshold   level   for   open
    //long/close short //position
    thresholdclose   //threshold   level   for   close
    //long/open short //position
Output: The fuzzy logic recommendation D
    // (range [-1..1]).
BEGIN
    if
        CheckPerformanceLevel()
    then
        BeginLearningProcess();
        p_USD/GBP_M+3      :=      Multi-layerPerceptron
        (q_FTSE100,          q_S&P500,          q_GOLD,
        q_OIL,i_Inflation, i_NCI, i_COE, i_IP);
        //Prediction of USD/GBP value change
        if
          p_USD/GBP_M+3 > thresholdopen
        then
          D:= heuristic_open(p_USD/GBP_M+3);
        else if
          p_USD/GBP_M+3 < thresholdclose
        then
          D:= heuristic_close(p_USD/GBP_M+3);
        otherwise  D:= heuristic_do_nothing
    END
```

From the point of finance, the fundamental analysis agent is founded on money flow interpretation. For instance, if the S&P 500 is falling and FTSE 100 is rising, one can suppose that investors exchange their S&P shares for USD, then they exchange USD to GBP, then they buy FTSE 100 shares. So if they buy GBP for USD, the value of GBP to USD should rise.

In A-Trader, an evaluation of selected agents is performed with the use of the following measures (ratios):

- rate of return (ratio x_1),
- the number of the transaction,
- gross profit (ratio x_2),
- gross loss (ratio x_3),
- the number of profitable transactions (ratio x_4),
- the number of profitable consecutive transactions (ratio x_5),
- the number of unprofitable consecutive transactions (ratio x_6),
- Sharpe ratio (ratio x_7)

$$S = \frac{E(r) - E(f)}{|O(r)|} \cdot 100\% \quad (1)$$

where:
$E(r)$ – arithmetic average of the rate of return,
$E(f)$ – arithmetic average of the risk-free rate of return,
$O(r)$ – standard deviation of rates of return.

- the average coefficient of volatility (ratio x_8) is the ratio of the average deviation of the arithmetic average multiplied by 100% and is expressed:

$$V = \frac{s}{|E(r)|} \cdot 100\% . \quad (2)$$

where:
V – average coefficient of variation,
s – average deviation of the rates of return,
E(r) – arithmetic average of the rates of return.

- the average rate of return per transaction (ratio x_9), counted as the quotient of the rate of return and the number of transactions.

For the purpose of the comparison of the agents' performance, the following evaluation function was elaborated:

$$y = (a_1 x_1 + a_2 x_2 + a_3 (1 - x_3) + a_4 x_4 + a_5 x_5 + \ldots$$
$$+ a_6 (1 - x_6) + a_7 x_7 + a_8 (1 - x_8) + a_9 x_9) \quad (3)$$

where x_i denotes the normalized values of the ratios. Coefficients a_1 to a_{10} may be also determined by the investor in accordance with his/her preferences (for instance, the user may determine whether is interested in the higher rate of return with a simultaneous higher risk level or lower risk level, but simultaneously agrees to a lower rate of return). These functions allow for determining the best strategies for the user in a given time period. Coefficients a_1 to a_{10} can be used also for creating different users' profiles (allow for personalization of A-Trader).The function is given the values

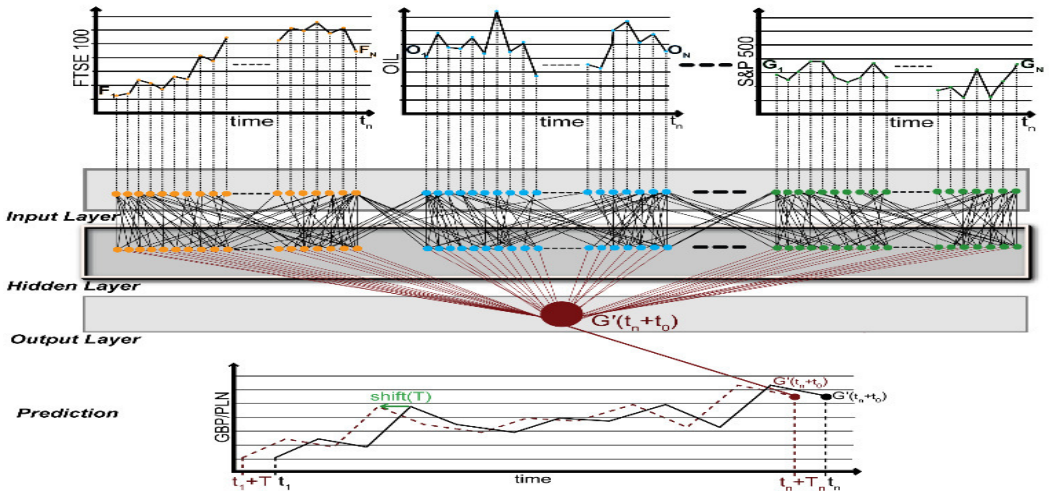


Fig. 2 Fundamental Analysis Agent

from the range [0..1], and the agent's efficiency is directly proportional to the function value.

V. EXPERIMENTS

The agent's performance analysis was carried out for data within the M1 period of quotations from the FOREX market. For the purpose of this analysis, the test was performed in which the following assumptions were made:

1. USD/GBP quotes were selected from randomly chosen periods (each - 1440 quotations), notably:
   - 15-03-2016, 0:00 am to 15-03-2016, 23:59 pm,
   - 16-03-2016, 0:00 am to 16-03-2016, 23:59 pm,
   - 22-03-2016, 0:00 am to 22-03-2016, 23:59 pm,
2. At the verification, the trading signals (for open long/close short position equals 1, close long/open short position equals -1) were generated by the strategies *FundamentalStrategy* and *Technical Strategy*.
3. It was assumed that decisions' probability levels for open/close position are determined by the genetic algorithm (on the basis of earlier periods).
4. It was assumed that the unit of performance analysis ratios (absolute ratios) are pips (a pip is equivalent to the final number in a currency pair's price).
5. The transaction costs were directly proportional to the number of transactions.
6. The capital management - it was assumed that in each transaction the investor engages 100% of the capital held at the leverage 1:1. The investor may define another capital management strategy.
7. The results obtained by the tested agents were compared with the results of the *Buy-and-Hold* benchmark (a trader buys a currency at the beginning and sells a currency at the end of a given period).

Table 1 presents the results obtained in the particular periods. In general, it may be noted that strategies generated not only profitable decisions. However, the rate of return cannot be the only measure taken into account in the performance evaluation. Very important are also other ratios, among other risk involved in the investment.

The evaluation function provides the fast choice of the best strategy. It may be noted that the values of efficiency ratios of particular strategies differ in each period. Values of this function oscillate in the range from 0.02 to 0.51. Therefore use of this function allows for reducing the deviation of the values of the ratios.

The results of the experiment allow us to state that the ranking of strategies' evaluation differs in particular periods. In the first period, the *TechnicalStrategy* was the best strategy, the *FundamentalStrategy* was ranked higher than *B&H benchmark*. In the second period also the *TechnicalStrategy* was ranked highest, but *FundamentalStrategy* was ranked lower than *B&H*. Considering the third period, it may be noted that the *FundamentalStrategy* was the best strategy, although the *Rate of Return* of this strategy was not highest.

The highest value of evaluation function of *FundamentalStrategy* (in third period) results from the highest *Average Rate of Return per Transaction* and low risk measures' values. The *B&H* benchmark was ranked lowest in two periods, and in the first and third periods it generated the losses. It should be noted that in the second period, the upward trend was observed, therefore *B&H*'s *Rate of Return* was positive. The first and the third periods show a downward trend, and therefore the *B&H*'s *Rate of Return* is negative. Taking into consideration all the periods, it may be stated that there is no one strategy ranked highest most often. Also, strategies achieving the highest *Rate of Return* were not always ranked in the highest positions. *TechnicalStrategy* always was characterized by greater number of transactions than *FundamentalStrategy*. However, *FundamentalStrategy* is characterized by highest Rate of Return per Transaction, than *TechnicalStrategy*. Both, *TechnicalStrategy* and *FundamentalStrategy* are characterized by low level of risk. This results mainly from the fact that these strategies take into consideration decisions generated by large numbers of agents.

During performing the experiments, a problem with neural network learning, was also observed. Values of the input

TABLE I.
PERFORMANCE ANALYSIS RESULTS

Ratio	FundamentalStrategy			TechnicalStrategy			B & H		
	Period 1	Period 2	Period 3	Period 1	Period 2	Period 3	Period 1	Period 2	Period 3
Rate of return [pips]	9	-66	28	56	132	48	-143	89	-160
The number of transactions	4	7	5	39	52	42	1	1	1
Gross profit [pips]	11	61	43	125	124	93	0	89	0
Gross loss [pips]	17	37	28	65	81	49	-143	0	-160
The number of profitable transactions	3	3	3	27	38	27	0	1	0
The number of profitable consecutive transactions	2	2	2	11	6	4	0	1	0
The number of unprofitable consecutive transactions	1	3	2	2	2	1	1	0	1
Sharpe ratio	0.55	1.8	0,78	0.91	2.60	1.38	0	0	0
The average coefficient of volatility [%]	1.66	0.36	1.94	1.86	0.24	0.92	0	0	0
The average rate of return per transaction	2.25	-9,43	5,6	1.43	2.54	1.14	-143	89	-160
Value of evaluation function (y)	**0.43**	**0.19**	**0.51**	**0.49**	**0.43**	**0.47**	**0.08**	**0.38**	**0,02**

vectors was very similar, thus learning process was not always convergent. It may result from long-time fluctuations of fundamental indicators in relation to short-time fluctuation of currencies quotations.

VI. Conclusion

The fundamental analysis implemented as investment strategy in A-Trader can be used as confirmation of decisions generated by other strategies. However, investing only on the basis of FundamentalStrategy does not allow us to achieve a satisfactory rate of return. Often, fundamental analysis indicators are time-shift in relation to currency value (for example, changes in oil prices or S&P500 quotations presented in fig 1 go faster than the USD/GBP trend). Thus, although we see on the chart a correlation between fundamental analysis indicators and currency quotations, the Mann-Kendall test has not confirmed this correlation, due to the time-shift. This implies the need for further research work on developing the fundamental strategy to take into consideration the time-shift phenomenon and on improving methods for neural network learning adjusted to fundamental indicators' values.

References

[1] B. LeBaron, "Active and Passive Learning in Agent-based Financial Markets", *Eastern Economic Journal*, vol. 37, pp. 35-43, 2011.

[2] L. Mendes, P. Godinho and J. Dias, "A Forex trading system based on a genetic algorithm", *Journal of Heuristics* 18 (4), pp. 627-656, 2012.

[3] J.R. Thompson, J.R. Wilson and E. P. Fitts, "Analysis of market returns using multifractal time series and agent-based simulation", in *Proceedings of the Winter Simulation Conference* (WSC '12). Winter Simulation Conference, Article 323, 2012.

[4] C. D. Kirkpatric and J. Dahlquist, *Technical Analysis: The Complete Resource for Financial Market Technicians*, Financial Times Press, 2006.

[5] C. Lento, "A Combined Signal Approach to Technical Analysis on the S&P 500", *Journal of Business & Economics Research,* 6 (8), pp. 41–51, 2008.

[6] H. C. Aladag, U. Yolco and E. Egrioglu, "A new time invariant fuzzy time series forecasting model based on particle swarm optimization", *Applied Soft Computing*, 12 (10), pp. 3291-3299, 2012.

[7] P. Singh and B. Borah, "Forecasting stock index price based on M-factors fuzzy time series and particle swarm optimization", *International Journal of Approximate Reasoning*, 55 (3), pp. 812-833, 2014.

[8] O. Badawy and A. Almotwaly, "Combining neural network knowledge in a mobile collaborating multi-agent system", *Electrical, Electronic and Computer Engineering*, ICEEC '04, pp. 325, 328, 2004, DOI: 10.1109/ICEEC.2004.1374457.

[9] M. Aloud, E.P.K. Tsang and R. Olsen, "Modelling the FX Market Traders' Behaviour: An Agent-based Approach", in *Simulation in Computational Finance and Economics: Tools and Emerging Applications*, B. Alexandrova-Kabadjova, S. Martinez-Jaramillo, A. L. Garcia-Almanza and E. Tsang (eds.), IGI Global, 2012, pp. 202-228.

[10] P. R. Kaltwasser, "Uncertainty about fundamentals and herding behavior in the FOREX market", *Physica A: Statistical Mechanics and its Applications*, 389 (6), pp. 1215-1222, March 2010.

[11] J. B. Glattfelder, A. Dupuis and R. Olsen, "Patterns in high-frequency FX data: Discovery of 12 empirical scaling laws", *Quantitative Finance*, 11 (4), pp. 599-614, 2011.

[12] R.P. Barbosa and O. Belo, "Multi-Agent Forex Trading System", in *Agent and Multi-agent Technology for Internet and Enterprise Systems, Studies in Computational Intelligence*, vol. 289, 2010, pp. 91-118.

[13] F. H. Westerhoff, "Multi-Asset Market Dynamics", *Macroeconomic Dynamics*, 8/2011, pp. 596—616, 2011.

[14] S. Johnson, "Push to tap into EM currency returns", *Financial Times*, February 2011.

[15] A. Kritzer, "Forex For Beginners. A Comprehensive Guide to Profiting from the Global Currency Markets", Apress, 2012.

[16] J. Kumar, T. Rao and S. Srivastava, "Economics of Gold Price Movement-Forecasting Analysis Using Macro-economic, Investor Fear and Investor Behavior Features", in S. Srinivasa, V. Bhatnagar (eds.), *Big Data Analytics*, Lecture Notes in Computer Science, Volume 7678, Springer-Verlag, Berlin, Heidelberg 2012, pp. 111-121, DOI: 10.1007/978-3-642-35542-4_10.

[17] J. Korczak, M. Hernes and M. Bac, "Fuzzy Logic as Agents' Knowledge Representation in A-Trader System", in E. Ziemba (ed.), *Information Technology for Management, Lecture Notes in Business Information Processing,* vol. 243, Springer International Publishing, 2016, pp. 109-124.

[18] J. Korczak, M. Hernes and M. Bac, "Performance evaluation of decision-making agents' in the multi-agent system", in *Proceedings of Federated Conference Computer Science and Information Systems (FedCSIS)*, Warszawa, 2014, pp. 1171 – 1180. DOI: 10.15439/2014F188.

[19] M. Hernes and N.T. Nguyen, "Deriving Consensus for Hierarchical Incomplete Ordered Partitions and Coverings", *Journal of Universal Computer Science* 13 (2) , pp. 317-328, 2007.

[20] M. Hernes and J. Sobieska-Karpińska , "Application of the consensus method in a multi-agent financial decision support system", *Information Systems and e-Business Management* 14 (1), Springer Berlin Heidelberg, 2016, DOI: 10.1007/s10257-015-0280-9.

[21] P. D. McNelis, "Neural Networks in Finance: Gaining Predictive Edge in the Market" (Academic Press Advanced Finance Series). Academic Press, Inc., Orlando, FL, USA, 2004.

[22] F. Ahrens, "U.S. Trade Imbalance Widens Slightly: So What?" Washington Post, May 2009.

[23] S. Johnson, "Push to tap into EM currency returns", *Financial Times*, February 2011.

[24] J. Beckmann, A. Belke and M. Kühl, "The dollar-euro exchange rate and macroeconomic fundamentals: a time-varying coefficient approach", *Review of World Economics* 147 (1), pp 11-40, 2011.

[25] P. Singh and B. Borah, "Forecasting stock index price based on M-factors fuzzy time series and particle swarm optimization", *International Journal of Approximate Reasoning*, 55 (3), pp. 812-833, 2014.

[26] P. van Foreest and C. G. de Vries, "The Forex Regime and EMU Expansion", *Open Economies Review*, 14 (3), pp. 285-298, 2003.

[27] http://forexmarketexplained.com/

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1175–1178
DOI: 10.15439/2016F456
ACSIS, Vol. 8. ISSN 2300-5963

Minimizing Total Completion Time in Flowshop with Availability Constraint on the First Machine

Paweł Markowski, Michał R. Przybyłek
Polish-Japanese Academy of Information Technology
Warsaw, Poland
Email: {pawel.markowski, mrp}@pjatk.edu.pl

***Abstract*—A real-world company asked us to solve emerging problem related to the flow of the documents: some of the documents had been lost. We used methods of process mining to solve this problem. We found an error in the information system. Moreover, our analysis of the flow of the documents led to the conclusion that the business processes did not meet the company's needs, and with our help they were redesigned and reimplemented. This research gave an additional important insight to process mining methodology — in order to extract knowledge about business processes that is useful for the decision makers, it is crucial to create algorithms capable of collaborating with human experts.**

I. Introduction

IN CONTEMPORARY business enterprises the complexity of real-world processes has to be perceived as one of the greatest obstacles in achieving effectiveness. More and more enterprises face with the problem of redesigning their business processes in order to survive in the global economy. The competitive market creates the demand for high quality services at lower costs and with shorter cycle times. Even relatively small companies are frequently confronted — in their daily routines — with processes of very high complexity. In such environment business processes must be identified, described, understood and analyzed to find inefficiencies, which cause financial losses.

One way to achieve it is modeling. Business modeling is the first step towards defining a software system. It enables companies to look afresh at how to improve organization and to discover the processes that can be solved automatically by software that will support the business. However, as it often happens, such a developed model corresponds more to how people think of the processes and how they wish the processes would look like, than to the real processes as they take place.

Another way is by extracting information from a set of events gathered during executions of a process. Process mining [1], [2], [3], [4], [5], [6], [7], [8], [9] is a new and prosperous technique that allows to extract a model of a business process based on information gathered during real executions of the process. The methods of process mining are used when there is no enough information about processes (i.e. there is no a priori model), or there is a need to check whether the current model reflects the real situation (i.e. there is a priori model, but of a dubious quality). One of the crucial advantages of process mining over other methods is its objectiveness — models discovered from real executions of a process are all about the real situation as it takes place, and not about how people think of the process, and how they wish the process would be like. In this case, the extracted knowledge about a business process may be used to reorganize the process to reduce its time and cost for the enterprise.

Table I shows a typical event-log gathered from a flow of documents in a company. This event-log is a list of entries that contain information about observable actions of a process:

- "Actor" is someone who triggered the event; in our example an "Actor" is described by his full name (we left visible only two first letters of actors' surname)
- "Time Stamp" is the exact time of the event; in example from Table I "Time Stamp" indicates the exact date and time when the event occurred
- "Event" is an observable action of the event (we shall assume, that we are given only some rough information about the real actions); in example from Table I there are four possible types of events:
  - "*Created*" — starting point of approval process triggered by new document that was created in the system
  - "*GL app*" — first level of acceptance ; "Actor" is directly responsible for the area, which document content concerns. Approval confirms that the description coincides with reality
  - "*Quality app*" — second level of acceptance ; "Actor" is a Quality Engineer and confirms that the change does not have negative impact in quality area
  - "*APU app*" — 3rd and last approval level ; "Actor" is an Area Supervisor/Manager and confirms that the organization is ready for this kind of described change and costs regarded with it were included in forecast (money, which have to be spent are booked)
  - "*Approved*" — document has been confirmed by all required employees and is registered in SAP
  - "*Rejected*" — document has been rejected in approval path ; rejection reason attached to this document
- "Case ID" is an instance of the process that executed the event

The column "Case ID" allows us to divide a list of events on collections of events corresponding to a particular executions of the process (i.e. to a particular instance of the process). The

978-83-60810-90-3/$25.00

Table I
AN EVENT LOG GATHERED FROM THE FLOW OF DOCUMENTS.

Actor	Time Stamp	Event	Case ID
Ko* Milosz	2015-11-27 15:28	Created	1
Ko* Milosz	2015-12-08 11:57	Created	4
Gr* Grzegorz	2016-01-04 16:05	Created	2
Gr* Grzegorz	2016-01-04 16:08	GL app	2
Ch* Arkadiusz	2016-01-04 18:30	Created	3
Ch* Arkadiusz	2016-01-04 18:31	GL app	3
Po* Aleksander	2016-01-05 11:15	Quality app	2
Po* Aleksander	2016-01-05 11:18	Quality app	3
Ko* Artur	2016-01-05 11:22	APU app	3
Se* Edyta	2016-01-05 19:38	Approved	3
Si* Aneta	2016-01-11 16:08	Created	5
Si* Aneta	2016-01-11 16:19	GL app	5
Po* Aleksander	2016-01-12 10:23	Quality app	5
Ko* Artur	2016-01-12 10:56	APU app	5
Se* Edyta	2016-01-13 11:05	Approved	5
Ko* Artur	2016-02-01 08:02	APU app	2
Se* Edyta	2016-02-01 18:51	Approved	2
Ja* Iwona	2016-02-15 16:31	Created	6
Ja* Iwona	2016-02-15 16:32	GL app	6
Do* Marek	2016-02-15 17:14	Created	7
Ch* Arkadiusz	2016-02-15 20:57	Created	9
Ch* Arkadiusz	2016-02-15 20:59	GL app	9
Ma* Wiktor	2016-02-15 21:12	Created	10
Ma* Wiktor	2016-02-15 21:13	Created	11
Tu* Lukasz	2016-02-16 10:14	Quality app	8
...	...	...	...

column "Time Stamp" makes it possible to linearly order each of the collections.

Figure 1 shows a model recognized from the log presented on Table I.

The aim of this paper is to show how methods of process mining can be applied to solve difficult problems in a real-world company. A company asked us to solve emerging problem related to the flow of documents: some of the documents had been lost. In the first case an error in the information system was identified and fixed. It also turned out, that the flow of the documents had been broken and after our analysis the whole process was redesigned and reimplemented.

Other examples of process mining usability in real-world companies are described in:

- *Process Mining in Healthcare*[11]
- *Process Mining Methodology for Health Process Tracking Using Real-Time Indoor Location Systems*[12]
- *A general divide and conquer approach for process mining*[17]
- *Process Mining Applied to the Test Process of Wafer Steppers in ASML*[13]

Health care process is modern subject for process mining researchers. One of the best environments for flow tracking are areas where human behaviors influence is the highest possible. Human errors, flow deviations, workarounds can be detected by process mining tools.

Our research gave an additional important insight to process mining methodology — in order to extract knowledge about business processes that is useful for the decision makers, it is crucial to create algorithms capable of collaborating with human experts. Optimization algorithms have been researched in the area of decision support [10], but much less is known about algorithms for process mining in this context. One of the most significant methods is an explanation and justification of a mined model. This requirement stems from real business cases, where the best model becomes useless if it is not accepted by the decision maker. The lack of acceptance of a business process model can be due to several reasons: undisclosed user preferences; new constraints; a different evaluation; or simply a misunderstanding of the proposed model.

The paper is structured as follows. We checked ProM tools usability in Section II, where present complete document flow analysis. Readers can find there complex problem resolved after data extraction and analysis. We described results after document flow update that are effect of our common work with a company experts in *conformance checking* process [16]. We conclude the paper in Section III.

II. DOCUMENT FLOW ANALYSIS

Flow of the documents, which approvals are required is important due to financial control requirements. Unsupported flow occurrence might be a source of abuses, human errors or problems to meet customer expectations. Corporations have to take care about transparency of financial status required by internal procedures and headquarters policies. Internally developed solution has to ensure document approvals in line with employees permissions. This kind of software is vulnerable to errors. Bug can be source of problems like embezzlement results, communication errors or wrong conclusions. Part of organization that is responsible for programming or management has to ensure that the implemented program works according to the established process.

A. Background

Real-world company has a problem with documents flow. They had mismatch between documents' flow system results and Financial Department's reports. Company's manager contacted us and ordered a complete service of process analysis. The company has implemented Microsoft Sharepoint to ensure proper flow of documents. Each document contains information about produced parts and represents their market value in currency. Developed addon is programmed with optional paths, which are depended on the value of created document. There is additional flow rule that the specialist opinion is needed in particular cases. This specialist is quality engineer. The higher value of a document is, the longer approvals path is required and the people involved in the acceptance process of the document have to be higher in the organization structure. Document creation or approval are triggers for system that generates event logs.

Financial controller and area supervisor have found a problem with budget forecasts. Anomaly was detected by differences in spending calculated by Financial Department and other department manager. Rules in the system looked fine, because each spending value was bigger than specified amount that had to be approved by the department manager. Operation Manager was informed about the situation and commissioned

the analysis. The analysis detected source of anomalies and after that Operation Manager gave instructions how to prevent such anomalies in the future.

Figure 1 presents a diagram mined by ProM Casual Activity Graph. A diagram is consistent with the flow implemented in Sharepoint. It shows that there are 3 levels of approval, but flow depends on specificity of a particular document, which path might be different. Collected logs contains 3 roles in approval system.

B. First level of process analysis CSV to XES

- Events log extraction from Sharepoint application to CSV file.
- CSV text file have to be imported to ProM and transformed to XES
- Analysis of imported data has to be preceded by questions, which may show proper direction of next mining methods

In process mining it is important to understand the data that have been collected. Without this knowledge it may be hard to extract any reasonable conclusion. The analysis of received data should be supported by a person who has theoretical knowledge about the assumptions. Too big deviations in comparison to process draft in organization might warn us about low data quality, which can be source of negative consequences in further analysis.

C. XES log summary

After CSV event log transformation to XES format ProM gives access to the document summary. One of the views is XES log summary where we found information about wrong starting point with 0,06% occurrences value. APU app is 3rd level approval in Figure 1. This is a point, which has to be verified during process exploration. It looks that a process was started from 3rd level acceptance even before its creation.

D. Directly follows graph extracted by ProM

ProM addon generates diagrams where nodes are events and edges are connections between them [15]. This visualization presents process steps flow, which are connected by arrows with labels presenting quantity of occurrences. Figure 1 is a directly follows graph mined from XES log. It transfers event logs into visualization of process steps.

Connection from 3rd role in acceptance process with an event announcing the creation of the document is an anomaly in this flow. The assumption is that the document has to be created before approval. Both in XES summary (Section II-C) and in Figure 1 is visible unusual event.

E. Extracting single case visualization

Process mining helped us with understanding the above anomaly, and we used an inductive visual miner (IvM [14]), which is dedicated ProM tool. It is able to extract information about process for every single instance and to visualize it. IvM allows to track the flow for every process step by graphical user interface. Inductive visual miner includes a functions

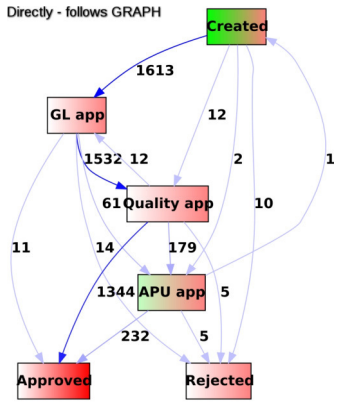


Figure 1. Process discovered by ProM from event log presented in Table I.

based on click event that allow us to extract additional information about process steps or single cases.

Figure 1 APU app(roval) was performed before document creation. Logs visualization shows that the next approval steps were performed even with this anomaly. Analyze of this case led to modifications in path structure. Additional check conditions were added to ensure proper document flow. The source of this problem was in improper database management and software bug. System changed first and second level of approval fields to empty but APU acceptance field did not. Modified document has APU acceptance from previous flow cycle. Trigger for process of documents' approval is document creation. After that instance of process have to be verified by authorized employees. Graph on Figure 1 shows that the created documents are addressed to GL app (first level of approval) in most of cases. This people are able to physical confirmation of compliance created document with the actual situation in the factory. They are working in 3 shift model so verification of documents is possible 24h per working days.

F. System validation after updates

Event logs were imported to ProM two months after system update. In order to confirm that a document approval path is in compliance with procedures. XES event log summary shows single starting point "Created". This summary confirms that the starting point is common and correct for each process instance. Process mining tools help corporation experts in analysis approvals path. Process approval path was redesigned to be in compliance with organization structure. This change is reducing human error risk by change that requires minimum 2 employees approval. Each step in approval path has a possibility to reject a document as like as in previous system version. Obligatory Quality approval step increased description coincidence with reality, because person responsible for that in 1st step knows about further Quality verification.

III. CONCLUSIONS

We investigated how methods of process mining can be applied to solve difficult problems in a real-world company. The paper shows that ProM functionality enables us to take

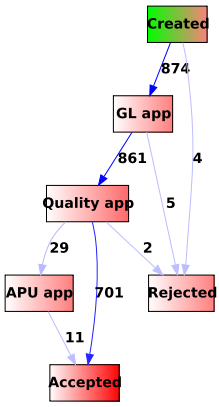


Figure 2. Document flow after system update

a look on processes not only from the designer perspective but also from the perspective of the production environment. IT systems are integral part of companies and sources of important data. Such data should be collected by system administrators, as they are valuable input in process extraction. Event logs gathered during execution of business processes present the reality in an objective way. Therefore, it is crucial to use them in further process optimization. They allow to look at the business process from the production perspective. Process mining has to be supported by people, who understands the relationship between discovered actions. The corporation that we worked with, solved emerging problem related to the flow of documents: some of the documents had been lost. An error in the information system was identified and fixed. It also turned out, that the flow of the documents had been broken and after our analysis the whole process was redesigned and reimplemented. The results obtained in process mining were the most important inputs for the system developers. The developers were able to prepare patches that fixed the problems. Process extraction realized by process mining tools was helpful in problem understanding and fixing. User friendly flow's representation combined with experts knowledge can be the key to successful system patch preparation and implementation.

Our research gave an additional important insight to process mining methodology — in order to extract knowledge about business processes that is useful for the decision makers, it is crucial to create algorithms capable of collaborating with human experts. Optimization algorithms have been researched in the area of decision support, but much less is known about algorithms for process mining in this context. One of the most significant methods is an explanation and justification of a mined model. This requirement stems from real business cases, where the best model becomes useless if it is not accepted by the decision maker. The lack of acceptance of a business process model can be due to several reasons: undisclosed user preferences; new constraints; a different evaluation; or simply a misunderstanding of the proposed model. In future work we should focus more on human-computer interaction in process mining algorithms and systems.

REFERENCES

[1] W.M.P. van der Aalst, *Process Mining: Discovery, Conformance and Enhancement of Business Processes*, Springer Verlag, 2011.
[2] A.J.M.M. Weijters, W.M.P. van der Aalst, *Process Mining: Discovering Workflow Models from Event-Based Data*, Proceedings of the 13th Belgium-Netherlands Conference on Artificial Intelligence, Maastricht, pp 283-290, 2001.
[3] A.K.A de Medeiros, B.F. van Dongen, W.M.P. van der Aalst and A.J.M.M. Weijters, *Process Mining: Extending the alpha-algorithm to Mine Short Loops*, BETA Working Paper Series, WP 113, Eindhoven University of Technology, Eindhoven, 2004
[4] W.M.P. van der Aalst, A.H.M. ter Hofstede, B. Kiepuszewski, A.P. Barros, *Workflow Patterns*, BPM Center Report BPM-00-02, BPMcenter.org, 2000.
[5] W.M.P. van der Aalst and A.H.M. ter Hofstede, *Workflow Patterns: On the Expressive Power of (Petri-net-based) Workflow Languages*, BPM Center Report BPM-02-02, BPMcenter.org, 2002.
[6] W.M.P. van der Aalst, A.J.M.M. Weijters, L. Maruster, *Workflow Mining: Discovering Process Models from Event Logs*, BPM Center Report BPM-04-06, BPMcenter.org, 2004.
[7] M.T. Wynn, D. Edmond, W.M.P. van der Aalst, A.H.M. ter Hofstede, *Achieving a General, Formal and Decidable Approach to the OR-join in Workflow using Reset nets*, BPM Center Report BPM-04-05, BPMcenter.org, 2004.
[8] W.M.P. van der Aalst, A.K. Alves de Medeiros, A.J.M.M. Weijters, *Process Equivalence in the Context of Genetic Mining*, BPM Center Report BPM-06-15, BPMcenter.org, 2006.
[9] W.M.P. van der Aalst, M. Pesic, M. Song, *Beyond Process Mining: From the Past to Present and Future*, BPM Center Report BPM-09-18, BPMcenter.org, 2009.
[10] A. Wierzbicki, M. Makowski, J. Wessels, *Model-Based Decision Support Methodology with Environmental Applications*, Springer, Series: Mathematical Modelling: Theory and Applications, Vol. 9, Berlin Heidelberg, 2000.
[11] Mans, R.; van der Aalst, W.; Vanwersch, R. *Process Mining in Healthcare*, Springer Briefs in Business Process Management; Springer International Publishing: Cham, Germany, 2015
[12] Carlos Fernandez-Llatas, Aroa Lizondo1, Eduardo Monton, Jose-Miguel Benedi, Vicente Traver *Process Mining Methodology for Health Process Tracking Using Real-Time Indoor Location Systems* Sensors, 11/2015; 15(12):29821-29840. DOI: 10.3390/s151229769
[13] A. Rozinat, I.S.M. de Jong, C.W. Gunther, and W.M.P. van der Aalst *Process Mining Applied to the Test Process of Wafer Steppers in ASML* IEEE Transactions on Systems, Man and Cybernetics - Part C, 39:474–479, 2009.
[14] Leemans, S.J.J., Fahland, D., Aalst, W.M.P.v.d. *Process and Deviation Exploration with Inductive Visual Miner* Proceedings of the BPM Demo Sessions 2014 Co-located with the 12th International Conference on Business Process Management (BPM 2014), Eindhoven, The Netherlands, September 10, 2014. (2014) 46
[15] Leemans, S., Fahland, D., van der Aalst, W. *Exploring processes and deviations* Business Process Management Workshops (2014)
[16] W. M. P. van der Alst *A general divide and conquer approach for process mining* Federated Conference on Computer Science and Information Systems, 2013, pp. 1-10.
[17] P. Homayounfar *A general divide and conquer approach for process mining* Proceedings of the Federated Conference on Computer Science and Information Systems, 2012, ISBN 978-83-60810-51-4, pp. 1135–1140

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1179–1188
DOI: 10.15439/2016F235
ACSIS, Vol. 8. ISSN 2300-5963

Project Communication Management Patterns

Karolina Muszyńska
University of Szczecin, Faculty of Economics and Management
ul. Mickiewicza 64, 71-101 Szczecin, Poland
tel. +48914441911
e-mail: karolina.muszynska@usz.edu.pl

***Abstract*— In present, dynamically developing organizations, that often realize business tasks using the project-based approach, effective project management is of paramount importance. Numerous reports and scientific papers present lists of critical success factors in project management, and communication management is usually at the very top of the list. But even though the communication practices are found to be associated with most of the success dimensions, they are not given enough attention and the communication processes and practices formalized in the company's project management methodology are neither followed nor prioritized by project managers. This paper aims at supporting project managers and teams in more effective implementation of best practices in communication management by proposing a set of communication management patterns, which promote a context-problem-solution approach to communication management in projects.**

I. Introduction

COMMUNICATION management is deemed by many as one of the most important knowledge areas in project management and a very complex one at the same time. It is affected by many factors, like characteristics of project stakeholders, project environment, project communication structure, communication properties, physical and psychological barriers [1]. Research has shown that there is a direct connection between communication and a project's outcome, which is determined by the design of the communication environment of the project [2]. Project communication and networking skills are considered to be the life blood of project management leadership [3] and awareness of the potential offered by efficient communication is an essential prerequisite for success in the business world [4].

Project management methodologies, frameworks and sets of principles like Project Management Body of Knowledge, Prince 2, Adaptive Project Framework, Agile Software Development, Scrum, and others, include rules, hints and procedures regarding various communication management aspects, which in most cases should be sufficient to properly manage communication in a project team. The reason why this is not always the case is that, many IT companies do not actually follow any of these methodologies when realizing projects (for example, 30% of the companies surveyed in [5]) and those that do, tend to concentrate on other project management knowledge areas, which seem more important, like scheduling, cost management, etc. There are also numerous cases of project failures [6], that could be linked to poor communication management (among others [7], [8]).

Thus it seems important to constantly promote good communication management practices and look for new ways to support project managers and team members in better realization of communication and documentation processes in their projects. That is the main goal behind the communication management patterns proposed in this paper – to give project teams an additional tool in a form of a list of patterns for controlling, managing and effective realization of communication and documentation processes. The idea behind patterns is that they provide general, reusable solutions to common problems, and as such are useful for the project team. Communication management patterns described in this paper are based on two main sources of information - communication management best practices identified in subject literature and results of a survey conducted among IT project managers.

The following, second section of the paper, provides evidence on the significance of project communication management knowledge area based on existing literature. In section III, the definition of a pattern is given and some examples of pattern usage are described. The essential fourth section of the paper begins with a definition of the project communication management pattern and on that base subsequent communication management patterns are defined. Conclusions and future research directions end the paper.

II. Significance Of Project Communication Management

The successful implementation of a project depends on its appropriate management in a number of areas, as described in detail by e.g. Kerzner [9], Schwalbe [10], and Meredith and Mantel [11]. One of the areas of project management identified within numerous methodologies and frameworks is

This publication was financed from the funds of the Faculty of Economics and Management, University of Szczecin.

communication management, which is considered to be of crucial importance to the success of a project (among others [12], [13]), in particular IT projects [14] especially those carried out by dispersed teams [15], [16], [17], [18], [19]. On the one hand importance of this knowledge area is emphasized by most stakeholders, but on the other hand, the communication processes and practices formalized in the company's project management methodology are neither followed nor prioritized by project managers [20]. For instance, recent research on utilization of project communication management methodologies in industrial enterprises in Slovak Republic revealed that 66% of the surveyed enterprises had not prepared any written document (methodology, process steps, etc.) to manage project communication [21].

Also Papke-Shields and co-authors, in their research on the use of project management practices and the link thereof to project success, discover that practices related to communication are not given enough attention, while at the same time communication practices are found to be associated with most of the success dimensions [22]. Most of the communication process in a project is usually done without proper planning, driven mostly by personalities and preferences rather than by needs, protocols and procedures [23].

Effective communication techniques and appropriate leadership styles are emphasized by Nguyen as the success factors for building and managing high performance global virtual teams [24]. Communication management is highly influenced in intercultural project teams by such factors as language, race, age, gender, religion, beliefs, habits, etc., whose analysis is essential if the project is to be accomplished with success [25].

According to PMI's Pulse research, 55 percent of project managers agree that effective communication with all stakeholders is the most critical success factor in project management [26]. Effective project communications ensure that the right information reaches the right person at the right time and in a cost-effective manner. Communication is the key to keeping team members, managers, and stakeholders informed and on track to pursue the project objectives, as well as to identifying issues, risks, misunderstandings, and all other challenges to project completion. Effective communication is a critical element of team effectiveness, both in traditional and virtual teams [27] and it is much more than message exchange, rather a way in which project managers generate the grounds for a project [28].

III. Definitions And The Use Of Patterns

One of the general definitions of a pattern states that it is "a regular and intelligible form or sequence discernible in the way in which something happens or is done" or "an excellent example for others to follow" [29]. Design patterns are used to represent knowledge that is based on experiences captured in several real world projects and is widely accepted. This representation is often used for describing and presenting the gained knowledge. There are similar concepts to the concept of a pattern – success factor, success models, success measures, reference architectures, best practices, worst practices, barriers, facilitators or incentives [30].

Different definitions for a pattern exist, but they all include a common ground – patterns are general, reusable solutions to common problems and are dependent on their context [31]. They are based on the philosophy of pattern languages, first proposed by architect Christopher Alexander, which is now widely applied in many other professional areas to encompass creative human actions ([32] and works cited therein).

The following subsections describe the structure and format of patterns defined in three knowledge areas – design patterns in software engineering, knowledge management patterns and collaboration patterns. These examples have served as a reference for the structure of communication management patterns defined in the subsequent section.

A. Design Patterns in Software Engineering

In the software discipline, a pattern describes a problem, which occurs recurrently and supports a solution to that problem in a given context. The pattern has four essential elements: the pattern name, the problem, which describes when to apply the pattern (it explains the actual problem and its context), the solution, which describes the elements that make up the design, their relationships, responsibilities and collaborations (this is an abstract description, a template of a solution), and the consequences – the results and trade-offs of applying the pattern.

To describe each design pattern, a consistent format has been used, including specific sections like: pattern name, intent (explains what the design pattern does, what particular issue/problem it addresses), motivation (a scenario illustrating a design problem and how the pattern solves it), applicability (situations when the design pattern can be applied, examples of poor designs where the pattern could be applied), structure (a graphical representation of classes in the pattern, accompanied with interaction diagrams), participants (classes participating in the design pattern and their responsibilities), collaborations (how the participants collaborate to carry out the responsibilities), consequences (the trade-offs and results of using the pattern), implementation (pitfalls, hints, techniques useful in pattern implementation), sample code, known uses (examples of the pattern found in real systems, related patterns (which patterns are closely related to each other). Each defined design pattern is described according to the above mentioned sections, which makes them easier to use, learn and compare [33].

B. Knowledge Management Patterns

Knowledge management patterns state lessons learned and best practices for the structuring of knowledge, the design of

knowledge management systems, and the development of underlying ontologies. Patterns in knowledge management represent also a form of language that helps knowledge engineers to communicate about knowledge and knowledge management systems.

A knowledge pattern is defined as a general, proven, and beneficial solution to a common, reoccurring problem in knowledge design, i.e., the structuring and composition of the knowledge or the ontology defining metadata and potential relationships between knowledge components. Knowledge management patterns are described in seven groups regarding different aspects of knowledge: content, usage, ontology, presentation, transfer, knowledge management systems organization and social knowledge management. Each pattern is described according to a template including the following sections: name, issue (problem addressed by the pattern), q-effect (what knowledge quality aspects are affected by the pattern and if it is a positive, negative or neutral effect), solution (principal solutions underlining the pattern), causes (basic causes of the pattern) [30].

C. Collaboration Patterns

Collaboration patterns strive to verbalize tips on achieving teamwork that creates new values. They are used to help teams achieve creative collaboration through their interactions and discover methods to engage in effective teamwork.

Each collaboration pattern comprises of two main parts – the first one, giving a brief idea of the pattern, including the following elements: name, one-liner (short description of the pattern), illustration and quotes related to the pattern, and the second part, offering more details, with the following sections: context, problem, forces, solution, actions and consequences, which describe how things can change when a certain pattern is applied [34].

There are also other views on collaboration patterns, like the Schadewitz's design patterns for cross-cultural collaboration [35].

According to the classification proposed by Verginadis et al. [36], the project communication management patterns, specified in the subsequent section, could be considered as a specific kind of patterns in collaborative work, and precisely classified as "an approach that aims to directly assist participants", and "which requires manual intervention".

IV. Project Communication Management Patterns

A. Definition of a Project Communication Management Pattern

The definition of the project communication management pattern proposed in this subsection is a result of the analysis of patterns and their frameworks developed in different disciplines, and combining selected aspects of these patterns with project communication management characteristics and practices.

Project communication patterns have been grouped into four categories according to the communication management practice categories described in [37] – informational (regarding generation, collection, dissemination, storage, and disposition of project information), strategic (connected with communication planning and project environment), emotional (concerning the building of trust and relationships) and practical (connected with clear and positive communication and behavior rules). Within each category, several communication management patterns were defined. Each pattern comprises of the following sections: pattern name, context, problem, solution, q-effect (what communication quality aspects are affected by the pattern and if it is a positive or a negative influence), applicability (situations, teams and projects where the communication management pattern should be applied), participants (parties participating in the communication/documentation process and their responsibilities), consequences (the trade-offs and results of using the pattern), implementation (pitfalls, hints, techniques useful in the pattern implementation), and related patterns.

For specifying q-effect the following communication quality aspects were considered: clearness and cohesion, adequate level of detail, right time, meeting needs of communicating participants, engaging right people, guarantee of uniform understanding of the content, communication workflow supporting openness, redundancy and feedback.

Solution within a pattern describes what actions should be undertaken to realize the pattern and the communication management goal that it supports.

The project communication management patterns described in the subsequent subsections are based on two main sources of knowledge. The first source are communication management best practices described in literature [38]–[43], [16] and thoroughly discussed in [37], and the second source are the opinions gained from practitioners (mostly project managers realizing IT projects) from 10 national and international IT companies operating in Poland. A structured interview with both closed and open-ended questions was used to obtain their opinions.

B. Informational Communication Management Patterns

Within the informational category, three communication management patterns have been specified: Communication schedule, Project knowledge center and Diversity of communication means. According to the survey, carried out among practitioners, all the following patterns are known to almost all of them and used in their companies. In one case it has been stated that the Communication schedule pattern is an intrinsic element of the communication plan, which is prepared at the start of the project realization.

Table I provides characteristics of the three project communication management patterns from the informational category, developed according to the defined template.

For the sake of brevity the pattern section names were abbreviated as follows: Cx (context), Pr (problem), S (solution), Q (q-effect), A (applicability), P (participants), Cq (consequences), I (implementation) and RP (related patterns).

TABLE I.
COMMUNICATION MANAGEMENT PATTERNS WITHIN THE INFORMATIONAL CATEGORY

Communication schedule	
Cx	The project team is dispersed, some team members are in different time zones; according to the project communication plan project partners should inform each other of the project status to get feedback and encourage involvement.
Pr	Communication between team members is too scarce, team members limit communication to sending reports, while direct communication takes place only in emergency situations.
S	Prepare a communication time schedule, including bilateral communication between particular team members, as well as multilateral audio/video conferences among wider forum of team members; communication participants possibilities and preferences concerning communication medium and time zone shifts should be taken into account.
Q	Positive on the following communication quality aspects: up to date information on the project tasks status; redundancy, feedback. Negative on the following communication quality aspects: in case of multilateral audio/video conferences too many participants taking part may cause the communication to be ineffective and irritating (technical problems are more likely to appear) and not engaging the right people.
A	The pattern can be used for any kind of project and team, although it is especially useful for dispersed teams and bigger projects longer than three months.
P	Any team member that is included in the communication time schedule; team members are responsible for adhering to the time schedule or timely informing about any derogations; a very important issue in realizing this pattern is engaging only the concerned team members in multilateral conferences (thematic groups).
Cq	Ensures regular communication among team members, adjusted to their working day schedules and communication preferences, and keeps everybody informed about the status of project tasks and encourages instant feedback.
I	Setting up a communication time schedule requires time, effort, cooperation and goodwill of team members, so that it is adhered to and beneficial; the more parties and localizations the more difficult it becomes; it should be agreed upon during the project kick-off meeting, accompanied by a clear message of its goal and instructions of realization; using such tools as shared calendars and communication matrix can be useful.
RP	*Clear rules at the start*
Project knowledge center	
Cx	Communication in the project team is performed in different ways; many people use e-mail as the primary communication medium, and send various elements of the project documentation this way. Others prefer communication via Instant Messaging (IM) and attaching files to conversations. Still others would rather talk on the phone and deliver files on a pen drive.
Pr	Many elements of the project documentation remain only in mail boxes/computers/pen-drives of individual team members and project knowledge in their heads; different versions of the same documents are created and their subsequent synchronization is very cumbersome; some information is lost or finding it is time-consuming; certain project knowledge is lost when a team member leaves the team.
S	Ensuring a project repository – project knowledge center, where project files are placed, stored and shared; a web portal with a wiki feature, group-work tool or project management software with the file versioning and change tracking functionalities can be used for that purpose; to wean team members from sending project documents as attachments to e-mails, a special function could be embedded in the e-mail program, which would display a message asking the user if the attached file should not rather be placed in the repository, instead of being sent.
Q	Positive on the following communication quality aspects: clearness and cohesion, up-to-date project documentation, adequate level of detail and communication workflow supporting openness. Negative on the following communication quality aspects: in case the repository is unordered and unclear the adequate level of detail, clearness and cohesion are no longer attained.
A	The pattern should be used for any kind of project and team, because even small projects and small teams produce project documentation which should be made available to the team and the cumulated knowledge may prove useful in future projects.
P	All team members producing any kind of project documentation or content; team members are responsible for uploading any project-related documentation in an orderly manner, established upfront or developed in the initial phase of the project realization.
Cq	Ensures a common project documentation reference center available to all team members, taking into account given user access rights, with up-to-date project documentation and orderly history. Project documentation is not hidden in the mail boxes of individual team members and a project knowledge base is being built. Team members do not have to communicate or send specific information separately to all interested parties but just link them to the appropriate place in the project knowledge center.
I	Setting up a project knowledge center requires using appropriate tools and setting certain procedures, so that the repository is easy to use and is effective in storing and sharing information; it is usually a software application chosen and used by the project team for many projects. Problems which may arise concern effective organization of the repository and a need for training team members how to use versioning tools. The chosen tool usually requires configuration effort and expertise and some systems are expensive. Sometimes documents sharing between the customer and the developer is not at all possible due to security issues.
RP	-

Diversity of communication means	
Cx	It happens that tools used in the project team impose the way of communication among team members, limiting it mainly to written communication (e.g. to have a permanent evidence of discussions and arrangements). Team members hardly talk to each other personally; this is particularly common in the case of international and distributed teams.
Pr	Focusing mainly on one way of communication, whether oral or written, hampers the implementation of the project, in the first case because of the transience of oral arrangements, in the second case, because of possible problems with understanding the intentions or the lack of instant response. In the case of written real time communication (using IM) the typing speed can be a problem, while for oral communication a poor knowledge of a foreign language can be a barrier.
S	The solution to the problems associated with using mainly one communication means is to promote diversity in the ways of communication, and while preserving the principles of the *Project knowledge center* pattern, emphasize the importance of oral communication, which should support mutual understanding between team members and unite the project team.
Q	Positive on the following communication quality aspects: meeting needs of communicating participants (as some team members are comfortable with written communication, while others need to communicate also orally), guarantee of uniform understanding of the content, redundancy and instant feedback (in case of oral communication). Negative on the following communication quality aspects: in case of excessive diversity, project documentation consistency and cohesion is hard to maintain.
A	The pattern should be used carefully, taking into account the communication culture of the project team, level of project language knowledge of team members and security issues (some communication tools may be considered as not secure).
P	All team members, and especially the project manager should take into account personal predispositions of each member to avoid forcing them to use communication means which they are not comfortable with.
Cq	Diversity of communication, on the one hand enriches and facilitates communication among team members, but on the other hand, if not appropriately managed, may cause communication chaos, with some information being lost in oral conversations or time wasted during too frequent and ineffective meetings.
I	The written form of communication, especially the part concerning communicating project results and producing project documentation, should be arranged at the beginning and organized into effective and easy to understand and follow procedures, which should be realized by all team members (see *Clear rules at the start* pattern). One of the most effective oral communication means are stand-up meetings, known from the agile project management approach, which are a quick and effective way to find out what is the status of project tasks, who needs help, or who is not working properly. It is also a perfect way for team members to get to know each other. It is however important that the project manager, or leader does not dominate these meetings. In case of dispersed teams video-conference stand-up meetings can be organized. In case of traditional meetings their costs (time, travel) must be taken into account and planned in advance.
RP	*Clear rules at the start, Project knowledge center*

C. Strategic Communication Management Patterns

Another three communication management patterns have been defined within the strategic category: Clear rules at the start, Cultural and language competencies and Client's power scope. These patterns were also acknowledged and used by most of the surveyed practitioners, although Cultural and language competencies pattern was not used by more than one third of them because their project teams were not culturally or linguistically diverse. Others argued that both Clear rules at the start, as well as Client's power scope, are part of the communication plan and need not to be separately described, but that is the role of patterns to highlight specific problematic areas to which solutions are proposed – in the case of Clear rules at the start pattern, preparing a high-quality communication plan is actually the suggested solution.

Table II provides an overview of the three project communication management patterns from the strategic communication management practice category.

TABLE II.
COMMUNICATION MANAGEMENT PATTERNS WITHIN THE STRATEGIC CATEGORY

Clear rules at the start	
Cx	It sometimes happens that while planning various aspects of the project (project tasks, responsible team members, schedule and budget), the area of communication and documentation management is neglected. There is no regular contact with the client to inform them about the progress of the project and for keeping in touch for quick reaction to possible changes and new requirements. If there are no assigned tasks and people associated with the management and implementation of communication, nobody will take care of it.
Pr	There are no designated persons and tasks related to planning and managing communication and documentation processes. Team members feel no need to communicate the status of their tasks, nor do they feel responsible for informing the client about the status of the project.
S	Development of a clear, practical and high-quality communication plan with assigned persons responsible for communication management, description of communication and documentation tasks – the ones to be carried out by specific individuals and those which are the responsibility of all members of the project team. In the case of distributed teams, it is particularly important to include the *Communication schedule* pattern. In the case of various language speaking teams, a common project communication language should be established.
Q	Positive on the following communication quality aspects: meeting needs of communicating participants. Negative on the following communication quality aspects: in case of excessive formalism and bureaucracy participants may be discouraged to communicate effectively.
A	The pattern should be used for any kind of project and team, although it is especially useful for teams with different working cultures and fixed price projects.

P	Every project stakeholder should be included in the communication plan. All persons assigned to any communication and documentation tasks should be clearly informed of their responsibilities at the beginning of the project realization.
Cq	Ensures that all team members and project stakeholders know their communication and documentation responsibilities. Client is instantly informed about the status of the project tasks. It is, however, important to let the communication plan evolve and alter throughout the project, to make it better tailored to the given project and team.
I	Preparing a high-quality communication plan requires time and effort, so that it is then easy to realize and not burdensome for the project team; too much formalism may discourage the team; the communication plan should be communicated already during the project kick-off meeting, or at least during the initiation phase of the project.
RP	*Communication schedule*
Cultural and language competencies	
Cx	In the case of multicultural teams or teams whose members come from different language areas, difficulties in communication, due to problems with mutual understanding, may appear. This situation may affect the members of one team or members of two collaborating teams (the client's side team and the developer's side team).
Pr	Problem with the lack of cultural and/or language competence of team members, which hinders communication and mutual understanding, and thereby successful realization of the project.
S	Team members should be prepared for the environment in which they are going to work and familiarized with the rules and customs prevailing in the country of other team members. Necessary language skills should also be checked to ensure comfortable communication.
Q	Positive on the following communication quality aspects: clearness and cohesion, meeting needs of communicating participants, guarantee of uniform understanding of the content.
A	The pattern is suitable for teams working in multicultural and international projects.
P	All team members engaged in an international/multicultural project, who are responsible for communication and documentation tasks.
Cq	Culturally and linguistically competent team members facilitate communication and make the internal cooperation easier and more efficient.
I	Having culturally and linguistically competent team members is not always possible or easy to achieve. Learning a foreign language is a long process and getting to know different culture is difficult. If there is at least one competent person in the project team they can train the team. It is also a good idea to promote and use "project culture", which is above local cultures of particular team members. In case of language problems in written communication, such tools as online translators or spell-checkers can be used. For oral communication a linking-person, who can freely communicate with both parties, is a solution.
RP	-
Client's power scope	
Cx	A strategic issue in project implementation is to clarify the scope of power on the client's side. This is particularly necessary in case of problems with precise determination of requirements or their frequent changes. In order to ensure smooth communication between the project team and the client's team, it is important to establish the competence scope.
Pr	The project team does not know who, on the client's side, is responsible for making and communicating decisions concerning the project, as well as who to contact in problematic issues.
S	Project manager must ensure that the client has appointed a person or persons who will be responsible on the client's side for making and communicating decisions throughout the project, and who should be contacted in problematic issues.
Q	Positive on the following communication quality aspects: clearness and cohesion, engaging the right people, communication workflow supporting openness.
A	The pattern is applicable to any team and project, because clear definition of responsibility for making decisions is always desirable.
P	All team members, project manager, client's team.
Cq	Clearly defined responsibility for certain competence areas. Clear decision-making procedure. Clearly defined deadlines for document verification and approval. Effective problem reporting.
I	Client's power scope should be defined at the very beginning of the project realization (communication aspects should be included in the communication plan – *Clear rules at the start* pattern). It is however important to be flexible and ready for negotiations of responsibilities and authorities. It is a good practice to assign representatives on both sides – mutual counterparts. Using the same standards for communication or process templates can also be helpful. Problems in implementing the pattern may arise in the case of dominant position of the client and reluctance to compromise, fear of making decisions, not properly chosen/prepared team on the client's side, conflicts within the client's team.
RP	*Clear rules at the start*

D. Emotional Communication Management Patterns

The emotional category of communication management practices comprises of another three patterns: Fostering direct communication, Visits and team rotations and Appreciating the team. Most of the practitioners, who were surveyed, knew all three patterns and used them in their project teams. One of them was however very sceptic towards the Fostering direct communication pattern, claiming that the pattern should actually be quite the contrary, because people tend to waste a lot of time on unproductive and ineffective talks and meetings.

The above mentioned patterns from the emotional communication management practice category are described in table III.

TABLE III.
COMMUNICATION MANAGEMENT PATTERNS WITHIN THE EMOTIONAL CATEGORY

Fostering direct communication	
Cx	Absorbed in work and rushed by deadlines, project team members often do not have time for direct talks with each other or with team members on the client's side. It also happens that the management restricts such direct contacts (chats in the hallway or through IM), treating it as a waste of time and delaying tasks.
Pr	Team members are alienated and feel discomfort associated with inability to satisfy human needs, these associated with direct contact with another person. Such reduction of direct communication restrains the team from uniting, understanding each other's needs and hinders comprehension.
S	Project manager should promote direct communication between team members, as well as with members of the team on the client's side. This may be achieved in various ways, through the use of *Communication schedule* pattern that takes into account the direct methods of communication (in the case of distributed teams, at least audio- or videoconferencing), through the use of social networking tools ("virtual water cooler"), where team members could talk informally, and through regular project meetings (stand-up meetings, reviews).
Q	Positive on the following communication quality aspects: meeting needs of communicating participants, guarantee of uniform understanding of the content, communication workflow supporting openness, redundancy, feedback.
A	The pattern should be used for any kind of project and team, although in the case of distributed teams "direct" usually means audio or videoconferences, as face-to-face meetings are costly and time-consuming.
P	All project team members should be able to take advantage of this pattern, although project manager should track the impact of direct communication on project performance.
Cq	May prove very beneficial to the project and the team if properly used; both formal and informal direct communication fosters better mutual understanding, team uniting, issues resolving. It must be however properly managed and monitored to prevent team members from wasting too much time and delaying realization of tasks – this mainly concerns poorly prepared meetings.
I	As far as formal communication is concerned, stand-up meetings can bring much profit, because they convey both information and emotions and let the team get to know each other better. Informal communication can be supported by social networking tools, informal chat-rooms or a common meeting room (in case of local teams). Use of the *Visits and team rotations* pattern is also a method of fostering direct communication and letting different team members get to know each other. Access to direct communication should also be made easy by supporting a list of team members' phone numbers, IM contact details, etc.
RP	*Communication schedule, Visits and team rotations, Diversity of communication means*
Visits and team rotations	
Cx	A very important aspect in project realization is the mutual trust and understanding of team members. If the direct contact of the contractor's team with the client's team is limited to the kick-off meeting and a few other project meetings, then it is hard to achieve.
Pr	Lack of trust and willingness to communicate within the project team, because of the lack of direct contact and familiarity of team members.
S	A possible way to solve this problem are regular visits of individual team members at the client's/contractor's site, as well as delegating a team member to the client's/contractor's site to facilitate communication. In the latter solution, rotation can also be used, so that different team members can get to know each other, and thus break the communication barrier.
Q	Positive on the following communication quality aspects: meeting needs of communicating participants, communication workflow supporting openness, feedback.
A	The pattern is suitable for certain projects and environments, because usually a single team member has too scarce knowledge of the project and sending the whole team is neither possible nor effective; beneficial mostly in big projects with distributed teams.
P	Only willing team members should be chosen for delegation to other locations, to avoid discontent and frustration experienced by people forced to leave their home city and family for a longer period of time. Shorter visits should be realized by all key team members.
Cq	Building non-professional relations among team members fosters effective and direct communication (relation with *Fostering direct communication* pattern). Delegated team members facilitate communication between the client's team and the contractor's team.
I	Realization of the pattern should be preceded by an analysis of predispositions and willingness of individual team members to delegations, so that appropriate plan of visits and team rotation can be developed and included in the budget. In reasonable circumstances bonuses or family delegations can be offered.
RP	*Fostering direct communication*
Appreciating the team	
Cx	In the course of project realization team members notice errors or possibilities for better solutions to various problems. However they do not always have the opportunity to express their views, to give advice or share opinions, or they do not know how and where it can be done. With some team members this may cause frustration and decrease motivation, others may not care about it, but the failure to take their opinions into account may prove to be unfavorable for the project.
Pr	The project management does not enable team members to share opinions, to formulate proposals or comments related to the implementation of the project. They cannot express their feelings, thoughts and remarks, and feel unappreciated and their motivation to work decreases.
S	Encouraging team members to share their thoughts, remarks and opinions. This can be achieved, inter alia, by reserving a project portal section for this purpose. This section can also be used by the project management to formulate requests for support and advice, which would give the team a sense of appreciation of their value and trust.
Q	Positive on the following communication quality aspects: meeting needs of communicating participants, communication workflow supporting openness, feedback.

A	The pattern is applicable to any team and project, because every team member should have an opportunity to share their thoughts and opinions and all team members need to feel appreciated. It is especially beneficial in long-term projects where constant improvement of work quality should take place.
P	All team members should have the opportunity to take advantage of this pattern. Project manager should be open to remarks from the team.
Cq	Appreciated project team is motivated to work towards successful realization of the project; useful remarks and suggestions are collected and may be applied to foster better project development; alarming situations are exposed and appropriate actions can be undertaken.
I	The pattern may be realized in many different ways – devoting time during project meetings for team opinions, remarks, suggestions; reserving a project portal section for this purpose or a thematic mailbox; organizing surveys, retrospection sessions. This pattern is connected with *Clear rules at the start*, *Communication schedule* and *Fostering direct communication* patterns, because all of them strive for letting team members communicate what they want, need or should communicate in a way which is the most suitable for them. The effort of organizing and analyzing surveys, mailboxes, retrospection sessions or portal sections should be included in the budget and schedule plan, to avoid situation that all information is collected in vain.
RP	*Clear rules at the start, Communication schedule, Fostering direct communication*

E. Practical Communication Management Patterns

Within the last, practical category, two more communication management patterns have been identified: Basic communication principles and Synchronous working environments. Only one of the surveyed practitioners has not heard about or used the first pattern, while the second one was recognized by all, but not used by more than one third, as it was deemed suitable mainly for big projects.

In table IV the two remaining patterns from the practical communication management practice category have been depicted.

TABLE IV.
COMMUNICATION MANAGEMENT PATTERNS WITHIN THE PRACTICAL CATEGORY

Basic communication principles	
Cx	This pattern applies to the basic principles, which should be observed by all project team members, and of which they should be aware in order to maintain transparency and positive nature of communication.
Pr	Misunderstandings, hostility or animosity among team members.
S	Reminding team members about the basic principles of transparent, effective and positive communication, and desired behavior, that is, among others: justifying requests, asking rather than telling, keeping promises and showing up for appointments (also virtual ones), writing positive emails (even criticisms and dissatisfaction can be expressed in a positive way); it is also a good practice to set the maximum time for response to an email, to ensure the dynamics of asynchronous communication.
Q	Positive on the following communication quality aspects: clearness and cohesion, meeting needs of communicating participants.
A	The pattern can be used for any kind of project and team, although it is especially useful for immature and inexperienced teams, or where there are many introverts, team members are age or culture diversified.
P	All project team members should communicate obeying the rules presented in this pattern, although for some of them, especially the experienced and mature ones, they may seem obvious.
Cq	Good atmosphere in the team, clear and positive relations among team members and their responsible behavior – all promoting successful project completion.
I	Usually the basic principles of transparent, effective and positive communication is something that every person knows and feels, and it should not be required to state it explicitly, but in the cases mentioned above it may be desired to bring them to the attention of some team members. If possible communication rules should be agreed upon together by the whole team, preferably during the kick-off meeting.
RP	-
Synchronous working environments	
Cx	In certain environments and usually big projects, it is desirable to have similar project setup and roles for better and easier synchronization of communication and work.
Pr	Cooperating teams in different locations greatly differ from each other both in terms of composition and way of working, making it difficult for communication and cooperation between them.
S	Provide a similar composition of the teams and work procedures in all locations in order to facilitate cooperation and communication.
Q	Positive on the following communication quality aspects: engaging the right people.
A	This pattern applies to big and long-term projects carried out by two teams in different locations – the client's and the contractor's team.
P	Team members playing similar roles in both teams.
Cq	Synchronized working environments on the client's and contractor's side, with defined roles and responsibilities; easier direct communication due to existence of counterparts.
I	Defining a process with roles definition, responsibilities, authorities and templates could be used to set up the synchronous working environments of the cooperating teams.
RP	-

V. Conclusion

The eleven project communication management patterns presented in this paper aim at supporting project managers and teams in effective implementation of communication management practices based on a context-problem-solution approach. The four groups they are arranged into address various aspects of project management and types of encountered problems.

There are at least two main future research directions that can enrich the body of knowledge on project communication patterns. The first is to assess implementation conditions, dependencies and effectiveness of the patterns specified in this paper. The second is to look for and identify more common project communication patterns. Both require conducting a more extensive survey among project-based companies.

Acknowledgment

Special thanks to all the professionals from the project management field who devoted their precious time to share their experiences, knowledge and opinions on the topic of communication management patterns in project teams in their companies.

References

[1] V. Damasiotis, P. Fitsilis, and J. F. O'Kane, "Measuring communication complexity in projects," in *Proceedings of the Management of International Business and Economic Systems (MIBES-ESDO) 2012 International Conference*. Larissa, Greece: School of Management and Economics, 2012, pp. 100-114.

[2] M. M. Phillips, *Reinventing communication: How to design, lead and manage high performing projects*. Farnham, UK: Gower Publishing, Ltd., 2014.

[3] R. Burke, and S. Barron, *Project management leadership: building creative teams*. Chichester, UK: John Wiley & Sons, 2014.

[4] M. Charles, "The Ascent of Communication: Are We on Board?," in *The Ascent of International Business Communication*, L. Louhiala-Salminen, A. Kankaanranta, Eds. Helsinki, Finland: Helsinki School of Economics, 2009, pp. 9-23.

[5] K. Muszyńska. "Kształtowanie modelu komunikacji w zespole projektowym," Ph.D. dissertation, University of Szczecin, 2010.

[6] W. Al-Ahmad, K. Al-Fagih, K. Khanfar, K. Alsamara, S. Abuleil, and H. Abu-Salem, "A Taxonomy of an IT Project Failure: Root Causes," *International Management Review*, vol. 5, no. 1, pp. 93-104, 2009.

[7] K. Conboy, "Project failure en masse: a study of loose budgetary control in ISD projects," *European Journal of Information Systems*, vol. 19, no. 3, pp. 273-287, 2010. DOI= http://dx.doi.org/10.1057/ejis.2010.7.

[8] R. Stoica, and P. Brouse, "IT project Failure: A proposed four-phased adaptive multi-method approach," *Procedia Computer Science*, vol. 16, pp. 728-736, 2013. DOI= http://dx.doi.org/10.1016/j.procs.2013.01.076.

[9] H. R. Kerzner, *Project management: a systems approach to planning, scheduling, and controlling*. Hoboken, NJ: John Wiley & Sons, 2013.

[10] K. Schwalbe, *Information technology project management*. Boston, MA: Cengage Learning, 2013.

[11] J. R. Meredith, and S. J. Mantel Jr., *Project management: a managerial approach*. Hoboken, NJ: John Wiley & Sons, 2011.

[12] G. Purna Sudhakar, "A model of critical success factors for software projects," *Journal of Enterprise Information Management*, vol. 25, no. 6, pp. 537-558, 2012. DOI= http://dx.doi.org/10.1108/17410391211272829.

[13] D. F. Ofori, "Project Management Practices and Critical Success Factors–A Developing Country Perspective," *International Journal of Business and Management*, vol. 8, no. 21, pp. 14-31, 2013. DOI= http://dx.doi.org/10.5539/ijbm.v8n21p14.

[14] V. Holzmann, and I. Panizel, "Communications management in Scrum projects," in *Proceedings of the European Conference on Information Management & Evaluation*. Reading, UK: Academic Conferences and Publishing International Limited, 2013, pp. 67-74.

[15] J. Han, and W. Jung, "How Geographic Distribution Affects Development Organizations: A Survey on Communication between Developers," *International Journal of Software Engineering & Its Applications*, vol. 8, no. 6, pp. 241-251, 2014.

[16] T. Niinimäki, A. Piri, C. Lassenius, and M. Paasivaara, "Reflecting the choice and usage of communication tools in global software development projects with media synchronicity theory," *Journal of Software: Evolution and Process*, vol. 24, no. 6, pp. 677-692, 2012. DOI= http://dx.doi.org/10.1002/smr.566.

[17] B. Sidawi, "Potential use of communications and project management systems in remote construction projects: the case of Saudi Electric Company," *Journal of Engineering, Project, and Production Management*, vol. 2, no. 1, pp. 14-22, 2012.

[18] K. Tone, M. Skitmore, and J. K. W. Wong, "An investigation of the impact of cross-cultural communication on the management of construction projects in Samoa," *Construction Management and Economics*, vol. 27, no. 4, pp. 343-361, 2009. DOI= http://dx.doi.org/10.1080/01446190902748713.

[19] P. Wagstrom, and J. Herbsleb, "Dependency forecasting in the distributed agile organization," *Communications of the ACM*, vol. 49, no. 10, pp. 55-56, 2006, DOI= http://dx.doi.org/10.1145/1164394.1164420.

[20] M. Monteiro de Carvalho, "An investigation of the role of communication in IT projects," *International Journal of Operations & Production Management*, vol. 34, no. 1, pp. 36-64, 2013. DOI= http://dx.doi.org/10.1108/IJOPM-11-2011-0439.

[21] J. Samáková, J. Sujanová, and K. Koltnerová, "Project communication management in industrial enterprises," in *European Conference on Information Management and Evaluation*, Reading, UK: Academic Conferences International Limited, 2013, pp. 155-163.

[22] K. E. Papke-Shields, C. Beise, and J. Quan, "Do project managers practice what they preach, and does it matter to project success?," *International Journal of Project Management*, vol. 28, no. 7, pp. 650-662, 2010. DOI= http://dx.doi.org/10.1016/j.ijproman.2009.11.002.

[23] A. M. Pop, I. Pop, and D. D. Dumitrascu, "An Analysis Model Of The Communication Features In Research Project Management," *Revista Economica*, vol. 65, no. 4, pp. 49-64, 2013.

[24] D. S. Nguyen, "Success Factors for Building and Managing High Performance Global Virtual Teams," *International Journal of Sciences: Basic and Applied Research*, vol. 9, no. 1, pp. 72-93, 2013.

[25] G. Isern, "Intercultural Project Management for IT: Issues and Challenges," *Journal of Intercultural Management*, vol. 7, no. 3, pp. 53-67, 2015. DOI= http://dx.doi.org/10.1515/joim-2015-0021.

[26] Project Management Institute, "The high cost of low performance: the essential role of communications," 2013, http://www.pmi.org/~/media/PDF/Business-Solutions/The-High-Cost-Low-Performance-The-Essential-Role-of-Communications.ashx, access date: 6.05.2016.

[27] V. E. Pitts, N. A. Wright, and L. C. Harkabus, "Communication in Virtual Teams: The Role of Emotional Intelligence," *Journal of Organizational Psychology*, vol. 12, no. 3/4, pp. 21-34, 2012.

[28] P. Ziek, and J. D. Anderson, "Communication, dialogue and project management," *International Journal of Managing Projects in Business*, vol. 8, no. 4, pp. 788-803, 2015. DOI= http://dx.doi.org/10.1108/IJMPB-04-2014-0034.

[29] Oxford Dictionaries. *Language matters*. http://www.oxforddictionaries.com/definition/english/pattern, access date: 6.05.2016.

[30] J. Rech, R. L. Feldmann, E. Ras, A. Jedlitschka, and B. Decker, „Knowledge Patterns and Knowledge Refactorings for Increasing the Quality of Knowledge," in *Knowledge Management, Organizational Memory and Transfer Behavior: Global Approaches and Advancements,* 1st ed., M. E. Jennex, Ed., London, UK: IGI Global,

2008, pp. 281-328. DOI= http://dx.doi.org/10.4018/978-1-60566-140-7.ch017.

[31] A. M. Ernst, "Enterprise architecture management patterns," in *Proceedings of the 15th Conference on Pattern Languages of Programs*, ACM, 2008. DOI= http://doi.acm.org/10.1145/1753196.1753205.

[32] S. Lukosch, and T. Schümmer, "Groupware development support with technology patterns," *International Journal of Human-Computer Studies*, vol. 64, no. 7, pp. 599-610, 2006. DOI= http://dx.doi.org/10.1016/j.ijhcs.2006.02.006.

[33] E. Gamma, R. Helm, R. Johnson, and J. Vlissides, *Design patterns: elements of reusable object-oriented software*. Upper Saddle River, NJ: Pearson Education, 1994.

[34] T. Iba, *Collaboration patterns: a pattern language for creative collaborations*. Japan: CreativeShift Lab, 2014.

[35] N. Schadewitz, „Design patterns for cross-cultural collaboration," *International Journal of Design*, vol. 3, no. 3, pp. 37-53, 2009.

[36] Y. Verginadis, N. Papageorgiou, D. Apostolou, and G. Mentzas, „A review of patterns in collaborative work," in *Proceedings of the 16th ACM international conference on Supporting Group Work*, New York, NY, ACM, 2010, pp. 283-292. DOI= http://dx.doi.org/10.1145/1880071.1880118.

[37] K. Muszyńska, "Communication management in project teams–practices and patterns," in *Managing Intellectual Capital and Innovation for Sustainable and Inclusive Society, Proceedings of the MakeLearn and TIIM Joint International Conference*, V. Dermol, A. Trunk, G. Đaković, M. Smrkolj Eds., ToKnowPress, 2015, pp. 1359-1366.

[38] S. Apud, and T. Apud-Martinez, "Effective Internal Communication in Global Organizations," 2008, http://www.iabc.com/ effective-internal-communication-in-global-organizations, access date: 6.05.2016.

[39] A. Bilczynska-Wojcik, "Communication management within virtual teams in global projects". Ph.D. dissertation, Dublin Business School, 2014.

[40] J. Douras, "Techniques to Build Respect and Trust with a Remote Workforce," 2010, http://www.iabc.com/techniques-to-build-respect-and-trust- with-a-remote-workforce, access date: 6.05.2016.

[41] L. Layman, L. Williams, D. Damian, and H. Bures, "Essential communication practices for Extreme Programming in a global software development team," *Information and Software Technology*, vo. 48, no. 9, pp. 781-794, 2006. DOI= http://dx.doi.org./10.1016/j.infsof.2006.01.004.

[42] F. Y. Y. Ling, S. P. Low, S. Q. Wang, and T. K. Egbelakin, "Foreign firms' strategic and project management practices in china," in *Proceedings of Construction Management and Economics: Past, Present and Future*. Reading, UK: University of Reading, 2007.

[43] S. Modi, P. Abbott, and S. Counsell, "Exploring communication challenges associated with Agile practices in a globally distributed environment," 2012, http://raiseconference.org/wp-content/uploads/2012/10/ModiCounsellRAISE-paper-resubmission-final.pdf, access date: 6.05.2016.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1189–1198
DOI: 10.15439/2016F99
ACSIS, Vol. 8. ISSN 2300-5963

Integrating Semantic Web Services into Financial Decision Support Process

Ilona Pawełoszek
Częstochowa University of Technology,
Faculty of Management
Poland
Email: ipaweloszek@zim.pcz.pl

***Abstract*—The operating environment of Small and Medium-sized Enterprises (SMEs) is more uncertain and risky comparing to big enterprises. The SMEs managers often face insufficient funds and lack of domain expertise to undertake important financial decisions. Moreover they are often unaware of opportunities offered by financial institutions. Therefore there is a growing need for intelligent and interactive web-based tools integrating data and providing information from heterogeneous sources, oriented on risk-detection and decision support for small and medium enterprises. In this paper the semantic Web services platform has been proposed, which integrates the internal data from the company's databases and external Web resources to perform dynamic evaluation of the financial position of enterprise. The designed platform detects signals indicating the need for action, and it composes a process choosing available Web services aiming at improving the company's financial situation.**

COMPETITIVE challenges on a global market require decision makers of small and medium enterprises (SMEs) to have up-to-date and relevant knowledge of the economic situation of the company and its environment. Managers must have the possibility to identify and analyze all indicators that may have impact on the operations of the enterprise and moreover they have to react to market changes and take the appropriate actions. In a world teeming with overwhelming amount of data accessible by the World Wide Web analyzing information becomes very difficult. Discovering and understanding all dependences between various financial ratios while taking into account economic trends is crucial because they alert one to anomalies and threats [1].

Unfortunately, many of the managers of the SMEs lack expertise or meaningful experience in financial domain as well as the potentials the Web services can bring to the business. Furthermore, SMEs operate in a definitely more uncertain and risky environment than big enterprises. A complex and dynamic market changes have much more impact on SMEs' financial situation than on big companies. Tolerance of mistakes is narrower [2], [3]. Usually financial expertise is either not available or too expensive for the managers of SMEs. Big companies, in contrast, have at their disposal strategic consultation and elaborated procedures to solve problems. The SMEs cannot achieve this level of expertise due to the lack of skilled personnel and financial resources. The decision makers often act intuitively so the rationality of their decisions may be questionable. Especially in exceptional and unusual cases intuition alone can be unreliable [4]. According to Dane and Pratt [5], the effectiveness of intuitive decision-making depends on the decision-maker's level of expertise on the subject at hand.

Therefore there is a growing need for intelligent and interactive Web-based tools integrating data and providing information from heterogeneous sources, oriented on risk-detection and decision support.

The diversity of solutions and opportunities in the field of integrating data from many sources have grown with the emergence of the semantic Web services which on the one hand provide intelligent search capabilities and on the other, they offer business process automation. The semantic Web shares many goals and issues with Decision Support Systems (DSS), e.g., being able to precisely interpret data, in order to deliver relevant, reliable and accurate information to a manager. Semantic Web technologies have been used in DSS during the past decade to solve a number of different tasks, such as information integration, sharing, Web service annotation, discovery, knowledge representation and reasoning [6].

There are a lot of publications discussing technical issues and challenges in this field. However a relatively small number of literature describes the subject from the managerial point of view, particularly: how to use Web services in an innovative way to improve the agility and efficiency of business processes and how this technology can be used to bring competitive advantage for the SMEs..

In this paper the semantic Web services platform has been proposed, which integrates the internal data from the company's databases and external Web resources to detect signals (factor changes) indicating the need for decision action, then it composes a process of available Web services aiming at improving the company's financial situation.

The work contributes to the domain of early warning systems and decision support systems.

The structure of the paper is as follows. Section 2 presents a brief state of the art of the domain of Web services and financial early warning systems. Section 3 explains the advantages of semantic Web services comparing to traditional approach. In the section 4 the components of the Web services-based platform for decision support are presented. In order

to illustrate the idea of using Web services in a business process an example of financial decision-support system is presented in the section 5. In the example, a part of early warning system and financial ontology will be integrated into semantically rich business processes.

I. Overview of the literature

Semantic Web services (SWS) have been extensively discussed in the literature in many contexts [7], [8]. The theoretical benefits of semantics as well as their potential impact on operational management are well known concepts. Semantic Web solutions integrated with the company's IT systems may support managers in composing the instances of business processes. However the automation is usually possible in streamlined processes with limited number of alternative flows.

In the research related to Web services, several platforms and languages have been presented that allow easy integration of heterogeneous systems. In particular Universal Description, Discovery, and Integration (UDDI) [9], Web Services Description Language (WSDL) [10], Simple Object Access Protocol (SOAP) [11] and REST [12], which define standard ways for service discovery, description and invocation. There are other initiatives such as Business Process Execution Language for Web Service (BPEL4WS) [13] and DAML-S Service Model [14] that are focused on service compositions where a flow of a process is known a priori.

There is a growing consensus that simple functional descriptions of Web services are not sufficient to develop intelligent and dynamic processes, characterized by the high degree of heterogeneity, autonomy, and distribution of service providers on the Web [15]. Due to the increasing availability of Web services that offer similar functionalities with different characteristics there is a need for more sophisticated discovery processes to match user requests [16].

The key issue to address is modeling of composition of Web Services, which should be integrated into real business processes. The focus is on the activities of discovery and selection that are required to identify the relevant Web services and to include them in a business process [17]. The idea is that a number of SWSs can meet some basic requirements specified by a manager. Then, the manager needs to be supported in choosing which one of the above SWSs better fulfills his needs. The selection and invocation of the best (according to a set of criteria) service can be automatic, or left to the discretion of a human requester [18].

Ontologies are increasingly used in describing complex data, processes and services on the Web. Here the concept of using semantics to support decision making has been proposed with particular emphasis on early warning models. There are a lot of methods and techniques of early warning and warning forecasting models that can be used by the managers of SMEs. The choice of the method is an open question because there is no synthetic indicator that would aptly describe the financial condition of the company. The prospecting works illustrate models of [19]: E.I. Altman, M. Tamari, R.J. Taffler, M. Blum, S. Appetiti, R. Edminster, E.B. Deakin, M. Zmijewski, W. H. Beaver.

Most of the aforementioned methods generate a warning signal informing mainly about threats inside the company, but also about unnoticed opportunities for further development of the company. Therefore it can be a single information or a set of information thanks to which one can predict future threats in the firm's development.

Currently organizations have a growing need for intelligent systems that can assist managers by gathering and analyzing information, making recommendations, supporting business decisions, and implementing business workflows [20]. Therefore interoperability and transparency of the technology are very important, so the managers could concentrate on the substantial aspects of the decision making process.

As the surveys conducted in different countries show [21], [22], [23], many managers of small businesses do not have grounded financial knowledge, many of them are self-taught, learn from an accountant, a consultant, or a bookkeeper. Today many important financial services companies use Web APIs to offer their customers, their staff, and their business partners new tools that streamline operations. Therefore it is possible to harness the exposed Web services to compose a process or a few variants of processes that fulfill the company's current needs. The proposition presented hereby is based on ontologies as the means of knowledge representation and interpretation of current company's financial situation. The possible actions that can be taken are suggested on the base of accessible financial Web Services. The services can be dynamically composed into the company's processes.

The next section presents the advantages of using semantic Web services in contrast to the traditional Web services approach.

II. From traditional to semantic web services

In traditional scenario Web services initially composed by the developer using Business Process Execution Language for Web Service (BPEL). Currently, the services are discovered by matching the service request parameters to the predefined keywords in service descriptions. This means that the process is known a-priori without any possibility to choose the best from some alternative process flows.

Semantic Web services are discovered by high level match-making techniques [24], whereas non-semantic Web services discovery methods use information retrieval techniques [25] based on keyword matching.

The traditional solution has serious limitations due to natural language descriptions, which may or may not be available and their analysis requires human intervention. The service publishers sometimes leave the service descriptions blank in the UDDI registries. In such a case even keyword matching based approach does not work.

To address the syntactical limitations of business processes and services the semantic enhancement of Web services descriptions with ontologies can be a solution. Semantic Web services enable a dynamic composition of Web services. In this case, Web services are composed at run-time: the participants of a composite Web service are discovered

dynamically based on a variety of concerns: availability, load-balancing, cost, quality of service (QoS), etc. [26 p.251].

According to a definition by Moreau et al. [27] Semantic Discovery is the process of discovering services capable of meaningful interactions, even though the languages or structures, of their descriptions may be different. Moreover, since ontologies and Web services are developed independently the service request and advertisement can be annotated with multiple ontologies, thus facilitating better efficiency [28].

In the Web services based DSS complex scenario, it is difficult to find specific Web services to meet the requirements of the decision makers. The situation becomes even more complicated when there is no single Web service which could satisfy all of the requirements but it can only be achieved with a combination of several Web services [29].

Many business applications are built on the base of different Web services available on the Internet. These applications are highly dependent on discovering relevant and efficient web service. The discovered Web service must match with the input, output, preconditions, effects and service quality parameters specified by the user. Modeling of composition of Web services is based on algorithms for comparing available services with identified business process requirements.

Web services developed by different vendors are usually published on the Internet using Universal Description, Discovery and Integration (UDDI) [30] which are XML-based registers of services. Search in UDDI is based on keyword matching which is not efficient as huge number of Web services may match a keyword and it is difficult to find the right one. Modern approaches take advantage of semantic Web concept where Web service matching is done using ontologies. Discovering Web services automatically without manual work is an important concern [31].

Web services can be developed using different protocols, and tools. The chosen standard implies the way the service can be discovered and invoked. There are several application programming interface (API) standards in use, of which the most current is Representational State Transfer (REST) [12]. REST is a network architecture paradigm relying on standard transport protocols like HTTP, without the use of an additional messaging layer. A service call is handled via its URI. "REST provides a set of architectural constraints that, when applied as a whole, emphasizes scalability of component interactions, generality of interfaces, independent deployment of components, and intermediary components to reduce interaction latency, enforce security, and encapsulate legacy systems" [32]. Services developed in accordance with the REST paradigm are often called RESTful services. The World Wide Web is the key example of the REST design.

REST has a lower barrier to entry versus other approaches such as SOAP, RPC or CORBA. As a programming approach, REST is a lightweight alternative to the aforementioned solutions in terms of complexity of coding. Another advantage is good performance and simplicity of HTTP used to make calls between machines. A single URL is enough to call a service and the HTTP reply is the raw result data — not embedded inside any code so it can be directly used without the need of parsing. Moreover the RESTful services can be invoked using a Web browser alone so it is easy to write and test applications.

However there is no standard way to publish and discover REST services (such as Web Services Description Language (WSDL) in SOAP approach where all the service descriptions are stored in a centralized UDDI registry). Many attempts were made in order to resolve this issue (the popular ones are WADL and RAML projects), but none of them has been commonly accepted by the community of REST developers. Therefore there is a challenge of adding semantics to the RESTful services [33] and it is still an active field of research and development.

Companies can benefit from exploiting semantic Web services in many aspects of their business. On the strategic level, the SWS give the possibility to build highly maintainable applications and profit from a loosely coupled architecture which facilitates the cooperation between the company and its business partners. On a tactical level, a company may reuse the ontologies that have been created for SWS in the other areas such as knowledge management and Business Intelligence. On the operational level the company can benefit from streamlined operations and data integration.

III. Components of the platform for supporting financial decision-making

The World Wide Web allows managers to access the information from the large database repositories globally. In general the information can be used to reduce the risk of financial decisions. The large amount of information (some examples are provided later in Table II) and diversity of structures of data sources make it difficult to find the right information, therefore semantic Web technologies are crucial to improve the efficiency of discovery, automation, integration and reuse of data. Semantic Web technologies also provide support for interoperability problem which cannot be resolved with Web technologies of the previous generation.

The overall framework of the semantic Web services-based decision support system can be schematically illustrated on Figure 1. The source of information for business processes are transactional systems such as ERP, data warehouses, BI and the environment of the company. The external information is imported into the system by the means of Web services, which are dynamically discovered on the Internet according to the criteria defined in the business process models. The information from the internal and external sources is combined and presented to the manager to support the decision making process and to suggest alternatives for the decision maker to consider.

However the broad scope of Web services and opportunities offered by Web financial platforms creates a possibility to automate the process of consulting and decision making. The schema of the process of decision making is presented on figure 2.

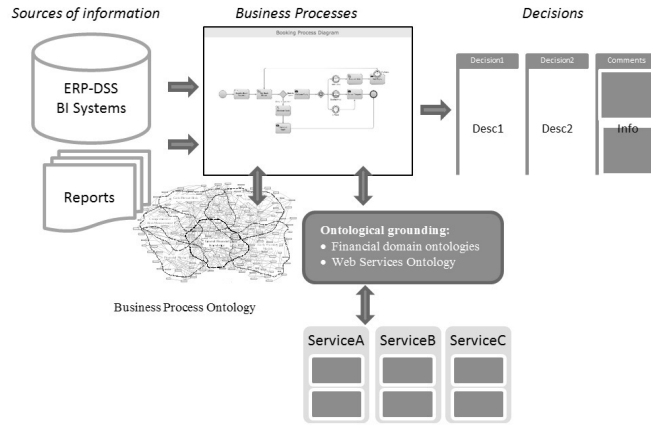


Fig. 1 Functional components of the platform

The process depicted above is written in human-readable BPMN syntax, but it can be translated into an executable BPEL process. That is not simple because of conceptual mismatch between the two process modeling languages which differ in their expressive power [34].

Some of the activities in the process can be executed automatically by discovering, selecting and invoking a proper Web service. This steps in the process diagram are marked with the sprocket icon (in BPMN) which means automation by calling the Web service (or set of Web services) that are able to perform an action predefined in the diagram.

The request for the financial data for the analysis can call the services which pull out the data from the enterprise systems and external Web services that provide data about market (for example stock prices of competitors, annual earnings, prices of products or other financial ratios). The data provided that way can be then used to further analyses and interpretations of KPIs (Key Performance Indicators), which is the next step in the process. Extracting the data from external resources offering RESTful services can be done in multiple ways. There are many existing conceptual models for service descriptions such as: SAWSDL, WSMO-Lite, hRESTS or Swagger.

In order to reduce the amount of manual effort required for finding, combining and using services, it is necessary to provide a common vocabulary. An interesting and well documented approach to provide common semantic model of Web services is the Minimal Service Model (MSM). Driven by the idea of Semantic Web and Linked Data MSM is written in RDFS. The Minimal Service Model defines services as having a number of operations, each of which have input and output messages and faults. Web APIs are also supported through the definitions of URI template, and the HTTP method [35 p.246].

Having the framework for semantic description of Web services the next thing is to build the database of instances referencing to the Web services that can be potentially useful in the financial Decision Support System.

In the case of traditional SOAP-based architectures it is quite easy to find the services because they are described and published in WSDL files which are indexed by popular search engines. The service discovery can be done for example by using Google with the search parameter ".... filetype:wsdl". Also the registers (UDDI) can be used (e.g. SeekDa.com or ServiceFinder) that crawl the Web for WSDL files. However as far there is no straightforward unified way to find RESTful Web services because it is hard to distinguish between URI of RESTful Web service and other Web resources (e.g. Web sites URLs).

The most accurate way to find RESTful services is to browse the web pages of their providers. The well-known global data providers that may offer financial services or economic data that can be used to support managerial decisions are for example: Quandl, Xignite, Bloomberg, online banks and other financial institutions .

The mentioned Minimal Service Model ontology can constitute a data structure for a triple store (RDF database such as Virtuoso) of instances. The advantage of this RDF-based solution is its flexibility and openness for adding new data structures describing new frameworks of services description. Moreover the triple stores offer exhaustive searching capabilities by the possibility to use semantic query languages (e.g. SPARQL).

The next issue is to dynamically embed the selected services into the business process definition.

As it was mentioned before there are a few points in the process flow where the services can be used: to gather data needed for the assessment of the company's financial situation and for suggesting possible alternatives (selected Web services that are relevant) for decision makers.

Because business process descriptions are constructed in their specific modelling languages such as BPMN or BPEL there is a need to do semantic mapping between the process descriptions and the services from the RDF database. For example the REST URL to get StockQuotes of 5 companies (Microsoft, Inter Corporation, SODR Gold Trust, Silver Wheaton Corp Common Shares and MarkWest Energy Partners) may be the following:

```
https://www.westwind.com/West-
WindWebToolkit/Samples/Rest/StockSer-
vice.ashx?Method=GetStockQuotes&symbol-
List=msft,intc,gld,slw,mwe
```

As it can be seen there are some terms in the URL that can be ambiguous for automatic recognition and mapping to the terms used in the business process model. For example in business process model an activity could be named: "Compare the stock prices of the competitors". This is compound activity that should invoke a few services to provide required data. The above REST request on stock quotes of 5 companies could be used. So there is a need to "know" that GetStockQuotes is one of the tasks in the considered activity.

Also the abbreviations of the companies' names can be ambiguous (there can be full names used in the process model) so there is a need to provide a dictionary / thesauri of the names and abbreviations. The components of the proposed solution are illustrated on Fig. 3.

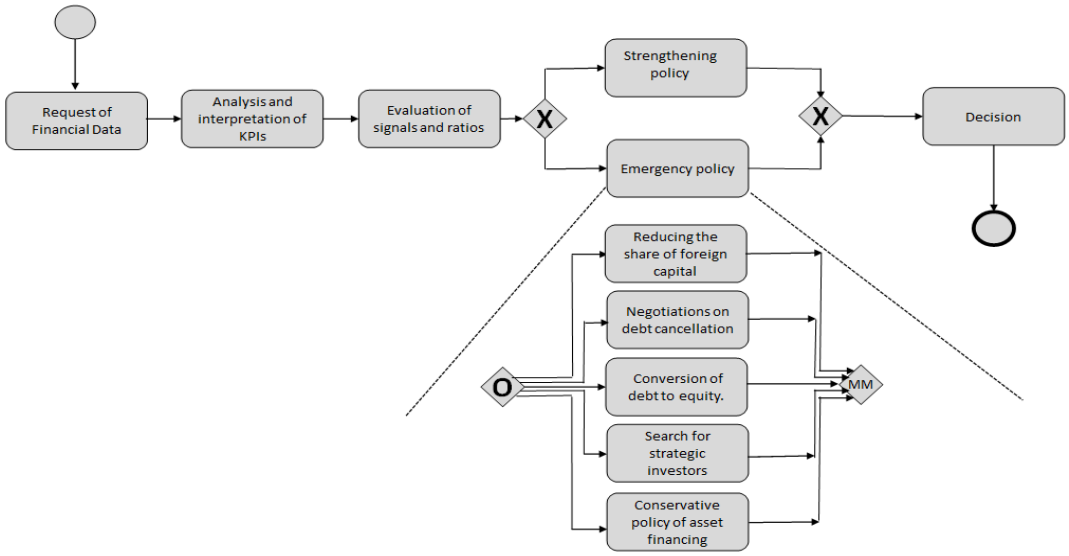


Fig. 2 The BPMN diagram of the process of financial decision-making

Irrespectively of the approach chosen (RESTful or SOAP-based Web services) the functioning of the platform can be described in the following steps:

1. The manager defines goal and requested QoS parameters
2. The system checks for available services (exposed by B2B partners, administration institutions, banks etc.). The ontology is used to identify the services that are relevant to the manager's request.
3. The SWS system extracts data about the user from IT Systems.

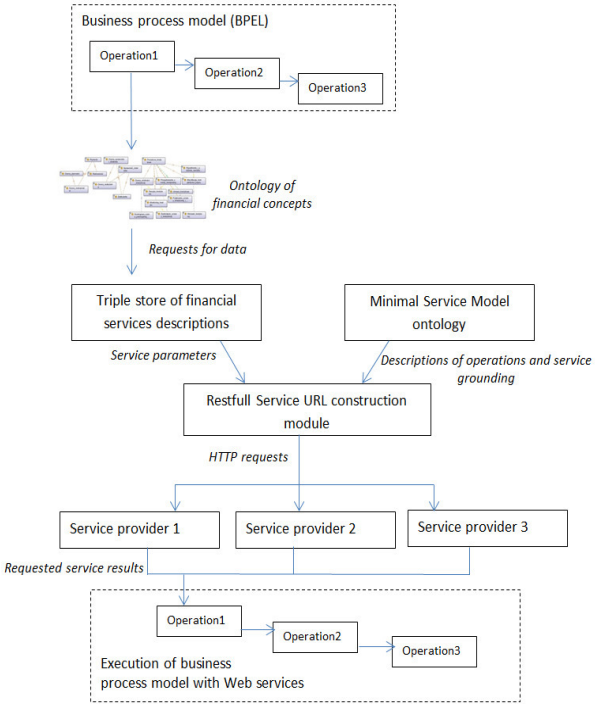


Fig. 3 The platform components with data flow

4. Reasoning capabilities are used to rate the services. In this step the user's profile and requirements are considered.
5. Highest rated services are chosen and are recommended for the manager.
6. The user receives offers ordered by their ratings.
7. The user can decide on the best way to perform a process. The manager can refer to the financial ontology at any time there is a need for additional explanations.
8. The corrections to the ontology can be made on the base of the manager's feedback.

The set of ontologies describing concepts, relations and functions related to the assessment of economic situation of the company was used. In the example described hereby, particularly the ontology of early warning signals pointing to financial difficulties were taken into account. The ontology facilitates the interpretation of warning signals by the manager. A part of the ontology has been presented on Fig. 2.

While conducting the analysis the manager has access to a domain ontology that includes concepts associated with models and methods of early warning and notions related with data from financial reports. These concepts can be linked with their instances – values in data bases, reports and documents to be used to determine the warning signals . The manager is supported by financial ontology while evaluating economic situation of the enterprise. The ontology helps to interpret the data coming from the ERP system, for example to determine and interpret ROA, ROE and many other ratios. The fragment of this ontology is illustrated on Fig.4.

The business processes depicted as a diagram in the center of Fig. 1 in the real scenario are defined in Business Process Execution Language - BPEL (also WS-BPEL, BPEL4WS) . BPEL provides the possibility to invoke Web services in a predefined sequence. Each business process defined in BPEL can be exposed as a Web service and included into other processes.

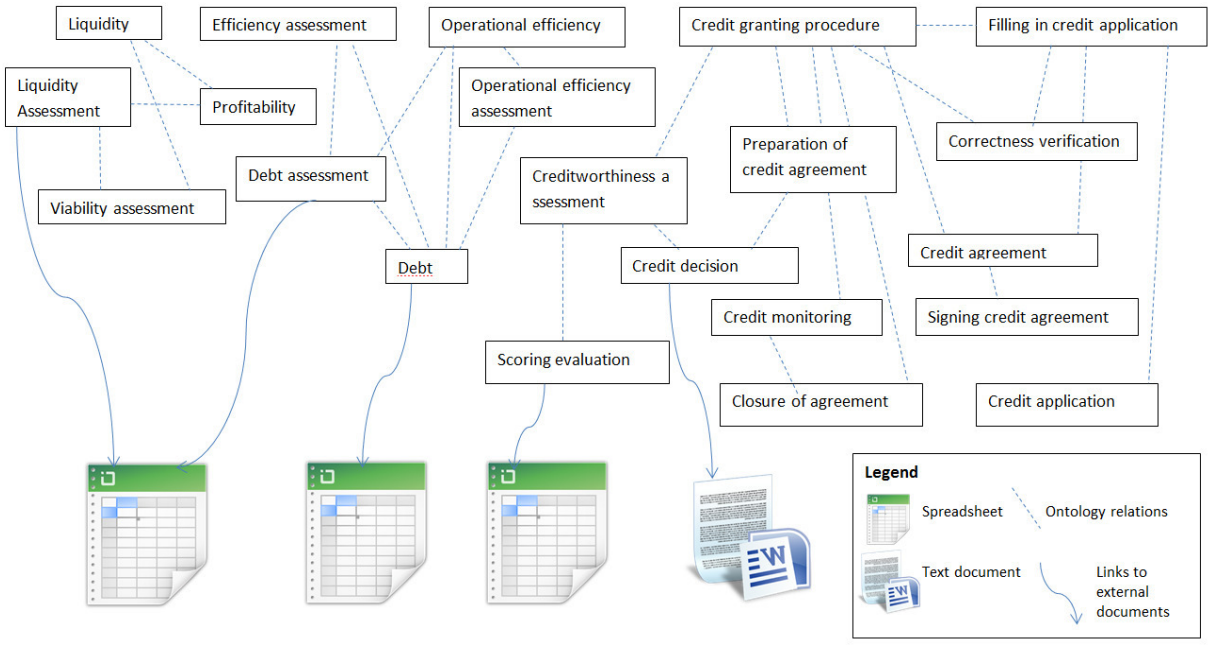


Fig. 4 Fragment of financial ontology with links to external documents

BPEL is using WSDL [36] descriptions to identify partner services, i.e. services are identified by port types and operations in the process models. As a result only services that implement a concrete interface can be used which is a major deficiency of BPEL [37].

Semantic Web Services (SWS) aim at eliminating this deficiency by automation of the development of Web service based applications through the use of ontologies. By providing formal descriptions with well-defined semantics, SWS facilitate the machine interpretation of Web Service – functional and not functional – properties. Significant results in the field of SWS are already available, in terms of reference ontologies, e.g. OWL-S [38] and WSMO [7]. The two ontologies represent different approaches to specify semantic information for Web services in order to enable automatic service discovery, composition and execution.

The OWL Profile ontology contains semantically enriched information about Web service and it is used as both an advertisement for the service and as a request to find a service on the basis of matching semantics. A first significant difference between the two approaches is that OWL-S does not separate what the user wants from what the service provides. In WSMO there is a property referred to as a goal, which specifies what the user wants and the Web service description defines what the service provides through its capability [39].

Common underlying idea behind all the semantic approaches is the dynamic selection, composition and mediation - on the basis of available semantic descriptions – of the most adequate Web resources (services and data) to accomplish the specified processes [40]. Current research efforts are investigating how SWS can be effectively applied in real world business scenarios and integrated with enterprise information systems.

It is easy to see the benefits of semantics in many application areas. For example in financial domain, a major area of financial business is the provision, selling, and management of mortgages. This is a compelling business: the provider (i.e. the bank) earns money by selling a mortgage contract, and the client is naturally interested in the cheapest offer that suites his needs. Usually, clients consider several offers and compare these in detail with respect to the monetary amount in question. For example, if several banks provide their mortgage offers via semantically described Web services, we can easily imagine an application that compares the offers and automatically selects the one which is most suitable for the customer's needs [7 p.168].

IV. Example Scenario of Financial Decision-Making

Semantic Web services can be used in many ways in an enterprise architecture, the most typical examples include supply chain management and business integration among multitude applications. One of the main advantages of SWSs is their reusability. The composition of new services by dynamically invoking the existing SWSs creates the possibility to perform very complex tasks in business processes. However, in this section the simple example of early warning system was chosen to illustrate the concept of semantic Web services integration with business processes.

The financial signals are the most popular and the broader group of quantifiable signals. They are generated on the basis of deviations from desired single values (or groups of values) of financial parameters or ratios. Selected warning signals designated on the basis of specific financial parameters are presented in Table I. It should be noted that these are example signals and their interpretation related to different areas of finances.

TABLE I.
EXAMPLE CRITERIA OF ASSESSMENT OF FINANCIAL CONDITION OF THE ENTERPRISE

Ratio	Evaluation criteria		
	Good	Average	Bad
Asset Turnover	>.40	.25 - .40	<.25
Current Ratio	>1.5	1.0 -1.5	<1.0
Debt to Asset	<.30	.30 - .50	>.60
Operating Expense	<.65	.65 - .80	>.80
Operating Profit Margin	>.15	.05 - .15	<.05
Return on Assets	>.10	.50 - .40	<.05

To illustrate, the case study concentrates on the values of ratios which indicate bad situation of the company. Moreover, it has to be stressed that during the analysis of the financial situation of the company many other ratios can be used. It is easier to determine potential problem areas when industry benchmarks are available, they can be provided by many Web portals (such as: finance.google.com, finance.yahoo.com, bizminer.com, bizstats.com).

The rank of generated warning signals can differ. Let us consider the case of using semantic services related to so called strong warning signal in the area of funding structure. This signal means high probability of financial difficulties resulting i.e. in bankruptcy.

To solve the problem the manager should take appropriate decisions associated with:

- reducing third party equity,
- negotiations on the cancellation of debt,
- the debt-to-equity swap,
- searching for strategic investors.
- conservative policy of financing current assets.

Currently financial advisory consultants offer their services to address the aforementioned problems. Their work is to find and suggest solutions and the best alternatives.

As it was mentioned before, Web services can be implemented also for taking actions as the response to the early warning signals and to put into practice the emergency policy. Ontology of financial knowledge is the foundation of the described platform. The comprehensive financial ontology takes on the role of the consultant – on the one hand it helps the manager to take decisions by improving the efficiency of analysis and increasing the capacity of understanding of financial data, and on the other hand it is needed for automatic discovery and selection of proper Web services which provide the managers with current data and financial services.

The activities included in the emergency policy are defined in the financial ontology. For example the warning signals may suggest negotiations on debt cancellation or reduction. The activities aiming at this goal may include looking for debt consolidation loans, so the system may find Web services of banks that offer debt consolidation loans and send requests to the banks' Web APIs. The requests would include parameters such as the amount of the company's debts and the deadlines of payment. Then the system can choose the best offers of banks, present them to the manager. After the manager's decision the system sends needed information to the selected bank to make finance agreement.

It is worth noticing that the aforementioned suggested actions would have strategic implications if implemented, so the decision to a high degree depends on how the given company views and solves important issues. There are many behavioral issues (such as risk aversion of the managers) that influence the choice of strategic actions. These issues cannot be resolved automatically, but the role of the proposed system is to present alternative ways of action.

The operations management problems are more relevant for automatic (or semi-automatic) handling, so the usability of the proposed system can be illustrated. The example of such problem can be liquidity analysis. The term refers to an enterprise's state of financial health. Liquidity is the ability of an organization to meet its short-term financial obligations. This issue is a major source of concern for small and medium business managers as bank loans are becoming too expensive to maintain.

Liquidity management encompasses the management of cash balances, including short-term funding and investments for excess cash. The alternatives for raising cash during temporary shortages and the opportunities to invest excess cash on a short-term basis are important to the functioning of an organization [41].

In the early-warning system the data for calculating liquidity ratios (ie. current ratio) are taken from financial statements such as: balance sheet, income statement and the statement of cash flow.

Early warning systems can spot liquidity risks by looking at both historical and projected financials. A low liquidity ratio could signal the company is suffering financial troubles.

However, a very high liquidity ratio isn't good either; it may indicate that the company is too focused on liquidity to the detriment of efficiently utilizing capital to grow and expand its business. According to the calculated values of the liquidity ratios the appropriate actions can be suggested by the system. A company can improve its liquidity ratios by raising the value of its current assets, reducing current liabil-

TABLE II.
EXAMPLES OF WEBSITES OFFERING RESTFUL WEB SERVICES IN FINANCIAL DOMAIN

URL	Description
http://www.bloomberg.com/company/	The established service provides free, unrestricted access to raw data for customers for its financial market information. The same publish/subscribe and request/response interactions available via its proprietary interface can be accessed via API. This functionality gives access to data on current market trades, either real-time or delayed, along with reference data on reference data, historical information, and records of intraday trading.
http://www.gaincapital.com/liquidity-api-trading.shtml	Capital offers a full range of CFD markets including indices, commodities, FX, bonds, interest rates and equities
https://angel.co/	AngelList is a community of startups and investors with the goal of making fund raising efficient. Angel investors are listed along with contact information which startups can use to set up introductions. The AngelList API provides developers with a RESTful interface to the AngelList data set. Data includes followers, reviews, startups and more. Responses are formatted in JSON and JSONP.
https://api.mattermark.com/	The Mattermark REST API allows to query expansive dataset of companies. It gives access to all data available by website interface, as well as the time series data about companies. It offers access to companies' profiles, investors and funding events.
https://www.lendingclub.com/developers/api-overview.action	Offers REST API for services such as: personal and business loans and investing.

ities by paying off debt, or negotiating delayed payments to creditors. There are many solutions for resolving liquidity problem that can be suggested by the system and appropriate Web services can be found.

The procedure of the system's functioning contains the following steps:

a) an overall assessment of the firm's financial condition,

b) referring to the financial early-warning ontology to find possible corrective and preventive actions,

c) browsing available and relevant Web services to find sources of financing (such as: credits, new investors, business angels etc.),

d) analysis of risk and costs of the alternatives,

e) presenting the simulations to the manager,

f) successive, iterative searching of the alternatives of the highest efficiency.

Over the last few years, RESTful Web services have become popular due to their simplicity in publishing and consuming Web resources and better alignment with modern Web applications. The examples of sites offering Web services in financial domain are presented in Table II.

The proposed solution for supporting financial decision making by exploiting Web services works in the Web environment and can also be exposed as a compound sem-antic Web service (i.e. in the cloud). The main challenge in building the platform is to find and provide the relevant Web services from many providers which could be useful in financial decision making. It is also very important to maintain and update the base whenever there are important changes brought by the service providers considering the service parameters and when new services and providers appear on the market.

The second challenge is semantic description of the Web services which on one hand should be matched with the concepts from financial ontology and on the other hand their technical details, input and output and parameters should be identified and described semantically.

It is worth to notice that there are attempts to unify access to Web Services particularly facing the problem of the lack of the registers of RESTful services. The solutions (for example iServe [42] are not domain-oriented and do not offer ontological support for matching the services with business process models. There is no Web platform that would offer a full range of specific services for supporting managers in taking financial decisions. There are many applications, which let managers to monitor financial KPIs and visualize them in the form of manager's cockpit. The detailed review of the solutions was presented in [43]. Although many functions were offered, none of them was integrated with external web services to support managers in taking actions by suggesting specific services offered by banks and other financial institutions.

V. CONCLUSION

Web services are currently a preferred way to architect and provide complex services, there are many technical considerations in the literature regarding this field and many competing standards arise. The usability and efficiency of the proposed platform to a large degree depends on available Web services exposed by financial institutions therefore there is a challenge of integrating services from many providers, where the semantic Web technology can help.

There are several limitations while using financial measures, which can be addressed by semantic Web services. Financial ratios highlight the situation of the enterprise but their interpretation is needed to give answers to problems. Financial ontologies can be valuable source of expertise for managers in this field. The measures of the economic condition of the company are as accurate as the data used to calculate them. Semantic Web services offered by trusted companies can provide data of good quality that can be used for industry-specific benchmarks.

At the moment, the development of the described platform is on its first, conceptual stage. Although, it should be noted that some ontologies have already been elaborated and applied in the area of early warning and decision support system InKoM [44]. Ontologies build with one purpose in mind, such as early warning system, can be reused in other areas because the knowledge codified this way is independent of technical solutions.

The main challenge and future work direction in the described approach is the segmental breakdown of financial policies and procedures into granular steps that can be performed by relevant Web services. Current work is also directed towards finding the most efficient solution for the Web services discovery and specification.

ACKNOWLEDGMENT

First of all I would like to express my gratitude to Jerzy Korczak, Wroclaw University of Economics, Poland, for providing access to the financial ontologies developed within the project InKoM and his valuable comments to the draft of the paper. I also thank Bogdan Franczyk, University of Leipzig and Maciej Pondel, Intratic Software Solutions for Business, for comments and suggestions on the technical details of RESTful Web services. I thank Jolanta Chluska, Czestochowa University of Technology, and Piotr Oleksyk Wroclaw University of Economics for consultation on the financial issues.

REFERENCES

[1] C. Olszak, "Wybrane technologie informatyczne w doskonaleniu rozwoju systemów Business Intelligence," in *Zastosowania systemów informatycznych zarządzania, Problemy Zarządzania*, special issue, W. Chmielarz, J. Kisielnicki, T. Parys and O. Szumski, Eds. Wydawnictwo Naukowe Wydziału Zarządzania Uniwersytetu Warszawskiego, 2011, pp. 85-96.

[2] J. Korczak and H. Dudycz, "Intelligent dashboard for SME managers. Architecture and functions," in *Proceedings of the Federated Conference on Computer Science and Information Systems* FedCSIS 2012, pp. 1003–1007.

[3] P. Gibcus, P.A.M. Vermeulen and J.P.J Jong, "Strategic decision making in small firms: a taxonomy of small business owners," in *International Journal of Entrepreneurship and Small Business*, vol. 7, no. 1, 2009, pp. 74-91.

[4] J. A. Howard, "Financial decision-making: the roles of intuition, heuristics and impulses," in *Journal of Modern Accounting and Auditing*, Vol. 9, no. 12, 2013, pp. 1596-1610.

[5] E. Dane, K. V.V. Rockmann and M.G. Pratt, "When dhould I trust my gut? Linking domain expertise to intuitive decision-making Effectiveness," in *Organizational Behavior and Human Decision Processes,* 119, no. 2, 2012, pp. 187-94.

[6] E. Blomqvist, "The use of semantic Web technologies for decision support - a survey," in *Semantic Web Journal*, 5(3), IOS Press 2014, pp. 177-201, http://www.semantic-web-journal.net/sites/default/files/swj299_0.pdf

[7] D. Fensel, H. Lausen, A. Polleres, J. De Bruijn, M. Stollberg, D. Roman and J. Domingue, Enabling Semantic Web Services: Web Service Modeling Ontology, Springer 2006.

[8] M. Ouzzani and A. Bouguettaya, Semantic Web Services for Web Databases, Springer Science+Business Media, LLC, 2011, DOI 10.1007/978-1-4614-1644-97

[9] T. Bellwood et al., "Universal Description, Discovery and Integration specification (UDDI) 3.0". 2002, http://uddi.org/pubs/uddi-v3.00-published-20020719.htm.

[10] R. Chinnici, et al., "Web Services Description Language (WSDL) 1.2", 2007, http://www.w3.org/TR/wsdl/.

[11] D. Box et al. "Simple Object Access Protocol (SOAP) 1.1". 2001, http://www.w3.org/TR/SOAP/.

[12] R.T. Fielding, Architectural Styles and the Design of Network-Based Software Architectures, PhD dissertation, Department of Information and Computer Science, University of California, Irvine, 2000, https://www.ics.uci.edu/~fielding/pubs/dissertation/rest_arch_style.htm

[13] Oasis, "Web Services Business Process Execution Language Version 2.0," 2007, http://docs.oasis-open.org/wsbpel/2.0/OS/wsbpel-v2.0-OS.pdf

[14] DAML, "DAML-S: Semantic Markup for Web Services," 2003, http://www.daml.org/services/daml-s/0.9/daml-s.html

[15] J. Cardoso and A.P. Sheth, "Introduction to semantic Web services and Web process composition". In Proc. of First Intl Workshop on Semantic Web Services and Web Process Composition (SWSWPC'04), San Diego, CA, USA 2004.

[16] C. Petrie. T. Margaria, H. Lausen and M. Zaremba eds., Semantic Web Services Challenge, Results from the First Year, Springer Science-Business Media, LLC, 2009.

[17] A. Karray, R. Teyeb and M. Ben Jemaa, "A heuristic approach for Web-service discovery and selection," in International Journal of Computer Science & Information Technology, 5(2) 2013.

[18] A. Carenini, et al., "Semantic Web service discovery and delection: a test bed scenario," in *Sixth International Workshop on Evaluation of Ontology-Based Tools and the Semantic Web Service Challenge (EON-SWSCQ8), CEUR Workshop Proceedings,* vol. 359, R García-Castro, A. Gómez-Pérez, C.J. Petrie, E. Delia Valle, U. Ktister, M. Zaremba and O. Shafiq eds., RWTH Aachen, Aachen, 2008, http://sunsitc.informatik.rwth-aachcn.de/Publications/ CEUR-WS/Vol-359/

[19] R. Siedlecki, "Prognozowanie trudności finansowych przedsiębiorstw z wykorzystaniem miary rozwoju Hellwiga," in *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu* nr 323, 2013, pp. 308-318.

[20] J. Collins, W, Ketter and M. Gini, "Flexible decision support in dynamic inter-organisational networks," in *European Journal of Information Systems* 19(4), 2010 pp. 436-448

[21] Sage, "Sage Canadian Small Business Financial Literacy Survey" Canadian-Small-Business-Financial-Literacy-Survey 2012, http://www.sage.com/na/~/media/site/sagena/documents/surveys/Sage-Canadian-Small-Business-Financial-Literacy-Survey

[22] F. K. Andoh and J. Nunoo, "Sustaining Small and Medium Enterprises through Financial Service Utilization: Does Financial Literacy Matter?", University of Cape Coast, Ghana 2012, http://www.uclan.ac.uk/research/explore/groups/assets/igfd_sustaining_small_and_medium_enterprises_through_financial_service_utilization.pdf

[23] N. Plakalović, "Financial literacy of SMEs managers", Make Learn, International Conference 2011, http://www.toknowpress.net/ISBN/978-961-6914-13-0/papers/ML15-086.pdf

[24] S. Srirama, M. Jarke and W. Prinz, Mobile host: "A feasibility analysis of mobile Web," in *The 4th International Workshop on Ubiquitous Mobile Information and Collaboration Systems*, 2006, pp. 942–953.

[25] J. Fuller, M. Krishnan, K.Swenson and J. Ricker, "Oasis asynchronous service access protocol (asap)," May 18 2005, http://www.oasisopen.org/committees/documents.php?wg_abbrev=asap

[26] J. Bergstra and M. Burgess, eds., *The Handbook of Network and System Administration*, Elsevier, 2007.

[27] L. Moreau, S. Miles, J. Papay, K. Decker and T. Payne, "Publishing Semantic Descriptions of Services", Proceedings of Global Grid Forum 9, Chicago IL, USA, 2003.

[28] S. Sowmya Kamath, V. S. Ananthanarayana, "Semantic Web Services - Discovery, Selection and Composition Techniques". Third International Conference on Computer Science & Information Technology, 2013 doi: 10.5121/csit.2013.3616

[29] L. Guo, Y.H. Chen-Burger, and D. Robertson, "Mapping a business process model to a semantic Web dervice model," in *Proceedings of the 2nd IEEE International Conference on Web Services*, 2004, pp. 746-749.

[30] Oasis, "UDDI Spec Technical Committee Draft", 2004 http://www.uddi.org/pubs/uddi_v3.htm#_Toc85907967

[31] D. Mukhopadhyay, A. Chougule, "A survey on Web service discovery approaches," in *Advances in Computer Science, Engineering & Applications*, D.C. Wyld, J. Zizka and D. Nagamalai, eds., AISC, vol. 166, Springer, Heidelberg 2012, pp. 1001-1012

[32] R. Douence R. et al., "Compositional Evolution of Secure Services using Aspects", 2010 http://cessa.gforge.inria.fr/lib/exe/fetch.php?media=publications:d1-1.pdf

[33] Y. Lee, "Semantic matching and resource discovery algorithms for RESTful Web services," in International Journal of Innovative Research in Computer and Communication Engineering Vol. 1, Issue 9, November 2013, http://www.ijircce.com/upload/2013/november/0A_Semantic.pdf

[34] J. Recker, J. Mendling, "On the translation between BPMN and BPEL: conceptual mismatch between process modeling languages," in CAiSE 2006 Workshop Proceedings - Eleventh International Workshop on EMMS AD. Luxembourg. June 5-6, pp. 521-532, 2006

[35] S. Dustdar and F. Li, eds., *Service Engineering: European Research Results*, 2011 Springer-Verlag/Wien, 2011.
[36] E. Christensen, F. Curbera, G. Meredith and S. Weerawarana, "Web Services Description Language (WSDL) 1.1" 2001, https://www.w3.org/TR/wsdl
[37] J. Nitzsche, T. van Lessen, D. Karastoyanova and F. Leymann, "BPEL for semantic Web services (BPEL4SWS)," in *Proceedings of the 3rd International Workshop on Agents and Web Services in Distributed Environments AWeSome'07* -- On the Move to Meaningful Internet Systems: OTM 2007 Workshops' , Springer-Verlag, 2007, pp. 179-188.
[38] D. Martin, et al, "OWL-S: Semantic Markup for Web Services W3C" Member Submission 22 November 2004, http://www.w3.org/Submission/OWL-S/
[39] J. De Bruijn, et al., "Relationship of WSMO to other relevant technologies" W3C Member Submission 3 June 2005 http://www.w3.org/Submission/WSMO-related/#owl-s
[40] S. Galizia, A. Gugliotta, C. Pedrinaci and J. Domingue, "Applying semantic web services," in *4th Workshop on Semantic Web Applications and Perspectives (SWAP 2007)*, Bali, Italy 2007 http://ceur-ws.org/Vol-314/48.pdf
[41] K.A. Horcher, *Essentials of Managing Treasury*, New Jersey, John Wiley & Sons 2005.
[42] Datahub, "iServe: Linked Services Registry" 2014 http://datahub.io/dataset/iserve
[43] P. Ziuziański and M. Furmankiewicz M., "Kokpit menedżerski jako narzędzie do wizualizacji danych w kontekście zarządzania wiedzą w organizacji," in *Zeszyty Naukowe Politechniki Białostockiej. Ekonomia i Zarządzanie*, z. 7(1), 2015, pp. 44-60.
[44] J. Korczak, H. Dudycz, M. Dyczkowski, "Inteligentny kokpit menedżerski jako innowacyjny system wspomagający zarządzanie w MŚP," in *Informatyka Ekonomiczna*, 1(31), 2014, pp. 288-303.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1199–1204
DOI: 10.15439/2016F542
ACSIS, Vol. 8. ISSN 2300-5963

Business Process Optimization with Big Data Analytics Under Consideration of Privacy

Silva Robak
Uniwersytet Zielonogórski, ul. prof. Z. Szafrana 4a, 65-516 Zielona Góra, Poland
Email: s.robak@wmie.uz.zgora.pl

Bogdan Franczyk
Uniwersytet Ekonomiczny we Wrocławiu, ul. Komandorska 118/120, 53-345 Wrocław,
Universität Leipzig, Germany
Email: franczyk@wifa.uni-leipzig.de,
bogdan.franczyk@ue.wroc.pl

Marcin Robak
Universität Leipzig, Germany
Email: robak@wifa.uni-leipzig.de

***Abstract*—One of the contemporary problems, and at the same time a big opportunity, in business networks of supply chains are the issues associated with the vast amounts of data arising there. The data may be utilized by the decision support systems in supply chains; nevertheless, often there are information privacy problems. The supply chains in cloud will need appropriate administration for support of privacy aspects of cooperating business units existing in big data ecosystems. In this paper we analyze the possibility of utilizing the big data technology for supporting business processes optimization with respect of the privacy regulations in supply chains under the usage of the big data analytics lifecycle. We present our approach on an example of a business process in logistics.**

I. Introduction

THE emergence of Big Data is creating significant new opportunities for business to achieve added value and competitive advantage. Nevertheless the huge volume of data, the complexity of new data types and structures and the speed of new data creation cause problems in utilizing it in established business solutions like SCM Supply Chain Management [1] to attain the competitive advantages. From the business point of view using Big Data in the viable ways in the logistics requires gradual convergence between the Big Data Analytics Lifecycles process stages on one side and established ways of business process modeling in logistics on the other. What's more the emergence of data deluge and diversity of data types cause problems to be solved in the context of business process modeling. Additionally the security and privacy of data in logistics have to be treated in a way appropriate for business stakeholders.

In the paper we will approach the problem of utilization of vast amounts of data in supply chains with respect of the data privacy issues. We will investigate the big data ecosystem, and the process of big data analytics. We consider the big data analytics process from a supply chain perspective with emphasis on showing the differences between the traditional approach and the approach by using big data. The proposal of the modeling of data as an integration solution for business process management in supply chains networks we have already presented in [2], and [3]. We also have considered the research problems associated with big data utilization in logistics and supply chains design and management in [4]. In this paper we mainly review possible changes in the business processes which result from the above stated requirements and further implications based on big data analytics lifecycle, regarding privacy aspects.

The supply chains define the network that comprehends all the organizations and activities associated with the flow and transformation of goods from the raw material stage, through to the end user, as well as the associated information flow [5]. In the paper we will concentrate on the networked supply chain activities and flow of information.

In the inter-organizational information systems, which link the companies to their suppliers, distributors and customers, a movement of information through electronic links takes place across organizational boundaries, between separately owned organizations. It requires not only the electronic linkage in form of basic electronic data interchange systems (as for purchase orders), but also interactions in complex cash and information management systems or by access to shared technical databases. So the problems with the privacy are very persistent in supply chains contexts.

A business process consists of one or more than one related activities that together respond to a business requirement for an action [5]. The processing steps in a workflow may undertake numerous transformations of data (geographic, technological, linguistic, syntactical and semantic transformations), communication is an important part of the process and (e-) business processes exist within certain environments. In the dynamic business environment, such as networks of venture participants involved in value chains in logistics, where data is forwarded among different enterprises, the appropriate arrangements of privacy aspects are essential.

Therefore, as stated previously, in our paper we will analyze how the big data analytics results can influence the business processes and their privacy aspects. For this aim the rest of the paper is organized as follows.

In Section 2 we characterize the main features of big data, the big data ecosystem elements. In Section 3 we summarize the big data analytics lifecycle process phases, key stakeholders and artifacts. In Section 4 we consider privacy regarded from the perspective of enterprises collaborating in logistics. We also give an example of a business process using big data analytics results in near-real-time in a supply chain. In the example we show how it can support privacy

978-83-60810-90-3/$25.00

from the perspective of enterprises. In the last Section we conclude our work.

II. Big Data

The big scale usage of available and generated data is made possible for organizations owing to cloud computing paradigms, such as Infrastructure as a Service (IaaS), Storage as a Service, which revolutionized the way the computing infrastructures are used [6]. As big data is referred to data that goes beyond the processing capacity of the conventional data base systems. In addition to this aspect that it is big, e.g. a huge number of small transactions, or (continuous) data streams from sensors, mobile devices etc., it may move too fast, or do not fit the structure of traditional (i.e., relational) database architectures.

According to [7] when we denote a big amount of data as "big data" it has to cover the three "Vs" (features) such as: volume, velocity and variety. Another authors (e.g. [8], [9]) add another V-features like value or veracity.

The first feature - volume of big data - denotes its massive character. The big volume of data is beneficial for the data analysts. It may improve the analytics models by having more cases available for forecasts and increase the number of factors to be considered in the models making them more accurate. Nevertheless, the volume feature bears potential challenge for the IT infrastructures to deal with big amounts of data, especially taking into account its second feature – velocity.

The second feature of big data is the velocity in which data flows into organization or the expected response time to the data. Big data may arrive quickly - in real-time, or near real-time. If data arrives too quickly the IT infrastructures of the organization may be not able to respond timely to it, or even to store all of it. Such situations may lead to data inconsistencies [10].

The third feature of big data is the variety of data. Big data may have diverse structures and forms, not falling into the rigid relational structures of SQL databases without loss of information. Some of data may be saved as blobs in inside traditional data bases. Therefore the IT infrastructures for big data are denoted as NoSQL, what means the data is "not only SQL" [9]. Several examples for diverse kinds of data are standard business documents, transactional records, and unstructured data in form of images, recordings, HTML documents (web pages), text messages and email messages, streams from meters and environments sensors, GPS tracks, click streams from Web queries, social media updates, data streams from machines' communication or wearable computing sensors, and many others.

The data deluge possibly useful for enterprises, especially involved in SCM is driven nowadays by many factors like data created by the traditional IT devices, mobile devices, Internet and social network users, GPS systems and sensor nets.

The flood of data comes into existence with the images and videos uploaded in the www, video surveillance in enterprises and cities, medical information recordings, mobile devices of the users (phone calls, text messages, mobile application usage, online games), smart devices (TV sets and receivers, smart private and industry buildings, electric grids), traditional devices (computers, game boxes, video games, e-books, etc.) and non traditional IT devices such as GPS navigation systems, earth data processing devices, radio-frequency identification (RFID) readers, ATMs, credit card readers, sensor nets and Internet of Things.

To each Gigabyte of the collected information the additional Petabyte meta-information is added. All this create a landscape of the emerging big data ecosystem.

The main player in the system are [11] the data devices, data collectors, data aggregators and data users, buyers. The data devices are the above stated originators of data deluge.

The main data collectors include information broker in Internet, the government using analytics services, medical institutions and their appliances, employers.

The marketers and private investigators can act as data aggregators by using the obtained data and transforming and packaging it for diverse stakeholders interested in conducting campaigns for users with high likelihood willing to get or buy a specific service.

The users and buyers of data are financial institutions, retail, phone and TV providers, media archives and so on. They also purchase data from third parties which enables more targeted marketing campaigns.

The collected data can be used for forecasting of the user behavior and suggesting and recommending services for user intended willing to pay for it.

Broadly speaking, the main types of data structures for big data are structured data (the most minor part), semi-structured data (slightly bigger ratio of big data), quasi-structured data (a big part of data) and unstructured data (building the majority of big data). The structured data type are such known from traditional databases and warehouse applications processing (OLAP, RDBMS, spreadsheets). The semi-structured data enables its relative easy parsing (e.g., XML-data files), while the quasi-structured data can be formatted only by using appropriate tools with much effort and possible inconsistencies occurring. On the other hand, the unstructured data does not possess an inherent structure.

III. Big Data Analytics Lifecycle Process

Conducting big data science ventures differs from approaches for projects using SQL data and Business Intelligence BI methods and tools applied for data analysis aims [3]. The approach used for big data is more explorative in its nature as well as there are additional key stakeholders involved in process stages.

The process can be easier handled by dividing it phases with clear defined milestones, involved stakeholder's responsibilities and artifacts. It is according to widely accepted divide and conquer principle in computer science. It does not strictly mean a waterfall process with no returns to the previous phases but the process is a guideline for the development process which is iterative and incremental in its nature.

The process for big data projects can be divided in phases such as [11]:

- Data discovery,
- Data preparation,
- Model planning,
- Model execution,
- Communicating results
- Operationalizing the results.

As mentioned above, several of the phases can be performed simultaneously as a project work. The iterative process in the subsequent phases can proceed forwards or sometimes backward, dependent on the yes/no decision at the milestones, or a deeper insight of the problem not enough realized in the earlier phases. It may include not understanding of the problem domain, stakeholder requirements or insufficient data available to solve the given project problems. First we regard the process stakeholders, then we give a description of the process stages with regard to the stakeholder roles and their important activities and deliverables within the phases.

In each phase there as workflows to fulfill the project aims in each phase and artifacts to be delivered by the project stakeholders.

The key stakeholders roles involved in big data projects [4] and [11] are the:

- Project sponsor,
- Business user,
- Project manager,
- Business intelligence analyst,
- Database administrator,
- Data engineer
- Data scientist.

The project sponsor sets priorities and metrics for the projects and establish the desired project outcomes. The business users as the experts in the business domain can provide guidance to the project requirements recovery and operationalizing the results of the big data analytics and also are those user who directly benefits from the results of the big data project. The project manager ensures the proper project scheduling according to key project's objectives and milestones. The business intelligence analyst is responsible for creating dashboards and reports and have knowledge in proper data feeds and sources. The database administrator is responsible for provision and configuration of the database environment in order to support the analytics needs of the project; he enables the access to the needed databases and data sets and ensures the desired security levels of the data repositories.

The first five roles are good known from the usual software projects, however the roles of data engineer and data scientist are new. The role of the data engineer is needed in context of the usage of analytic sandbox workspace and fulfills the needs of data preparation for data manipulation (extracts, transformation and loading). The data scientist provides expertise needed for accurate application of analytical techniques and enables the right choices for given business problems. We already explained in detail the skills and background of such roles as data engineer and data scientist in our paper [4].

The six stage process for big data analytics project starts with the first, the discovery phase. In this stage where the project team became acquainted with the business domain and accesses the resources needed for conducting the project. The team should familiarize itself with the project data. The accomplishment manner of this phase is dependent of the team experience in the past in similar projects. The main deliverables in this phase include framing of the business problem as an analytics challenge and formulating the initial hypotheses for data analytics.

In the second, the data preparation phase, the team prepares the analytics sandbox with the data required and conducts the ELTL operations (Extract, Load, Transform, Load) on the data in the sandbox workspace [11]. In this phase the team further familiarize with the available data and if required have to decide how to obtain the required, but at the moment not available data. It this phase the usage of technologies like Hadoop [12] may be needed.

In phase three, the model planning phase, the project team needs to decide the usage of the methods, techniques and workflows to be followed in the subsequent phases, especially in the next one. The dependencies between the variables have to be established, the key variables have to chosen and the most appropriate models types have to be selected.

In the next, the model building phase, the models recommend in the previous phase have to be developed and executed together with the appropriate data sets [4]. If the IT project environment will appear not to be sufficient for the project aims, the needed adaptations and hardware platform changes (parallelization or change to faster hardware) have to be fulfilled in this phase.

The fifth phase, which assignment is to communicate results consists of summarizing the project results such as key finding, quantification of business value and communicating the between the project stakeholders. In this phase the project aims success or failure will be decided according to the criteria determined in the first phase.

In the last operationalize phase, the final deliverables of the conducted project will be released in form of presentations, final reports, briefings, code, technical descriptions and documents etc. The form of the documents will be dependent on the recipient stakeholder type. The business value of the project should be conveyed to the key stakeholders. Moreover, after successfully running all the project phases the pilot project implementing the models in the production environment might be launched.

IV. Privacy From The Perspective Of Enterprises

From the perspective of enterprises data protection in business processes can be seen from two different perspectives. The first one is considering the security and privacy of sensitive business data which belongs to the enterprise or its supply chain [13]. It could be enterprises important operational data or the processed data of the customers, which is aggregated while doing own business or won in other man-

ner like big data analytics. From this perspective, where it is enterprises who can fall prey to data leaks, it is vital for them to protect important information which is needed for successful operation or maintaining of a competitive advantage. This situation was recently addressed by several solution frameworks or software platforms which take care for the data exchange and access management in supply chains. We refer to the examples of PREsTiGE platform or Aniketos [14], which organize and manage data access rights in business processes. However, they only treat explicit data and do not provide strategies for management of big data and its analysis.

Other point of view the protection of the privacy of individuals [15], i.e. current or potential clients, employees and other actors whose data is aggregated by enterprises in the course of their business processes. From this perspective – where enterprises are seen as potential benefiter of extrinsic information - the important points are developing and staying compliant with the privacy policy of the enterprise and above all staying in accordance with law in order to make safe both the individuals as well as enterprises.

Governments and several organizations developed regulations, guidelines and proposals for dealing with personal data. These are legal laws, or frame conditions for professional work with personal data.

The European Union's Data Protective Directive is most advanced and restrictive among world's data privacy laws. It contains several points which must be followed by entrepreneurs. In the USA, the Federal Trade Commission proposed a three step framework, consisting of the demands of privacy by design, simplified choice and greater transparency. Other recent development of American administration is Privacy Bill of Rights which includes demands individual control, transparency, respect for context, security, access and accuracy, focused collection, and accountability.

In the literature there are several privacy requirements concepts defined which are proposed for the processes which deal with personally-identifiable information (PII) and are component of the legal regulations. Shown below important hallmarks for dealing with PII in enterprises identified in the literature [16], shall be regarded in the big data applications.

Authorization verifies who has the right to access the data or can use specific activities. In the field of business processes and especially big data it is not an easy task to draw a clear line about such requirements like separation and binding of duties since the data and possible operations on them can be so different in their nature that it is not possible to model every feasible use option. Nevertheless binding of duty can be used for setting additional responsibilities for the data or activity user of the business process. In the same way some actions can be excluded which may help maintaining privacy in the big data field, i.e. by lowering the chances of de-anonymization. Authentication is accountable for the verification if the current user is one of the user authorized for the data or service use. Confidentiality postulates to keep the architecture, process and the whole environment in such state that the data stays protected from all non-authorized actors. Audit-ability is incorporated to that the process can be reviewed for keeping the privacy rights. Data integrity demands that the integrity of the original data has to be preserved after a processing failure. The data has to remain consistent, accurate and correct.

Since the demands regarding PII coming from the law regulations and internal privacy policies are very high therefore the procedures to fulfill them are correspondingly complex. One of the first questions which arise for enterprises is when the aggregated data has to be treated as personally identifiable. This is particularly complicated when big data comes in picture since with the ongoing technology development amounts of various data grow even more rapidly and cannot be thoroughly inspected as fast as they are collected. It demands elaborately designed analysis to assess if the conclusions drawn from the data will fall into the PII category. The technology advances enable easier identifying of specific person with growing amounts of data which only seem to have no correlation. In the end one can never be sure if apparently minor extension of the collected data won't enable the identification of individuals, and possession of such PII could harm the legal laws or cause negative effect on public opinion and therefore on the enterprises reputation.

Further step of this thinking is how the data can be processed that it will not fall in the PII category anymore.

In the next Section we propose the integration of big data analysis in business processes so that privacy will be regarded in everyday work of enterprises.

V. Business Process Modeling Practices For Supply Chains Architectures In Cloud

As stated in the above Section, privacy is an important part to regard in the business processes. While doing business the enterprises must adhere to legal law regulations and must comply with standards set in their own or in branch policies. It is often not apparent and not easily recognizable within standard business process if and when the stored data and usage of has to adhere to the above mentioned standards. It demands special steps within process which can help noticing potentially susceptible usage and unwilled consequences.

When thinking about the steps the first one would be to identify when the process and its data moves into the domain of PII. This could be done thanks to big data analysis, which can be seen a part of business process. Such analysis could be triggered by several stakeholders or events. Secondly if the data is identified as one which falls under the legislative of privacy, then actions must be undertaken in order to desensitize the contents and make it usable for business. Data after such processing must be replaced with the previous data.

It also must be analyzed why and how the data was aggregated to the level which does not comply with the standards and how this can be avoided in the future. It must be noted that the causes can be multifold and shall not be explicitly seen as enterprises misdemeanor, since these are often external circumstances which take the data out of balance. Such

circumstances are sourcing of data with previously not known contents, lack of overview on the available data, lack of thorough analysis of the new data in regard on relations with the already available data, changes in the permissions for the owned data, changes in the regulations and laws. As soon as the causes are known process changes must be made in order to avoid reoccurrence of similar situation in the future.

As an example let consider a web shop which plans to establish new service: delivery within one hour in a big city. The service itself will incorporate live tracking of traffic and weather situation, which will be used for delivery feasibility in the one hour time frame. To offer the service a warehouse will be established in the area which covers the service. The warehouse will be rather small, since in a big city area the costs are high. Because of the warehouse capacity restrictions, the products offered for sale will be studiously chosen. The options for opening other warehouse for better service or its expansion will be analyzed too. The goods shall bring good profit per warehouse area, so they must have high turnover or have a good profit margin. New opportunities for the goods will be continuously searched, i.e. the shop will also analyze their standard delivery shop and look for articles which are often ordered together. They will also review the preferences of the area population in order to offer better article range. The goods have to fit to the transport means (lorry, car, motorbike, cycle, etc.) which will be provided - at different cost - by external providers.

As described above the process uses big data and big data analytics at least two layers. First one is associated with the data related to transport used for the ordering process and assessing delivery time and service feasibility. Second is for optimization of product assortment at the warehouse or assessment if additional warehouse(s) would increase the profit.

First category extensively uses real-time data, needed for prompt delivery. This incorporates real time analysis of traffic data at the area of covered by the service, as well as observing current weather situation. Also the fleet of couriers is tracked at real-time, with detailed information about delivery duration. This feedback helps to choose right means of transport, e.g. a motorbike or a car. This data contains personal identifiable information – about the driver performance - as well.

Second level of big data analysis uses the data related to customers and predicts their buying preferences in order to adjust selection of the products available at the warehouse; it could show that new products or their varieties shall be offered, or that nearby area would be ideal candidate for expansion with additional transport means (i.e., e-bike) or with another warehouse. The data used for this analysis is won in multiple ways. It can be data feed in from social networks, where the engine is looking for activities of persons living in the area of warehouse delivery radius. It can look for the groups people living there belong, seek for their hobbies, music, films, sports, lifestyle, books, health interests. The machines can track and deliver the data of their clickstreams, friend lists, follows, likes and tweets. Also text mining of news sites and feeds, online newspapers, blogs, and public chats can be conducted in order to detect new social trends and needs. Although collecting such data may provide high-grade information about the needs of the inhabitants of the area, the risks for harming the privacy rights is clearly recognizable even for a layman. At the same time even for professionals it is very hard to interpret if privacy rights are harmed with ongoing data aggregation.

This shows that in both cases - transport and order data analysis, and population/trend analysis – it is not possible to assess in the real time, which information can be aggregated and stored, and when the thin border between preserving and breaking privacy is crossed. There is a need for deeper privacy rights compliance analysis for the aggregated data, which could be compared to the actual business analysis looking for business process optimization. In ideal case such privacy analysis should be done before the business analysis is conducted.

VI. Conclusion

Within the emerging big data ecosystem there are new groups of players, and also key roles for stakeholders. Big data analytics lifecycle is more exploratory in the nature as conventional processes, and requires a new approach. The process includes six main stages as: data discovery, data preparation, model planning, model execution, communicating results and operationalizing the results. The process itself is iterative and recurring while few phases can be carried out simultaneously as a project work.

The aim of this paper is to present the impact of big data analytics on business process modeling practices for supply chains architectures in which the modeling of privacy and security aspects play a significant role for the businesses, as they have to hold on to privacy laws like the European Data Protection Directive. The global players, as well as smaller businesses using big data, must be thoughtful about their data aggregation and analytics practices, in order to hold to those regulations. Some examples are capabilities of big data analytics which may interfere with privacy rights, like re-identification of (sensitive) data owners, profiling, amassing granular information about a person, and other uses, which go beyond the purpose and use restrictions, or the requirement of data minimization, data security, etc. Other open questions and discussions are the use of derived information based on personal data, as well as empowerment of consumers to manage their data. Supply chains are using nowadays huge amounts of data available in batch-time, online, as well as other diverse data.

As stated in [4] the data analytics involves descriptive analytics, predictive analytics and prescriptive analytics. Business Intelligence methods and tools using structured data, manageable data sets and traditional data sources can be utilized for diverse queries and providing answers for common questions of what happened in the past and why it did. Going beyond structured data requires usage of data science methods for conducting predictive analytics and data mining techniques for optimization, predictive modeling and forecasting. It will not only foster resolving reporting questions,

but also forecasting of what and why will happen. Moreover it may also support the operationalization of the key outputs of the analytic process.

Considering the data analytics process for big data described in Section 2, the sharing of the results within organization is conducted in the phase 6 (operationalization of analysis results), which is aimed at effective passing the outcomes of the analysis to the stakeholders, who are responsible to address them with appropriate actions and changes in the existing business process, so that the proposed changes and customizations can be efficiently integrated. The results of the analysis in form of describing and reporting change instructions and proposals for the business intelligence analysts will have a technical character and be in form of technical graphs like density plots, histograms etc.

If the data analysts shall detect that privacy rights may be harmed, then the suspect data sets shall be removed from the analysis. One must realize that removing parts of data will lead to results with lesser granularity but this is a price which must be paid to stay on the safe side and compliant with the privacy rights.

In the future we will investigate how big data analytics of privacy aspects could influence the established business process modeling methods, models and tools (e.g. BPMN [17]).

REFERENCES

[1] H. Baumgarten, *Das beste der Logistik*. Springer Verlag, Berlin 2008, http://dx.doi.org/10.1007/978-3-540-78405-0.

[2] S. Robak, B. Franczyk, and M. Robak, *Applying Linked Data concepts in BPM*, FedCIS 2012, IT4L. IEEE Conference Publications pp. 1105-1110.

[3] S. Robak, B. Franczyk, and M. Robak, *Applying Linked Data concepts in Supply Chains Management*, FedCIS 2013 IEEE Conference Publications pp. 1215-1221.

[4] S. Robak, B. Franczyk, and M. Robak, Research Problems Associated with Big Data Utilization in Logistics and Supply Chains Design and Management. FedCSIS 2014. Warsaw: Polish Information Processing Society, 2014. Annals of Computer Science and Information Systems, Vol. 3, pp. 92-93, http://dx.doi.org/10.15439/2014F472..

[5] M. P. Papazoglou, and P. M. A Ribbes, *E-business: organizational and technical foundations, John Wiley and sons*. London 2006, pp.88-90.

[6] D. Agrawal, S. Das and A. E. Abbadi, *Big data and cloud computing: current state and future opportunities*. EDBT 2011, March 22-24, 2011, Uppsala, Sweden. ACM 978-1-4503-0528-0/11/0003, http://dx.doi.org/10.1145/1951365.1951432.

[7] E. Dumbill, *What is big data? An introduction to the big data landscape*, Strata O'Reilly, 11 January 2012, http://strata.oreilly.com/2012/01/what-is-big-data.html

[8] S. Wrobel, *Big Data – Vorsprung durch Wissen*, Fraunhofer-Institut für Intelligente Analyse- und Informationsverarbeitungssysteme IAIS. Presentation, www.iais.fraunhofer.de

[9] I. Mitchell and M. Wilson, *Linked Data. Connecting and exploiting big data*, Fujitsu Services Limited, March 2012, www.fujitsu.com.uk.

[10] N. Marz and J. Warren, *Big data. Principles and best practices of scalable realtime data systems*. Manning Publications, MEAP Edition, Manning Early Access Program Big Data version 7, 2012.

[11] Data Science and Big Data Analytics. Discovering, Analysing, Visualizing and Presenting Data. EMC Education Services – Ed. Wiley 2014.

[12] The Apache Hadoop Project. http://hadoop.apache.org/core/ , 2009.

[13] Schwarzbach, B., Glockner, M., Pirogov, A., Rohling, M. M., & Franczyk, B., *Secure service interaction for collaborative business processes in the inter-cloud*. In Computer Science and Information Systems (FedCSIS), 2015 Federated Conference pp.1377-1386), http://dx.doi.org/10.15439/2015F282.

[14] Brucker, A. D., Malmignati, F., Merabti, M., Shi, Q., & Zhou, B. *A framework for secure service composition*. In Social Computing (SocialCom), 2013 Int. Conference pp. 647-652, http://doi.ieeecomputersociety.org/10.1109/SocialCom.2013.97.

[15] Pearson, S., *Taking account of privacy when designing cloud computing services* In Proceedings of the 2009 ICSE Workshop on Software Engineering Challenges of Cloud Computing pp. 44-52, http://doi.ieeecomputersociety.org/10.1109/CLOUD.2009.5071532.

[16] Mülle, J., Von Stackelberg, S., & Böhm, K. *Modelling and transforming security constraints in privacy-aware business processes*. In Service-Oriented Computing and Applications (SOCA), 2011 IEEE International Conference pp. 1-4, http://dx.doi.org/10.1109/SOCA.2011.6166257.

[17] OMG Consortium, *Documents Associated with Business Process Model and Notation (BPMN) Version 2.0*, Release date: January 2011, ww.omg.org/spec/BPMN/2.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1205–1213
DOI: 10.15439/2016F400
ACSIS, Vol. 8. ISSN 2300-5963

User specific privacy policies for collaborative BPaaS on the example of logistics

Björn Schwarzbach*, Michael Glöckner*, Arkadius Schier†, Marcin Robak* and Bogdan Franczyk‡
*Leipzig University,
Grimmaische Strasse 12, 04109, Leipzig, Germany
Email: {schwarzbach, gloeckner, robak}@wifa.uni-leipzig.de
†Fraunhofer Institute for Material Flow and Logistics IML
Joseph-von-Fraunhofer-Str. 2-4, 44227, Dortmund, Germany
Email: arkadius.schier@iml.fraunhofer.de
‡Wroclaw University of Economics
Kommandorska 118/120, 53-345, Wroclaw, Poland
Email: bogdan.franczyk@ue.wroc.pl

***Abstract*—Today's business is more and more organized in collaborative networks. Although decision makers know the benefits of collaboration, they are afraid of losing control of their data, which is one of the main impediments for Cloud Computing. We propose a novel cloud based approach for collaboration in business processes with guaranteed control of the privacy of the data. The platform ensures the compliance with the companies' privacy policies and laws. The paper shows the definition of privacy policies and how they are converted into a well established access control language. An example helps to clarify the methods.**

I. Introduction

ALMOST ten years ago, in 2008, Thomas J. Bittman, vice president of Gartner Research, published his view on future Cloud Computing development. Back in the beginning of Cloud Computing, cloud services were build on proprietary architectures of few dominant cloud service providers, e.g. Google, Amazon and Microsoft. The main problem in these times was the missing interoperability and compatibility of the cloud services of different cloud service providers. During the second phase the vertical supply chain distinguished itself as first ecosystems of smaller cloud companies emerged within the Cloud Computing market. New cloud service providers use the proprietary Cloud platforms of the dominant providers of the first phase to provide their own services. During the last phase these smaller cloud service providers unite to form horizontal federations. This union increased their earnings by expanding their capacities while reducing the costs at the same time through more efficient resource allocation. In parallel, open interoperability standards of service communication in intercloud-environment have been developed[1].

In the past years more and more companies adopted Cloud Computing by integrating cloud services into their supply chain. Especially in Germany the use of Cloud Computing in companies has increased from 2011 to 2014 by 16 percent, almost every second company is consuming at least one cloud service[2]. These cloud services reach from Infrastructure as a Service to Software as a Service. Especially the Software as a Service provides an tremendous number of services for every kind of task that can be achieved by software. Unfortunately these services are often provided by smaller cloud service providers while the cooperation between them, i.e. Bittman's third phase, is not well established. Every service has it's own interfaces with different message formats, even two services that provide the same functionality can differ in message format, behaviour, and constraints. According to an interview among approx. 120 german small and medium sized companies the target group of these services, i.e. these companies, does not have the knowledge how to tackle this problem.

In [3] we have proposed an architecture of a platform that enables those companies to consume cloud services of different clouds. The platform offers the features of a business process management system by orchestrating the individual cloud services in business processes that have been modelled by the consumers, i.e. the companies. Hence, we call this approach and the service provided by the platform Business Process as a Service.

In the interview mentioned before we discovered that most of the companies who do not consume cloud services are reluctant because of a fear of losing control of their data, which is even more important because of the hacks of global players (e.g. Sony) in the past months. But also those companies who consume at least one cloud service are concerned because of privacy issues. [4] comes to the same conclusion.

So one of the main challenges for such a platform is preserving the privacy of the data and the compliance with privacy laws while the business process is executed. This becomes even more important when multiple companies are involved in one business process and need to share data to each other. In [5] we have proposed an approach for secure service interaction, which has shown it's feasibility in multiple tests. The architecture proposed in [3] also provides a component

The work presented in this paper was funded by the German Federal Ministry of Education and Research under the projects PREsTiGE (BMBF 16KIS0082K) and LSEM (BMBF 03IPT504X).

for adding privacy policies to the business processes and the individual activities which are evaluated and enforced by the platform while the business processes are executed.

This paper discusses a new and flexible approach to define privacy policies for data that is transmitted while a business process is executed.

The remainder of this paper is structured as follows. After a brief presentation of the platform's architecture and the relevant components, the concepts behind the privacy policies for collaborative business processes is shown. The next section shows the translation of these policies into machine readable and evaluateable form. The algorithm for the translation is applied to an example based on a logistics use case. The paper is completed by a conclusion, which also reveals open tasks and questions.

II. Architecture of the platform

This section gives a very brief overview on the architecture of the platform for privacy preserving collaborative business processes.

The platform first presented in [3] is shown in Fig. 1. The components are represented as rectangles, their interfaces are shown as the lines between the components.

The most important component of the platform is the business process management system, which is located in the center. The business processes are modeled by the user with the configurator, which also enables the user to define process and activity related privacy policies and asign them to the appropriate objects. Privacy policies that are not related to one particular process or activity are defined in the privacy management component, which also stores the privacy policies defined in the configuator. The privacy management passes all known privacy policies to the identity and access management system (IAMS).

When the BPMS executes a business process and reaches an activity that needs some data as input to call the cloud service related to this activity, the BPMS queries the IAMS whether the service is allowed to access this data in the current context. If the IAMS grants access to the data (with potential obligations) the BPMS instantiates a gateway that takes care of a secure service interaction as described in [3]. The gateway also takes care of the obligations for data access.

The components of the platform provide RESTful web services to communicate. The communication is secured by SSL and client certificates. User data, except for the companies' core data, is held within the BPMS, there is no interface which offers the data to other components or external users other than through gateways, which ensures that no unauthorized entity can access data.

III. Privacy policies for collaborative business processes

This section describes in detail our approach for defining privacy policies in the context of collaborative business processes. One of our main requirements for the approach was to provide the companies with a tool that they could understand.

To define privacy policies that can be evaluated automatically and be used to decide whether a service is allowed to access some data or not we rely on use access control approaches. Basically there are four different types of access control. The mandatory access control and discretonary access control where applied in computer systems in the 70s of the last century. While mandatory access control describes security from the system itself by policies like "access is only granted from localhost", discretionary access control assigns each identity the appropriate access rights.[6] Mandatory access control is still used nowadays, e.g. SElinux is applying this approach [7].

In the late 80s and early 90s more and more users where using computer systems, hence assiging each individual user, i.e. identity, the correct access rights was not feasible any more. So in the beginning of the 90s role based access control emerged [8]. Role based access control assigns roles to identities and access rights are assigned to roles. This approach is used in Linux and Windows file systems and almost every modern software. Roles can be organized hierarically as shown in Fig. 2. [9], [10], [11]

Because of the well established application of role based access control our first approach for defining privacy policies was to apply role based access control.

During multiple workshops with local companies we discovered that the companies do not think about privacy identically. One common thing is that all companies separated the actors who want to access data into groups. But while some companies had have a very easy and strict approach for group setup, others could not clearly tell us which companies are member of which group. Instead they used phrases like "The driver of the truck while he is in the destination city is allowed to get the recipient's phone number to call the recipient to tell him his arrival time". This simple phrase contains the following information:

- The basic role of the person requesting access to the recipient's phone number is *driver*.
- The person requesting access to the recipient's phone number has to be located in the city where the shipment has to be delivered.
- Even if the first two conditions are met, the driver is only allowed to get the phone number if he want's to use it to call the recipient to announce his arrival.

This simple policy cannot be represented easily with roles because the location of the driver is changing over time. To tackle such requirements the research community followed two core approaches: extend role based access control with additional features, e.g. context or attributes, and creating a new access control model. [10]

Following the role based access control [12] has developed an access control model, which extends role based access control for virtual organizations. Unfortunately this model does not cover business processes, workflows, and cloud computing. Other approaches in this direction do cover business processes but leave out the cooperational aspect. Ref. [13], [14], [15] proposed and evaluated an extended role based access control

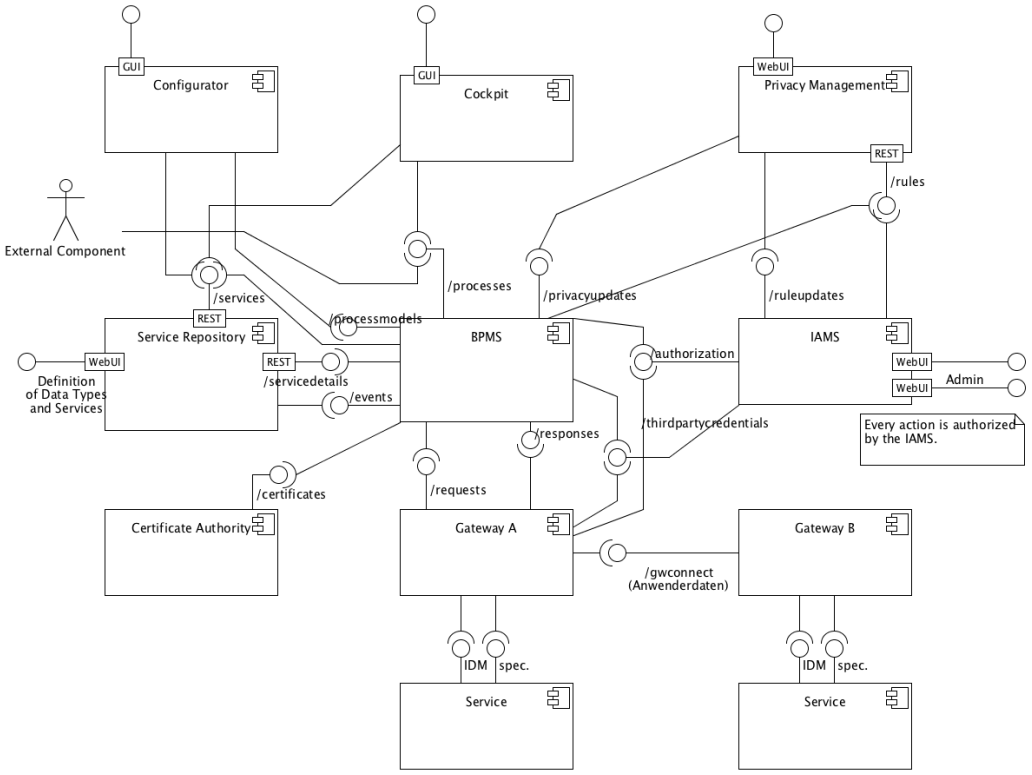


Fig. 1. Architecture of the platform for secure and privacy preserving collaborative business process as a service

model for team collaboration and workflows in the health sector.

All of the proposed models do not provide the flexibility in policy definition language that was needed by the participants of our workshops.

To achieve a maximum flexibility the research community developed a novel approach, the attribute based access control. In attribut access control policies are based on attributes of subjects and objects. According to [16] attribute based access control is:

"An access control method where subject requests to perform operations on objects are granted or denied based on assigned attributes of the subject, assigned attributes of the object, environment conditions, and a set of policies that are specified in terms of those attributes and conditions."[16]

The entity requesting access is called a subject. Typical attributes of subjects are their id, e.g. username, company name, and name of the department. The data that subjects want to access is called object or resources. A policy in attribute based access control is a triple of a subject, a resource, an operation, where the operation describes what access type the subject wants to have, and a result, e.g. grant or deny access. A policy can also comprise one or more conditions.

The policy "The driver of the truck while he is in the destination city is allowed to get the recipient's phone number to call the recipient to tell him his arrival time." consists of:

Subject The driver of the truck who is located in the destination city.
Resource Recipient's phone number
Operation Read
Condition Current activity in the workflow is "call recipient for dispatch notification"
Result Permit

The remainder of this section presents our approach on applying attribute based access control for privacy preservation to collaborative business process as a service. Privacy of data is always specified by the owner of the data, i.e. the entity who has created it.

First of all, in our platform business processes consist of activities that call external cloud based web services. Hence, there are two very basic roles in our platform. A process designer is an entity that models the business process, that is responsible for the correctness of the process itself, and that offers the resulting business process as a service to its customers. The second role is the service provider. A service provider is an entity that provides the external services that are beeing orchestrated in the business process by the process designer.

Our approach enables both roles to define their privacy policies independently from each other. It also includes privacy policies defined by law. Hence, the combined privacy policy consists of three columns as shown in Fig. 3 that can be

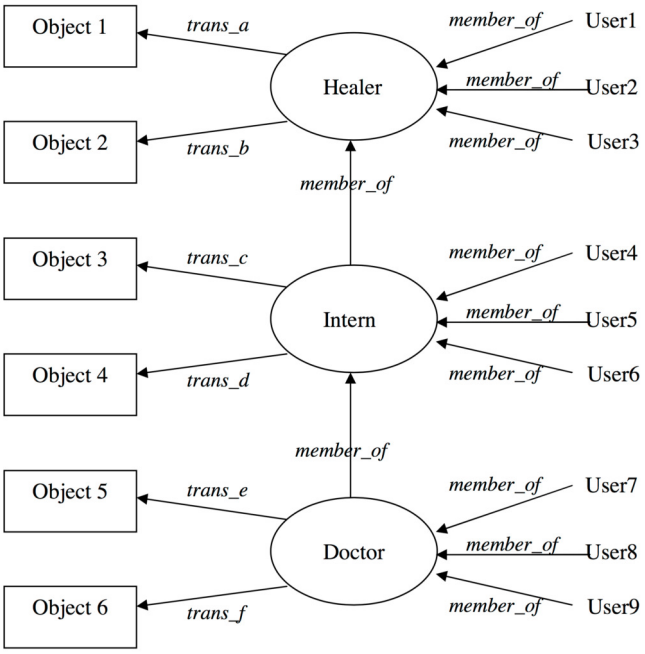


Fig. 2. Role based access control [11]

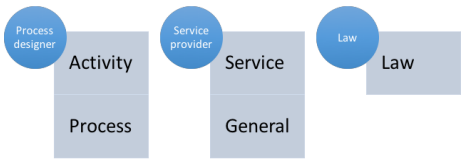


Fig. 3. Three columns of privacy policy process designer, service provider, and law

evaluated independently. The combined privacy policy results in permit if all three columns result in permit, else it results in deny.

To simplify the process of policy definition and to reduce redundancy in policies we provide each role with two levels of policies.

First a process designer can specify general privacy policies that are valid for the whole business process. E.g. a process designer may restrict access to all data to companies that are located in the Europe Union to ensure no data is transfered to other countries. Such privacy policies are visualized as tables where the objects are in the rows while the subjects are located in the columns. The cells contain either permit or deny depending on whether or not the subject is allowed to access the object. The subjects are defined by filters using attributes. So in this example the subject filter would be:

Companies meeting the condition: all locations have an attribute country with the a value that is in a list of the countries of the European Union.

The relevant section of the table for the privacy policy "Data can only be accessed by European companies." is shown in Fig. 4. Apart from the groups created by the process designer, every table does have an additional column *Default*. The

Object	Default	EU Companies
All data	deny	permit

Fig. 4. Privacy policy "Data can only be accessed by European companies." in table form

Object	Default	EU Companies
Parent	deny	permit
Child	permit	n/s

Fig. 5. Expandable objects in table to define a privacy policy for a child element different from the privacy policy for the parent element

algorithm to select the correct column when the evaluation of a policy, i.e. table, takes place is select the rightmost column whose filter does accept the subject, where Default accepts every subject.

The rows of the table represent the objects, i.e. the data the policy is about. The data is organized in an object hierarchy. Objects can be expanded to define policies for child elements as shown in Fig. 5. This table also states that EU Companies have unspecified access to the object Child. The evaluation algorithm handles *t/s* as if this column does not exist. The only column that is not allowed to have *t/s* is the *Default* column since else there would be no result for the evaluation of the policy.

The process level privacy policy applies to all data created by activities of the business process, i.e. it is assigned to all activities. If the process designer wants to define a different policy for a specific activity he defines a privacy policy on activity level. Privacy policies on acitivity level are evaluated before the process level privacy policies, i.e. activity level overrides process level. On activity level even the *Default* column can be set to *n/s*. If the evaluation of activity level policy results in *n/s* the policy on process level is evaluated.

The groups of subjects of a process designer's privacy policies can use both, companies and roles of the business process, as target. In case the process designer wants to use a business process role as the subject's filter, the systems shows up a list of the names of all swim lanes of the process. The process designer selects the appropriate entries and specifies the access rights as he does for company based filters.

The second role, i.e. the service providers, can define privacy roles that are applied to all data generated by their services in any business process. This is done on the level *General*. The definition of the policies follows the same concepts as for the process designer's policies. A service provider can override his general privacy policies by setting up a service specific privacy policy.

The third type of privacy policies are laws. Laws are provided by the platform provider as is and are not represented in a easy to read form as the process designer's and service provider's policy are.

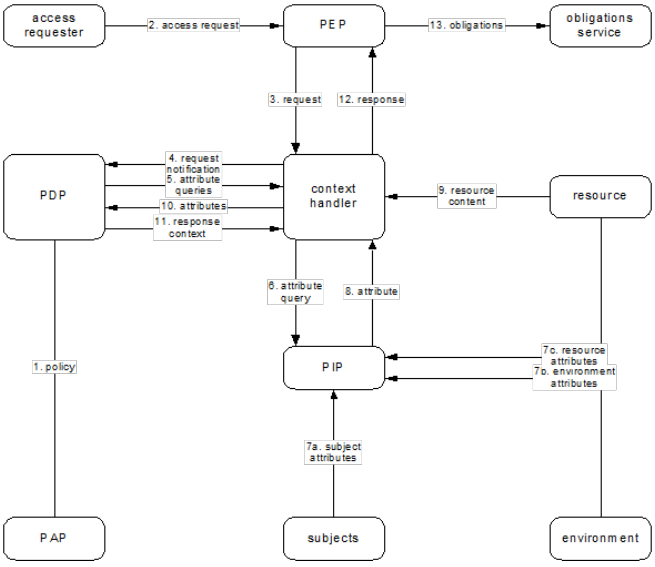


Fig. 6. Architecture and workflow of XACML 3.0 [18]

IV. Translation to XACML

This section shows how the process of transformation from table form to machine readable form of privacy policies works. As standard policy language and architecture we have selected XACML which is a well established and accepted language and architecture to define attribute based access control policies. XACML is available in version 3.0 since the beginning of 2013 [17]. Unfortunately the support by tools is not very well at this time. Most of the tools are not available for the current XACML version, others are available but are not passing the conformance tests. Fortunately AT&T published the source code of a XACML 3.0 implementation that is almost complete. The project is in a frozen state and will become an Apache Incubator Project [18].

The core architecture of XACML 3.0 is depicted in Fig. 6. Applied to our platform's architecture the PDP is the IAMS, the PEP is the BPMS and the PAP is the Privacy Management.

The document specified by the XACML language specification is xml based and specifies three main elements: *PolicySet*, *Policy*, and *Rule*.

The root element of the XACML document is either *PolicySet* or *Policy*. A *PolicySet* can contain any number of *PolicySets* or *Policies*, while a *Policy* can only contain *Rules*. A *Rule* is a single expression with an effect and a condition. Each level of these elements can have a target. Targets are used as a very easy and fast selector for applicable section of the XACML-document. Hence, target provide limited functionality but spead up evaluation time of the XACML-document tremendously.

XACML adopts the attribute based access control structure of subjects, resources, conditions, and actions. This simplifies the algorithm for translating out table based privacy policies to XACML based ones. The algorithm consists of n main steps:

1) Creation of missing tables on activity / service level

Algorithm 1 Evaluate all cells of a activity table to either permit or deny

Require: table is not empty
Ensure: $\forall$ cell $\in$ table : cell=permit or deny

```
expand all object
for all rows from top to bottom do
  for all cells from right to left do
    if cell is N/S then
      if cell in default column is N/S then
        if cell in corresponding column in process level
        table is N/S then
          cell ← value of the cell in default column of
          process level table
        else
          cell ← value of the cell in corresponding col-
          umn of process level table
        end if
      else
        cell ← value of the cell in default column
      end if
    end if
  end for
end for
```

2) Evaluation of each cell of the activity / service level tables
3) Translation of each activity / service level table into XACML
4) Combination of the XACML fragments into the target XACML documents

The algorithm is executed separately for process designer and for service provider privacy policies. For process designer view the algorithm first identifies all activities of the business process that have to privacy policy table asigned and creates such a table with all cells set to *n/s*. This ensures that all activities can be handles the same way. The second step is the most important one. In this step all *n/s* values are evaluated to either permit or deny according to the algorithm 1.

After this step there are privacy policy tables for each activity of the process containing only permit or deny. These tables are now translated into XACML policies. Each table is transferred into a *PolicySet*. The target of this *PolicySet* is set to "resource:generator:activityid equals activity-id". The resource:generator:activityid is an attribute that is derived at runtime from the context of the BPMS. The *PolicySet* contains a *Policy* per attribute of the objects and a *Rule* per column of the table. The target of the *Policy* is set to "resource-id equals objectname:attributename", where objectname ist the name of the object and attributename is the name of the attribute of the current row. The rule's effect is set to the cell's value. In rules we do not use target's but we use condition's instead, since condition's are more powerful. The condition of the rule is set according to the subject filter of the column.

When all tables are translated into XACML *PolicySets*, all *PolicySets* are put together in one *PolicySet*. This PolicySet

has "process-id equals id of the process" set as target. This *PolicySet* is the first of the three document types needed by our platform, the process related document. The platform holds one document of this type per business process.

The second document type handles the service provider's privacy policies. The platform creates one document that contains all privacy policies of all services of all service providers. The steps are very similar to the ones for the process designer's privacy policies. The following steps are executed for every service provider.

First it is ensured that there is a privacy policy table for every service. If one is missing, the system generates a table containing only *n/s* in all cells. After this step an algorithm similar to algorithm 1 is applied, with service provider level tables instead of process level tables. The result of this step is a set of privacy policy tables, one for each service, containing only permit or deny.

Each table is transformed into a *PolicySet* with the target set to "resource:generator:serviceid equals service-id", where service-id is replaced with the current service-id and resource:generator:serviceid is derived from the context of the BPMS at runtime. This *PolicySet* contains one *Policy* per attribute of the objects. The *Policies* contain one *Rule* per column as for the process based documents. All *PolicySets* of all Services are combined together into one *PolicySet* that is the second document type of the platform.

The third document type is produced by the platform provider and contains a *PolicySet* containing *Policies* for each law that is relevant to the platforms privacy preservation feature.

The process flow of this algorithm will be explained with an example in the next section.

V. Example

The algorithm presented in the previous section will be processed in this section based on the following simple use case of the logistics sector. Due to the big similarity between the process designer's and the service provider's version of the algorithm this example is focussing on the service provider's privacy policies.

In this use case there is only one resource object called *address* containing the child elements *street*, *zipcode*, and *city*. Furthermore there is a service provider called ACME that is offering two services ACME-DE and ACME-WW to the platform. The company ACME has created two subject filters, one is named "GoodRelations" and contains companies that ACME likes to work with, and one is named "NeverAgain", this filter contains companies who ACME does not want to work with again any more.

ACME defines the default privacy policy as follows. Companies are allowed to access the zipcode and the city of an address but not the street. Companies matching the filter GoodRelations are allowed to access the street, companies matching the filter NeverAgain are not allowed to access the zipcode. This general privacy policy is represented in Fig. 7.

Attribute	Default	GoodRelations	NeverAgain
Address			
Street	Deny	Permit	Deny
Zipcode	Permit	N/S	Deny
City	Permit	N/S	N/S

Fig. 7. Privacy policy for service providers on global level of ACME

Attribute	Default	GoodRelations	NeverAgain
Address			
Street	Deny	Permit	Deny
Zipcode	N/S	Permit	Deny
City	Permit	Permit	N/S

Fig. 8. Privacy policy for service providers for service ACME-DE of ACME

In addition ACME defines a privacy policy for its service ACME-DE as follows. Companies should not be able to access a streetname but should be able to access the city of an address. Companies that ACME has good relation with are allowed to access the whole address generated by the service ACME-DE and companies that ACME has made bad experiences with are not allowed to access streetnames or zipcodes of addresses generated by the service ACME-DE. No statement is made for zipcode for default companies and city for companies matching the filter NeverAgain. The resulting table form of the privacy policy is shown in Fig. 8.

ACME does not define a special privacy policy for the service ACME-WW.

The first step is to create a privacy policy table for every service. There is already a privacy policy table for ACME-DE but none for ACME-WW. Hence, the system creates such a table with only *N/S* in the cells as shown in Fig. 9.

In the next step the *N/S* entries of the privacy policy tables of each service are evaluated to either permit or deny. This is shown in Fig. 10 and Fig. 11. The arrows show the source of the entry.

While the process for ACME-DE is easy, the process for ACME-WW needs some explaination. For example the cell Zipcode / GoodRelations contains *N/S* in the service specific table. During evaluation the algorithm first looks up the value

Attribute	Default	GoodRelations	NeverAgain
Address			
Street	N/S	N/S	N/S
Zipcode	N/S	N/S	N/S
City	N/S	N/S	N/S

Fig. 9. Privacy policy for service providers for service ACME-WW of ACME

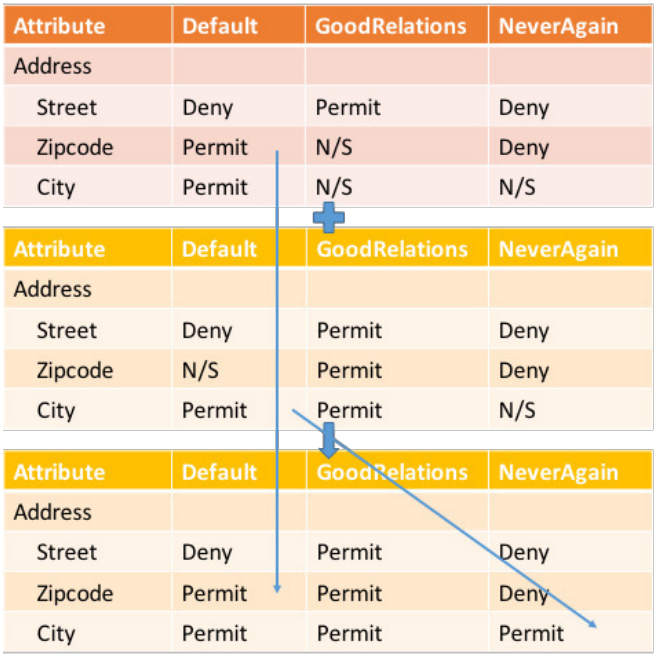


Fig. 10. Resulting privacy table for ACME-DE

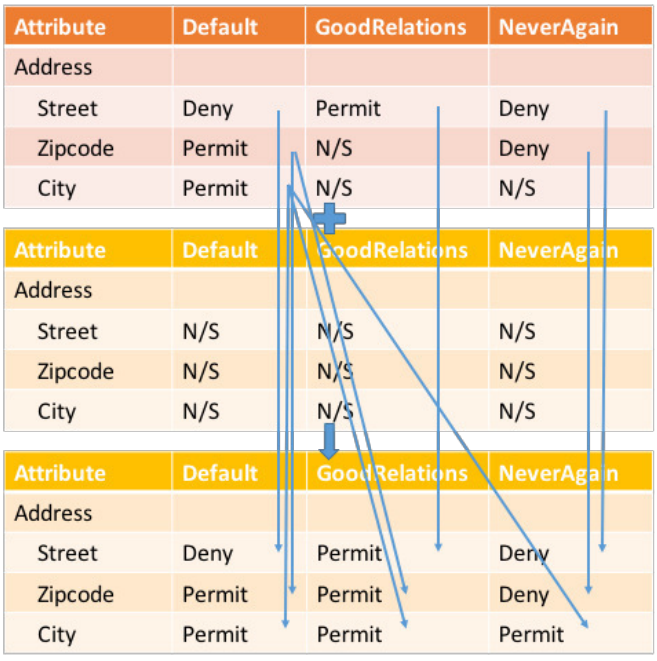


Fig. 11. Resulting privacy table for ACME-WW

in the default column for Zipcode in the service specific table. There it reads *N/S*, too. Hence, the algorithm looks up the value of the cell Zipcode / GoodRelations in the global level privacy policy table of ACME. There access control is set to *N/S* once again. So finally the algorithm falls back to the cell Zipcode / Default where it finds the resulting Permit.

For the transformation of the privacy policy tables to XACML we assume that the filters are based on company names, although they good have any complexity. The companies' names are kept in the XACML attribute *urn:prestige:iams:ldap:Company:Name*. The resulting XACML portion for the service ACME-DE is shown below.

```
<PolicySet xmlns="
    urn:oasis:names:tc:xacml:3.0
    :core:schema:wd-17" PolicySetId="
    urn:com:xacml:policy:id:8cadbdca
    -1592-4ff9-bf49-a9cccce064cf" Version=
    "1" PolicyCombiningAlgId="
    urn:oasis:names:tc:xacml:1.0:policy-
    combining-algorithm:first-applicable">
 <Description/>
 <Target>
  <AnyOf>
   <AllOf>
    <Match MatchId="
        urn:oasis:names:tc:xacml:1.0
        :function:string-equal">
     <AttributeValue DataType="http://www
         .w3.org/2001/XMLSchema#string">
         ACME-DE</AttributeValue>
     <AttributeDesignator Category="
         urn:oasis:names:tc:xacml:3.0
         :attribute-category:resource"
         AttributeId="
         urn:prestige:attribute:owner:service
         -id" DataType="http://www.w3.org
         /2001/XMLSchema#string"
         MustBePresent="true"/>
    </Match>
   </AllOf>
  </AnyOf>
 </Target>
 <Policy PolicyId="
    urn:com:xacml:policy:id:62aa47ff-cbe5
    -4bfa-afef-737cb8e10ad4" Version="1"
    RuleCombiningAlgId="
    urn:oasis:names:tc:xacml:1.0:rule-
    combining-algorithm:first-applicable"
    >
  <Target>
   <AnyOf>
    <AllOf>
     <Match MatchId="
         urn:oasis:names:tc:xacml:1.0
         :function:string-equal">
      <AttributeValue DataType="http://
          www.w3.org/2001/XMLSchema#string
          ">
          urn:prestige:data:address:street
          </AttributeValue>
      <AttributeDesignator Category="
          urn:oasis:names:tc:xacml:3.0
```

```
        :attribute-category:resource"
        AttributeId="
        urn:oasis:names:tc:xacml:1.0
        :resource:resource-id" DataType=
        "http://www.w3.org/2001/
        XMLSchema#string" MustBePresent=
        "true"/>
    </Match>
   </AllOf>
  </AnyOf>
 </Target>
 <Rule RuleId="
     urn:com:xacml:rule:id:0fff9941-8b89
     -455c-a009-26e9107e0902" Effect="
     Deny">
  <Description>NeverAgain for Street</
      Description>
  <Target/>
  <Condition>
   <Apply FunctionId="
       urn:oasis:names:tc:xacml:1.0
       :function:string-is-in">
    <Apply FunctionId="
        urn:oasis:names:tc:xacml:1.0
        :function:string-one-and-only">
     <AttributeDesignator Category="
         urn:oasis:names:tc:xacml:1.0
         :subject-category:access-subject
         " AttributeId="
         urn:prestige:iams:ldap:Company:Name
         " DataType="http://www.w3.org
         /2001/XMLSchema#string"
         MustBePresent="true"/>
    </Apply>
    <Apply FunctionId="
        urn:oasis:names:tc:xacml:1.0
        :function:string-bag">
     <AttributeValue DataType="http://
         www.w3.org/2001/XMLSchema#string
         ">NeverAgainCompanyName1</
         AttributeValue>
     <AttributeValue DataType="http://
         www.w3.org/2001/XMLSchema#string
         ">NeverAgainCompanyName2</
         AttributeValue>
    </Apply>
   </Apply>
  </Condition>
 </Rule>
 <Rule RuleId="
     urn:com:xacml:rule:id:db1e3bca-2cb3
     -42f6-bcfe-299c40189b70" Effect="
     Permit">
  <Description>GoodRelations for Street<
      /Description>
  <Target/>
  <Condition>
   <Apply FunctionId="
       urn:oasis:names:tc:xacml:1.0
       :function:string-is-in">
    <Apply FunctionId="
        urn:oasis:names:tc:xacml:1.0
        :function:string-one-and-only">
     <AttributeDesignator Category="
         urn:oasis:names:tc:xacml:1.0
         :subject-category:access-subject
         " AttributeId="
         urn:prestige:iams:ldap:Company:Name
         " DataType="http://www.w3.org
         /2001/XMLSchema#string"
         MustBePresent="true"/>
    </Apply>
    <Apply FunctionId="
        urn:oasis:names:tc:xacml:1.0
        :function:string-bag">
     <AttributeValue DataType="http://
         www.w3.org/2001/XMLSchema#string
         ">GoodRelationsCompanyName1</
         AttributeValue>
     <AttributeValue DataType="http://
         www.w3.org/2001/XMLSchema#string
         ">GoodRelationsCompanyName2</
         AttributeValue>
    </Apply>
   </Apply>
  </Condition>
 </Rule>
 <Rule RuleId="
     urn:com:xacml:rule:id:467ed7f7-d7b6
     -49e2-970c-a32fb5b66a8a" Effect="
     Deny">
  <Description>Default for Street</
      Description>
  <Target/>
 </Rule>
</Policy>
<!-- Skipping policies for zipcode and
    city -->
</PolicySet>
```

The resulting *PolicySets* for ACME-DE and ACME-WW will be combined and transferred by the Privacy Management to the IAMS. The same procedure applies to the business process related privacy policies.

VI. Conclusion

In this paper we have proposed a novel approach for defining privacy policies in business process and Cloud based scenarious. The definition is done by the end users with an easy to understand table based presentation and at the same time offers enough flexibility to fit the needs of the users. We have evaluated the approach with members of the target

group and found that it is feasible and easy to use. Especially the definition of global policies and local policies only where necessary was rated very good.

The technical implementation of the platform is working and fast enough even for a big number of policies. We have tested the platform with 500 services and a business process containing 40 activities. Every XACML request, i.e. request for privacy policy evaluation, was answered within a maximum of 27 ms over local network with no significant CPU load.

In the near future we will try to improve our platform especially in terms of visualization of privacy policies. Above all at the moment the definition of subject filters is either flexible or easy to use, depending on whether the user uses a code view or a list to select the companies from. We are planing to provide the user with a tool to define the filters using a set of attributes and a graphical editor to arrange those attributes.

Another task for the future is to perform system tests and experiments of the whole platform with companies in controlled laboratory and real world, as well.

REFERENCES

[1] T. Bittman, "The evolution of the cloud computing market," *Gartner Blog Network, http://blogs. gartner. com/thomas bittman/2008/11/03/theevolution-of-the-cloud-computing-market*, 2008.

[2] Statista, "Nutzung von cloud computing in unternehmen in deutschland in den jahren 2011 bis 2014," 2016. [Online]. Available: http://de.statista.com/statistik/daten/studie/177484/umfrage/einsatz-von-cloud-computing-in-deutschen-unternehmen-2011/

[3] B. Schwarzbach, A. Pirogov, A. Schier, and B. Franczyk, "Inter-cloud architecture for privacy-preserving collaborative bpaas," *QUIS14*, 2015.

[4] Statistisches Bundesamt, "12 % der unternehmen setzen auf cloud computing," 2014. [Online]. Available: https://www.destatis.de/DE/PresseService/Presse/Pressemitteilungen/2014/12/PD14\textunderscore467\textunderscore52911.html

[5] B. Schwarzbach, M. Glöckner, A. Pirogov, M. M. Röhling, and B. Franczyk, "Secure service interaction for collaborative business processes in the inter-cloud," in *2015 Federated Conference on Computer Science and Information Systems*, ser. Annals of Computer Science and Information Systems. IEEE, 2015, pp. 1377–1386.

[6] D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M. Y. Vardi, G. Weikum, N. Cuppens-Boulahia, F. Cuppens, and J. Garcia-Alfaro, Eds., *Data and Applications Security and Privacy XXVI*, ser. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012.

[7] H. Lindqvist, "Mandatory access control," *Master's Thesis in Computing Science, Umea University, Department of Computing Science, SE-901*, vol. 87, 2006.

[8] D. Ferraiolo, J. Cugini, and D. R. Kuhn, "Role-based access control (rbac): Features and motivations," in *Proceedings of 11th annual computer security application conference*, 1995, pp. 241–248.

[9] I. Zahid and N. Josef, "Towards semantic-enhanced attribute-based access control for cloud services," in *2012 IEEE 11th International Conference on Trust, Security and Privacy in Computing and Communications*, 2012, pp. 1223–1230. [Online]. Available: http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6296118

[10] X. Jin, R. Krishnan, and R. Sandhu, "A unified attribute-based access control model covering dac, mac and rbac," in *Data and Applications Security and Privacy XXVI*, ser. Lecture Notes in Computer Science, D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M. Y. Vardi, G. Weikum, N. Cuppens-Boulahia, F. Cuppens, and J. Garcia-Alfaro, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, vol. 7371, pp. 41–55.

[11] D. F. Ferraiolo and D. R. Kuhn, "Role-based access controls," *arXiv preprint arXiv:0903.2171*, 2009.

[12] A. Gouglidis and I. Mavridis, "domrbac: An access control model for modern collaborative systems," *computers & security*, vol. 31, no. 4, pp. 540–556, 2012.

[13] X. H. Le, T. Doll, M. Barbosu, A. Luque, and D. Wang, "An enhancement of the role-based access control model to facilitate information access management in context of team collaboration and workflow," *Journal of biomedical informatics*, vol. 45, no. 6, pp. 1084–1107, 2012.

[14] ——, "Evaluation of an enhanced role-based access control model to manage information access in collaborative processes for a statewide clinical education program," *Journal of biomedical informatics*, vol. 50, pp. 184–195, 2014.

[15] X. H. Le and D. Wang, "Development of a system framework for implementation of an enhanced role-based access control model to support collaborative processes," in *Proc 3rd USENIX Workshops on Health Security and Privacy*, 2012.

[16] V. C. Hu, D. Ferraiolo, R. Kuhn, A. Schnitzer, K. Sandlin, R. Miller, and K. Scarfone, *Guide to Attribute Based Access Control (ABAC) Definition and Considerations*. National Institute of Standards and Technology, 2014.

[17] OASIS, "extensible access control markup language (xacml) version 3.0," 2013. [Online]. Available: http://docs.oasis-open.org/xacml/3.0/xacml-3.0-core-spec-os-en.html

[18] AT&T, "At&t xacml 3.0 implementation," 2015. [Online]. Available: https://github.com/att/XACML

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1215–1222
DOI: 10.15439/2016F39
ACSIS, Vol. 8. ISSN 2300-5963

A declarative decision support framework for supply chain problems

Paweł Sitek
Kielce University of Technology
Al. 1000-lecia PP 7,25-314 Kielce, Poland,
Institute of Management and Control Systems
e-mail:sitek@tu.kielce.pl

***Abstract*—The author presents a novel declarative approach to modeling, solving and decision support for supply chain problems as a declarative decision support framework. The proposed framework makes it possible to ask different types of questions (general, specific, logical etc.). The implementation of the framework was performed in the CLP (Constraint Logic Programming) environment.**

To increase the efficiency of the framework, particularly in the area of optimization made its integration with MP (Mathematical Programming) environment. The paper also presents the implementation of illustrative model, using the proposed framework. In addition, an efficiency analysis of the presented solution in relation to the application of mathematical programming have been conducted.

I. Introduction

THE ssupply chain (SC) is commonly seen as a collection of various types of companies (raw materials, production, trade, logistics, transport, etc.) working together to improve the flow of products, information and finance. As the words in the term indicate, the supply chain is a combination of its individual links in the process of supplying products (material/products and services) to the market.

There is a considerable literature on the supply chain management problems [1,2].

The major difficulties that appear in supply chain management (SCM) include large amount of information and multiple constraints relating to each participant in the management process. These participants share many of those constraints and information items [3], which further complicates the management. The constraints have various characters and structures. The most common are the constraints related to resources, time, finance, transportation, environment, business, law and safety. They can be linear constraints, non-linear, binary integer and logical. Managers/Decision makers are typically interested in feasibility and/or optimality of the decisions they make in the environment with many constraints. The most natural way to support the decision makers is to enable them to ask questions and obtain answers within acceptable time.

Good environments for the modeling of constraints, questions and logical conditions include declarative environments, CLP (Constraint Logic Programming) in particular.

Our motivation was to develop a framework for the modeling and decision support for supply chain management problems. The use of this framework would help obtain quick answers to key questions (Is it possible…?, What If…?, What is the minimum/maximum..?) asked by managers/decision makers.

This paper proposes the concept of a declarative decision support framework for supply chain problems and presents its implementation in the CLP environment. The illustrative example shows the potential of the framework.

The remainder of the article is organized as follows. Section 2 presents problem statement, research methodology, contribution etc. The concept and implementation aspects of a declarative decision support framework are provided in Section 3. Computational examples, tests of the implementation platform and discussion are presented in Section 4. Possible extensions of the proposed approach as well as the conclusions are included in Section 5. .

II. Problem statement and methodology

Most of the SC decision and optimization problems are modeled and solved by operations research (OR) methods. The vast majority of the literature reviewed [1,2,3,4], have formulated SC models as linear programming (LP), integer programming (IP) and mixed integer linear programming (MILP) problems. Declarative environments such as CLP facilitate problem modeling and introduction of logical and symbolic constraints [5,6]. Unfortunately, high complexity and the multiple types of constraints of decision-making models as well as combinatorial nature contribute to poor efficiency of modeling in OR methods and inefficient optimization in CLP. Therefore, a new approach to modeling and solving such problems was developed [7,8,9,10]. A declarative environment was chosen as the best structure for this approach especially in modeling [5]. Mathematical programming environment was used for problem optimization [11]. This integrated approach is the basis for the creation of the implementation environment to support managers.

A. Problem description –illustrative example

The problem of supply chain management considered here refers to the supply chain in which:

- the supply chain consists of factories, distribution centers and customers (Fig. 1);
- customer orders are executed by deliveries from distribution centers;
- distribution centers are supplied by the factory;
- transport is multimodal (several modes of transport, a limited number of means of transport for each mode);
- the environmental aspects of use of transport modes are taken into account;
- different products are combined in one batch of transport;
- the cost of supplies is presented in the form of a function (in this approach, linear function of fixed and variable costs).

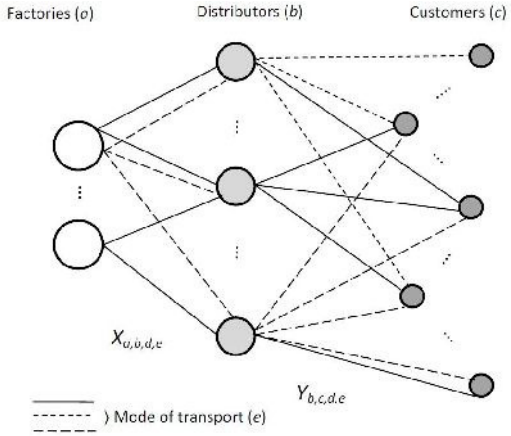


Fig. 1 The structure of supply chain for illustrative example

There are many decision and optimization problems in such supply chain management environment. Specific management problems are presented in the form of questions (possible questions (Q_i) for illustrative example are shown in Table I), which have to meet the reference set of constraints. Indices, and decision variables for the illustrative example have been reported in Table II. The set of reference constraints (1).. (22) for the illustrative example was created and its mathematical/formal notation is included in Appendix A. Brief description of reference set of constraints has been provided below.

Production capacity constraint (1) determines that all deliveries of product *d* produced by the manufacturer *a* and delivered to all distributors *b* using mode of transport *e* do not exceed the manufacturer's production capacity.

Demand constraint (2) covers all customer *c* demands for product *d* ($Z_{c,d}$) through the implementation of delivery by distributors *b* (the values of decision variables $Y_{b,c,d,e}$). The flow balance of each distributor *b* corresponds to balance constraint (3). The possibility of delivery is dependent on the distributor's technical capabilities – capacity constraint (4). Time constraint (5) ensures the terms of all deliveries are met. Transport cconstraints (6a), (6b), (7) guarantee deliveries with available transport taken into account. Constraints (8), (9), (10) set values of decision variables based on binary variables Tc_b, $Xa_{a,b,e}$, $Ya_{b,c,e}$. Dependencies (11) and (12) represent the relationship based on which total costs are calculated. In general, these may be any linear functions. The remaining constraints (13)..(22) arise from the nature of the decision variables.

TABLE I.
THE SET OF QUESTIONS (INCLUDING BUT NOT LIMITED TO)

Question	Description
Q_1	What is the minimum overall cost of timely supply?
Q_2	Can timely supply be realized at the set cost of transportation *Kt*?
Q_3	What is the minimum environmental cost *Ks* of timely supply ?
Q_4	What is the minimum cost of timely supply with the use of no more than *N* distribution centers?
Q_5	Is timely supply possible without the use of transport means *dx*?
Q_6	Is timely supply possible with the following numbers of transport means *d1,d2,d3*?
Q_7	Can timely supply be realized at the set production cost *Kp*?
Q_8	What is the minimum cost of supply execution if transport means *d1*, *d2* cannot be used simultaneously by distribution center *S1*?

Decision variables of this problem are shown in Table II.

TABLE II.
INDICES AND DECISION VARIABLES

Symbol	Description
Indices	
d	product type (d=1..D)
c	delivery point/customer/city (c=1..C)
a	manufacturer/factory (a=1..A)
b	distributor /distribution center (b=1..B)
e	mode of transport (e=1..e)
A	number of manufacturers/factories
C	number of delivery points/customers
B	number of distributors
D	number of product types
E	number of mode of transport
Decision Variables	
$X_{a,b,d,e}$	delivery quantity of product *d* from manufacturer *a* to distributor *b* using mode of transport *e*
$Xa_{a,b,e}$	if delivery is from manufacturer *a* to distributor *b* using mode of transport *e* then $Xa_{a,b,e}=1$, otherwise $Xa_{a,b,e}=0$
$Xb_{a,b,e}$	the number of courses from manufacturer *a* to distributor *b* using mode of transport *e*
$Y_{b,c,d,e}$	delivery quantity of product *d* from distributor *b* to customer *c* using mode of transport *e*
$Ya_{b,c,e}$	if delivery is from distributor *b* to customer *c* using mode of transport *e* then $Ya_{b,c,e}=1$, otherwise $Ya_{b,c,e}=0$
$Yb_{b,c,e}$	the number of courses from distributor *b* to customer *c* using mode of transport *e*
Tc_b	if distributor *b* participates in deliveries, then $Tc_b=1$, otherwise $Tc_b=0$

III. A DECLARATIVE DECISION SUPPORT FRAMEWORK FOR SUPPLY CHAIN MANAGEMENT PROBLEMS-CONCEPT AND IMPLEMENTATION

The declarative decision support framework was proposed for supply chain management problems. The concept is based on the declarative programming paradigm, which allows high level programming with the use of predicates and facts. Due to the character of problems in the supply chain management, CLP (Constraint Logic Programming) was selected from among many declarative options. The implementation of the framework was performed with the use of ECL^{i}PS^e [5,12].

The following general assumptions were applied:

- possibility of modeling constraints of any type;
- automatic generation of implementation models in the form of MILP models;
- data recorded as facts.

Figure 2 presents the general concept of the framework. The framework comprises several phases: modeling, presolving, generating and solving. It has two inputs and uses the set of facts. Inputs are the set of questions and the set of constraints to the reference model of a given problem. Based on them, the primary model of the problem is generated as a CLP model, which is then presolved. The built-in CLP method (constraint propagation [5,6]) and the method of problem transformation designed by the authors [8,9] (Section 3A) are used for this purpose. Presolving procedure results on the transformed model CLP^T. This model is the basis for the automatic generation of the MILP (Mixed Integer Linear Programming) model, which is solved in MP (with the use of an external solver or as a library of CLP). The general concept of the framework consists in modeling and presolving of a problem in the CLP environment with the final solution (including optimization) found in the MP environment. This approach is the result of experience as well as extensive research devoted to both environments and their integration [10,13,14,15,16,17]. In all its phases, the framework uses the set of facts having the structure appropriate for the problem being modeled and solved (Fig. 2). The set of facts is the informational layer of the framework, which can be implemented as a relational database, XML database, etc.

The functional layer comprises adequate sets of predicates: P_1 (CLP model generation), P_2 (CLP model presolving through constraint propagation and transformation, post-transformation generation of CLP^T model), and P_3 (generation of the final MILP^T model in the format of the MP solver).

The presolving phase is an important element of the framework as it makes it possible to simplify the model for the problem being solved and to reduce the combinatorial search space.

For the presolving phase to be effective, unfeasible combinations of model dimensions have to occur. In practice, unfeasible combinations of the index of decision variables and/or facts occur. The proposed framework uses constraint propagation and transformation for the presolving procedure. Constraint propagation is a concept and method that appears in constrained-based environments. Constraint propagation embeds any reasoning which consists in explicity forbiding values from some varable domain of a problem, because all constraints can not be satisfied otherwise [6].

Transformation transforms decision variables of the problem along with constraints and facts. The transformation of facts for the illustrative example is shown in Fig. 3, and the post-transformation variables are compiled in Table AII.

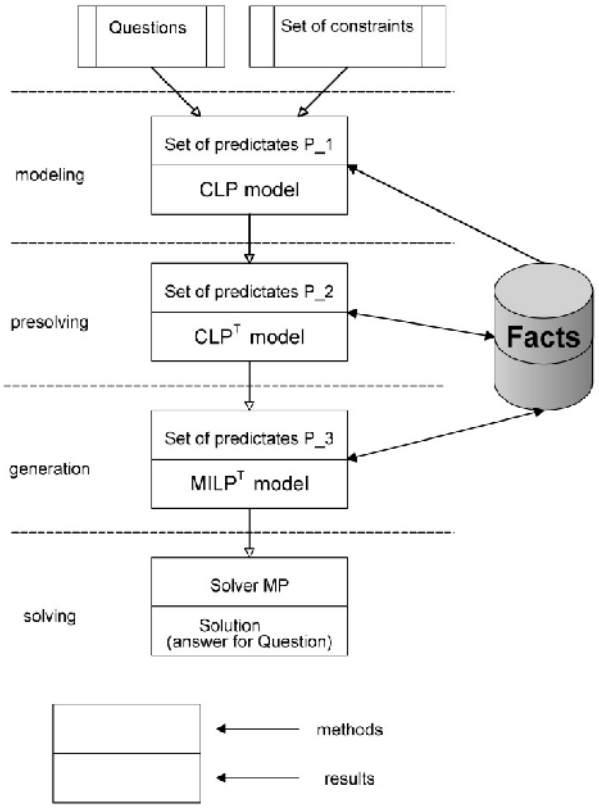


Fig. 2 A concept of a declarative decision support framework

A. Transformation of the problem-presolving phase

In the case of the problem presented, the transformation consisted in changing the problem representation from graph to routing. Instead of analyzing all possible trnasport connections from the factory to the distribution center and then from the center to the customer, only the feasible connections (factory-center-customer) were generated and named routes. This resulted in the removal of certain indices and in the aggregation of other indices for decision variables, parameters, etc., which eventually led to the reduction in the number of decision variables and constraints [7,8]. The new set of decision variable, constraints and facts was the basis for creating the CLP^{T}model.

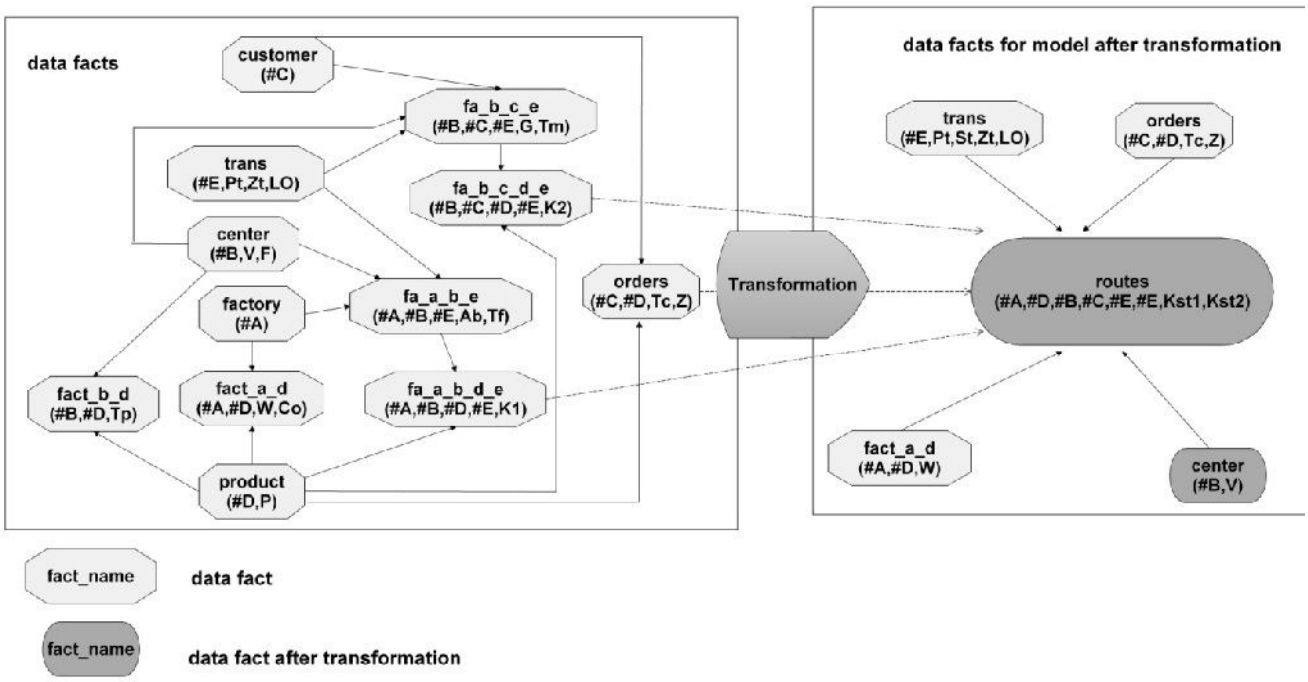


Fig. 3 Structure of facts for illustrative example before and after transformation

IV. COMPUTATION EXAMPLES FOR ILLUSTRATIVE MODEL

In order to verify and evaluate the proposed approach, many numerical experiments were performed. All the examples relate to the supply chain with five factories *(a=1..5),* four distributor centers *(b=1..4)*, fifteen customers *(c=1..15),* four modes of transport *(e=1..4),* and fifteen types of products *(d=1..15)* and fifty orders *(No)*.

All data instances for these experiments were recorded in the form of facts and included Appendix B. The structure of facts and their description has been shown in Fig. 3 and Table V.

Computational experiments consisted in asking questions Q_1..Q_8 to illustrative example. For each question was generated and solved suitable implementation model using declarative decision support framework. The answers to these questions are shown in Table III.

The answer to question Q_1 is the minimum entire cost of timely delivery all orders (Table III). This cost is the aggregate costs of the entire chain and consists of five elements (23). The first element comprises the fixed costs associated with the operation of the distributor involved in the delivery (e.g. distribution centre, warehouse, etc.). The second element corresponds to environmental costs of using various means of transport. Those costs are dependent on the number of courses of the given means of transport, and on the other hand, on the environmental levy, which in turn may depend on the use of fossil fuels and carbon-dioxide emissions. The third and fourth element determined by the cost of transportation. The last element is the cost of production. The answer to question Q_2 defines the feasibility of timely supply/delivery at the set (closed) transportation cost (25). Table II shows two versions of the question with different parameters. The answer to question Q_3 determines minimum environmental costs (26) of timely delivery (Table I).

Two versions of the answer to question Q_4 are shown in Table III. These versions define the minimum overall cost of timely delivery when two distribution centers are used (Q_4_A) and when one distribution center is used (Q_4_B).

The possibility of executing timely delivery without the use of selected transportation means is defined in the answer to question Q_5 (three versions of the question are shown in Table III). The answer to question Q_6 specifies the possibility of timely delivery at the set number of each transportation means (Table III).

The possibility of executing timely delivery at the established, closed production cost (24) is included in the answer to question Q_7. Table III shows three versions of the answer. The answer to question Q_8 specifies the minimum cost of timely delivery at the logical condition met, ruling out the possibility of that the selected center can be used concurrently by two different transportation means.

In the second phase of the experiments, a comparative analysis was performed for questions Q_1 and Q_8 (most compute-intensive of all) for different numbers of orders (No) in two environments (declarative decision support framework ($MILP^T$) and MP (MILP)) to evaluate the effectiveness and efficiency of the proposed framework relative to the classical MP environment.

Obtained more than 50-fold reduction of time searching for solutions (Table IV). This is due to the fact that the application of the framework has allowed the up to 50-fold

reduction of integer decision variables up to 20-fold reduction of constraints (Table IV).

TABLE III.
ANSWERS TO QUESTION FOR ILLUSTRATIVE EXAMPLE

Question	Parameters	Answer
Q_1	----	17805
Q_2_A	Kt≤1520, T≤15	YES
Q_2_B	Kt≤1400, T≤20	NO
Q_3	----	6250
Q_4_A	N=2	17945
Q_4_B	N=1	18126
Q_5_A	S_1=0,T=12	YES
Q_5_B	S_2=0,T=12	YES
Q_5_C	S_3=0,T=12	NO
Q_6_A	S_1=5, S_2=5, S_3=5, S_4=5,T=9	YES
Q_6_B	S_1=5, S_2=5, S_3=5, S_4=5,T=9	NO
Q_6_C	S_1=5, S_2=5, S_3=5, S_4=5,T=12	YES
Q_7_A	Kp≤9820, T≤8	NO
Q_7_B	Kp≤9820, T≤12	YES
Q_7_C	Kp≤39200, T≤8	YES
Q_8	----	17856

TABLE IV.
NUMERICAL EXPERIMENTS ON THE EFFICIENCY

No	Model	Vint	C	Answer	T
Q_1					
10	MILP	45286	29510	5500	26
	$MILP^T$	860	1590	5500	2
25	MILP	45286	33990	9100	101
	$MILP^T$	1268	1596	9100	2
50	MILP	45286	41990	17805	545
	$MILP^T$	2078	1603	17805	13
75	MILP	45286	49990	27695^*	600^{**}
	$MILP^T$	2996	1612	27175	56
Q_8					
10	MILP	45290	29518	5610	45
	$MILP^T$	864	1590	5610	3
25	MILP	45290	33998	9210	112
	$MILP^T$	1272	1604	9210	2
50	MILP	45290	41998	17856	588
	$MILP^T$	2082	1611	17856	14
75	MILP	45290	49998	28102	600^{**}
	$MILP^T$	3000	1620	27450	68

No-the number of orders; **Vint**-the number of integer decision variables, **C**-the number of constraints, **T**-time of finding solution (in seconds)

TABLE V.
STRUCTURE OF THE FACTS FOR ILLUSTRATIVE EXAMPLE

Fact	Description
product(#D,P).	products , D-product ID, P-the capacity occupied by product.
factory(#A).	factories , A-factory ID.
fact_a_d(#A,#D,W,Co)	products in factories, A-factory ID, D-product ID, W-production capacity at factory A for product D, Co-the cost of product D in factory A.
center(#B,V,F)	distribution centers, B-center ID, V-maximum capacity, F-the fixed cost of distribution center.
fact_b_d(#B,#D,Tp)	products in distribution centers, B-center ID, D-product ID, Tp-the time needed for distributor B to prepare the shipment of product D.
trans(#E,Pt,Zt,Od)	transportation, E-number of mode of transportation, Pt-the capacity of transport unit, St-payload ,Zt-the number of transport units using mode of transportation E, Od-the environmental cost
customer(#C)	customers, C-ustomer ID
fa_a_b_e(#A,#B,#E,Ab,Tf)	delivery from factory to distribution center using mode of transportation E, A-factory ID, B-center ID, E-number of mode of transportation, Ab-the fixed cost of delivery, Tf-the time of delivery
fa_b_c_e(#B,#C,#E,G,Tm)	delivery from distribution center to customer using mode of transportation E, B-center ID, C-customer ID, E-number of mode of transportation, G-the fixed cost of delivery, Tm-the time of delivery
fa_a_b_d_e(#A,#B,#D,#E,K1)	the variable costs of delivery of product, A-factory ID, B-center ID, D-product ID, E-number of mode of transportation, Ks3-the variable cost of delivery of product from factory to distribution center (optional)
fa_b_c_d_e(#B,#C,#D,#E,K2)	the variable costs of delivery of product, B-center ID, C-customer ID, D-product ID, E-number of mode of transportation, Ks4-the variable cost of delivery of product from distribution center to customer (optional)
orders(#C,#D,Tc,Z)	orders, C-customer ID, D-product ID, the cut-off time of delivery to customer of product, Z-customer demand for product

V. CONCLUSIONS

Two types of questions can be asked in the proposed declarative decision support framework.

General questions may require domain solution, which in practice determines the availability of capacity, the number of transport units, timely supply etc. The specific wh-questions will in practice define the best, fastest, cheapest, or the most expensive of the possible solutions. To obtain answers to these questions, optimization is necessary. Both

question types can contain logical conditions relating, for example, to the disjoint use of transport units, distributors, etc. The illustrative example shows only part of potential of the framework designed to increase both the speed and the size of the problems solved (Table IV).

This is particularly evident if we compare the possibilities of the framework in relation to the classical approach based on mathematical programming (Table IV).

Further work will consist in the implementation of more complex models [18], uncertainty, fuzzy logic etc. [19], and as a cloud internet application [20]. New questions will be implemented to broaden the scope of decision support.

It is also considering development of models to take account product demand interdependencies [21]. In the future it is planned to integrate framework with ERP and APS systems [22]. It is planned to also use a hybrid approach to optimize the use of graphs, for example, to image retrieval [23].

APPENDIX A

TABLE A1
SUMMARY PARAMETERS

Symbol	Description
	Input parameters
F_b	the fixed cost of distributor/distribution center *b*
P_d	the area/volume occupied by product *d*
V_b	distributor *b* maximum capacity/volume
$W_{a,d}$	production capacity at factory *a* for product *d*
$Co_{a,d}$	the cost of product *d* at factory *a*
$R_{b,d}$	if distributor *b* can deliver product *d* then $R_{b,d}=1$, otherwise $R_{b,d}=0$
$Tp_{b,d}$	the time needed for distributor *b* to prepare the shipment of product *d*
$Tc_{c,d}$	the cut-off time of delivery to the delivery point/customer *c* of product *d*
$Z_{c,d}$	customer demand/order *c* for product *d*
Zt_e	the number of transport units using mode of transport *e*
Pt_e	the capacity of transport unit using mode of transport *e*
$Tf_{a,b,e}$	the time of delivery from manufacturer *a* to distributor *b* using mode of transport *e*
St_e	the payload of transport unit using mode of transport *e*
$K1_{a,b,d,e}$	the variable cost of delivery of product *d* from manufacturer *a* to distributor *b* using mode of transport *e*
$R1_{a,b,e}$	if manufacturer *a* can deliver to distributor *b* using mode of transport *e* then $R1_{a,b,e}=1$, otherwise $R1_{a,b,e}=0$
$A_{a,b,e}$	the fixed cost of delivery from manufacturer *a* to distributor *b* using mode of transport *e*
$Koa_{b,c,e}$	the total cost of delivery from distributor *b* to customer *c* using mode of transport *e*
$Tm_{b,c.e}$	the time of delivery from distributor *b* to customer *c* using mode of transport *e*
$K2_{b,c,d,e}$	the variable cost of delivery of product *d* from distributor *b* to customer *c* using mode of transport *e*
$R2_{b,c,e}$	if distributor *b* can deliver to customer *c* using mode of transport *e* then $R2_{b,c,e}=1$, otherwise $R2_{b,c,e}=0$
$G_{b,c,e}$	the fixed cost of delivery from distributor *b* to customer *c* using mode of transport *e*
$Kog_{b,c,e}$	the total cost of delivery from distributor *b* to customer *c* using mode of transport *e*
Od_e	the environmental cost of using mode of transport *e*
CW	Arbitrarily large constant

$$\sum_{b=1}^{B}\sum_{e=1}^{E} X_{a,b,d,e} \cdot R_{b,d} \le W_{a,d} \forall a = 1..A, d = 1..D \quad (1)$$

$$\sum_{b=1}^{B}\sum_{e=1}^{e} (Y_{b,c,d,e} \cdot R_{b,d}) \ge Z_{c,d} \forall j = c..C, d = 1..D \quad (2)$$

$$\sum_{a=1}^{A}\sum_{e=1}^{E} X_{a,b,d,e} = \sum_{c=1}^{C}\sum_{e=1}^{E} Y_{b,c,d,e} \forall b = 1..B, d = 1..D \quad (3)$$

$$\sum_{d=1}^{D} (P_d \cdot \sum_{a=1}^{A}\sum_{e=1}^{E} X_{a,b,d,e}) \le Tc_b \cdot V_b \forall b = 1..B \quad (4)$$

$$Xa_{a,b,e} \cdot Tf_{a,b,e} + Xa_{a,b,e} \cdot Tp_{b,d} + Ya_{b,c,e} \cdot Tm_{b,c,e} \le Tc_{b,d} \quad \forall a = 1..A,\ b = 1..B, C = 1..C, d = 1..D, e = 1..E \quad (5)$$

$$R1_{a,b,e} \cdot Xb_{a,b,e} \cdot Pt_e \ge X_{a,b,d,e} \cdot P_e \quad \forall a = 1..A, b = 1..B, d = 1..D, e = 1..E \quad (6a)$$

$$R2_{b,c,e} \cdot Yb_{b,c,e} \cdot Pt_e \ge Y_{b,c,d,e} \cdot P_e \quad \forall b = 1..B, c = 1..C, d = 1..D, e = 1..E \quad (6b)$$

$$\sum_{a=1}^{A}\sum_{b=1}^{B} Xb_{a,b,e} + \sum_{b=1}^{B}\sum_{c=1}^{C} Yb_{b,c,e} \le Zt_e \forall e = 1..E \quad (7)$$

$$\sum_{a=1}^{A}\sum_{e=1}^{E} Xb_{a,b,e} \le CW \cdot Tc_e \forall e = 1..E \quad (8)$$

$$Xb_{a,b,e} \le CW \cdot Xa_{a,b,e} \forall a = 1..A, b = 1..B, e = 1..E \quad (9)$$

$$Yb_{b,c,e} \le CW \cdot Ya_{b,c,e} \forall b = 1..B, c = 1..C, e = 1..E \quad (10)$$

$$Koa_{a,b,e} = Ab_{a,b,e} \cdot Xb_{a,b,e} + \sum_{d=1}^{D} K1_{a,b,d,e} \cdot X_{a,b,d,e} \quad \forall a = 1..A, b = 1..B, e = 1..E \quad (11)$$

$$Kog_{b,c,e} = G_{b,c,e} \cdot Yb_{b,c,e} + \sum_{d=1}^{D} K2_{b,c,d,e} \cdot Y_{b,c,d,e} \quad \forall b = 1..B, c = 1..C, e = 1..E \quad (12)$$

$$X_{a,b,d,e} \ge 0 \forall a = 1..A, b = 1..B, d = 1..D, e = 1..E \quad (13)$$

$$Xb_{a,b,e} \ge 0 \forall a = 1..A,\ b = 1..B,\ e = 1..E \quad (14)$$

$$Yb_{b,c,e} \ge 0 \forall b = 1..B,\ c = 1..C,\ e = 1..E \quad (15)$$

$$X_{a,b,d,e} \in C \forall a = 1..A, b = 1..B, d = 1..D, e = 1..E \quad (16)$$

$$Xb_{a,b,e} \in C \forall a = 1..A,\ b = 1..B,\ e = 1..E \quad (17)$$

$$Y_{b,c,d,e} \in C \forall b = 1..B, c = 1..C, d = 1..D, e = 1..E \quad (18)$$

$$Yb_{b,c,d} \in C \forall b = 1..B,\ c = 1..C,\ e = 1..E \quad (19)$$

$$Xa_{a,b,e} \in \{0,1\} \forall a = 1..A,\ b = 1..B,\ e = 1..E \quad (20)$$

$$Ya_{b,c,e} \in \{0,1\} \forall b = 1..B,\ c = 1..C,\ e = 1..E \quad (21)$$

$$Tc_b \in \{0,1\} \forall b = 1..B \quad (22)$$

$$K = \sum_{b=1}^{B} (F_b \cdot Tc_b) + \sum_{e=1}^{E} Od_e (\sum_{a=1}^{A}\sum_{b=1}^{B} Xb_{a,b,e} + \sum_{b=1}^{B}\sum_{c=1}^{C} Yb_{b,c,e}) + \sum_{a=1}^{A}\sum_{b=1}^{B}\sum_{e=1}^{E} Koa_{a,b,e} + \sum_{b=1}^{B}\sum_{c=1}^{C}\sum_{e=1}^{E} Kog_{s,j,d} + \sum_{a=1}^{A}\sum_{d=1}^{D} (Co_{a,d} \cdot \sum_{b=1}^{B}\sum_{e=1}^{E} X_{a,b,d,e}) \quad (23)$$

$$K_p = \sum_{a=1}^{A}\sum_{d=1}^{D} (Co_{a,d} \cdot \sum_{b=1}^{B}\sum_{e=1}^{E} X_{a,b,d,e}) \quad (24)$$

$$K_t = \sum_{a=1}^{A}\sum_{b=1}^{B}\sum_{e=1}^{E} Koa_{a,b,e} + \sum_{b=1}^{B}\sum_{c=1}^{C}\sum_{e=1}^{E} Kog_{s,j,d} \quad (25)$$

$$K_s = \sum_{e=1}^{E} Od_e (\sum_{a=1}^{A}\sum_{b=1}^{B} Xb_{a,b,e} + \sum_{b=1}^{B}\sum_{c=1}^{C} Yb_{b,c,e}) \quad (26)$$

TABLE AII.
DECISION VARIABLES AFTER TRANSFORMATION

Decision variables	
delivery quantity of product *d* for route from manufacturer *a* to distributor *b* using mode of transport e_1 and from distributor *b to* customer *c* using mode of transport e_2	X^{T}_{a,b,c,d,e_1,e_2}
the number of courses from manufacturer *a* to distributor *b* using mode of transport *e*	$Xb^{T}_{a,b,e}$
the number of courses from distributor *b* to customer *c* using mode of transport *e*	$Yb^{T}_{b,c,e}$
if distributor *b* participates in deliveries, then $Tc_b=1$, otherwise $Tc_b=0$	Tc^{T}_{b}

APPENDIX B

```
%Products
product(d1,10,2).  product(d2,20,4).
product(d3,20,4).  product(d4,20,5).
product(d5,10,2).  product(d6,20,2).
product(d7,20,3).  product(d8,20,4).
product(d9,10,4).  product(d10,20,5).
product(d11,20,6). product(d12,20,2).
product(d13,30,3). product(d14,30,3).
product(d15,20,4).
%Factories
factory(a1). factory(a2). factory(a3).
factory(a4). factory(a5).
%Distribution Centers
center(b1,1000,100). center(b2,1000,200).
center(b3,1000,300). center(b4,1000,400).
%Products in distribution centers
fact_b_d(b1,d1,1).  fact_b_d(b1,d2,1).
fact_b_d(b1,d3,1).  fact_b_d(b1,d4,1).
fact_b_d(b1,d5,1).  fact_b_d(b1,d6,1).
fact_b_d(b1,d7,1).  fact_b_d(b1,d8,1).
fact_b_d(b1,d9,1).  fact_b_d(b1,d10,1).
fact_b_d(b1,d11,1). fact_b_d(b1,d12,1).
fact_b_d(b2,d5,1).  fact_b_d(b2,d6,1).
fact_b_d(b2,d7,1).  fact_b_d(b2,d8,1).
fact_b_d(b2,d9,1).  fact_b_d(b2,d10,1).
fact_b_d(b2,d11,1). fact_b_d(b2,d12,1).
fact_b_d(b2,d13,1). fact_b_d(b2,d14,1).
fact_b_d(b2,d15,1). fact_b_d(b3,d1,1).
fact_b_d(b3,d2,1).  fact_b_d(b3,d3,1).
fact_b_d(b3,d4,1).  fact_b_d(b3,d5,1).
fact_b_d(b3,d6,1).  fact_b_d(b3,d7,1).
fact_b_d(b3,d8,1).  fact_b_d(b3,d9,1).
fact_b_d(b3,d10,1). fact_b_d(b3,d11,1).
fact_b_d(b3,d12,1). fact_b_d(b3,d13,1).
fact_b_d(b3,d14,1). fact_b_d(b3,d15,1).
fact_b_d(b4,d1,1).  fact_b_d(b4,d2,1).
fact_b_d(b4,d3,1).  fact_b_d(b4,d4,1).
fact_b_d(b4,d5,1).  fact_b_d(b4,d6,1).
fact_b_d(b4,d7,1).  fact_b_d(b4,d8,1).
fact_b_d(b4,d9,1).  fact_b_d(b4,d10,1).
fact_b_d(b4,d11,1). fact_b_d(b4,d12,1).
fact_b_d(b4,d13,1). fact_b_d(b4,d14,1).
fact_b_d(b4,d15,1).
%Products infactories
fact_a_d(a1,d1,800,400). fact_a_d(a2,d2,800,400).
fact_a_d(a3,d3,800,400). fact_a_d(a1,d4,800,400).
fact_a_d(a2,d5,800,400). fact_a_d(a3,d6,800,400).
fact_a_d(a1,d7,800,400). fact_a_d(a2,d8,800,400).
fact_a_d(a3,d9,800,400). fact_a_d(a1,d10,800,400).
fact_a_d(a2,d11,800,400).fact_a_d(a3,d12,800,400).
fact_a_d(a1,d13,800,400).fact_a_d(a2,d14,800,400).
fact_a_d(a3,d15,800,400).fact_a_d(a4,d1,800,100).
fact_a_d(a4,d2,800,100). fact_a_d(a4,d3,800,100).
fact_a_d(a4,d4,800,100). fact_a_d(a4,d5,800,100).
fact_a_d(a4,d6,800,100). fact_a_d(a4,d7,800,100).
fact_a_d(a4,d8,800,100). fact_a_d(a4,d6,800,100).
fact_a_d(a4,d7,800,100). fact_a_d(a4,d8,800,100).
fact_a_d(a4,d9,800,100). fact_a_d(a4,d10,800,100).
fact_a_d(a4,d11,800,100).fact_a_d(a4,d12,800,100).
fact_a_d(a4,d13,800,100).fact_a_d(a4,d14,800,100).
fact_a_d(a4,d15,800,100).fact_a_d(a5,d1,800,100).
fact_a_d(a5,d2,800,100). fact_a_d(a5,d3,800,100).
fact_a_d(a5,d4,800,100). fact_a_d(a5,d5,800,100).
fact_a_d(a5,d6,800,100). fact_a_d(a5,d7,800,100).
fact_a_d(a5,d8,800,100). fact_a_d(a5,d6,800,100).
fact_a_d(a5,d7,800,100). fact_a_d(a5,d8,800,100).
fact_a_d(a5,d9,800,100). fact_a_d(a5,d10,800,100).
fact_a_d(a5,d11,800,100).fact_a_d(a5,d12,800,100).
fact_a_d(a5,d13,800,100).fact_a_d(a5,d14,800,100).
fact_a_d(a5,d15,800,100).
%Transportation modes
trans(e1,850, 5,500). trans(e2,600,10,300).
trans(e3,350,10,200). trans(e4,150,10,100).
%Deliveries from factories to distribution centers
fa_a_b_e(a1,b1,e1,270,1).fa_a_b_e(a2,b1,e1,300,1).
fa_a_b_e(a3,b1,e1,400,1).fa_a_b_e(a4,b1,e1,500,8).
fa_a_b_e(a5,b1,e1,400,8).fa_a_b_e(a1,b1,e2,170,2).
fa_a_b_e(a2,b1,e2,200,2).fa_a_b_e(a3,b1,e2,300,2).
fa_a_b_e(a4,b1,e2,400,8).fa_a_b_e(a5,b1,e2,300,8).
fa_a_b_e(a1,b2,e1,270,1).fa_a_b_e(a2,b2,e1,300,1).
fa_a_b_e(a3,b2,e1,400,1).fa_a_b_e(a4,b2,e1,500,8).
fa_a_b_e(a5,b2,e1,400,8).fa_a_b_e(a1,b2,e2,170,1).
fa_a_b_e(a2,b2,e2,200,1).fa_a_b_e(a3,b2,e2,300,1).
fa_a_b_e(a4,b2,e2,300,8).fa_a_b_e(a5,b2,e2,300,8).
fa_a_b_e(a1,b3,e1,270,1).fa_a_b_e(a2,b3,e1,300,1).
fa_a_b_e(a3,b3,e1,400,1).fa_a_b_e(a4,b3,e1,500,8).
fa_a_b_e(a5,b3,e1,400,8).fa_a_b_e(a1,b3,e2,170,1).
fa_a_b_e(a2,b3,e2,200,1).fa_a_b_e(a3,b3,e2,300,1).
fa_a_b_e(a4,b3,e2,400,8).fa_a_b_e(a5,b3,e2,300,8).
fa_a_b_e(a1,b4,e1,270,1).fa_a_b_e(a2,b4,e1,300,1).
fa_a_b_e(a3,b4,e1,400,1).fa_a_b_e(a4,b4,e1,500,8).
fa_a_b_e(a5,b4,e1,400,8).fa_a_b_e(a1,b4,e2,170,1).
fa_a_b_e(a2,b4,e2,200,1).fa_a_b_e(a3,b4,e2,300,1).
fa_a_b_e(a4,b4,e2,400,8).fa_a_b_e(a5,b4,e2,300,8).
%Customers
customer(c2).  customer(c3).  customer(c4).
customer(c5).  customer(c6).  customer(c7).
customer(c8).  customer(c9).  customer(m10).
customer(m11). customer(m12). customer(m13).
customer(m14). customer(m15).
%Deliveries from distribution centers to customers
fa_b_c_e(b1,c1,e4,30,1). fa_b_c_e(b1,c2,e4,30,1).
fa_b_c_e(b1,c3,e4,30,1). fa_b_c_e(b1,c4,e4,30,1).
fa_b_c_e(b1,c5,e4,30,1). fa_b_c_e(b1,c6,e4,30,1).
fa_b_c_e(b1,c7,e4,30,1). fa_b_c_e(b1,c8,e4,30,1).
fa_b_c_e(b1,c9,e4,30,1). fa_b_c_e(b1,m10,e4,30,1).
fa_b_c_e(b1,m11,e4,30,1).fa_b_c_e(b1,m12,e4,30,1).
fa_b_c_e(b1,m13,e4,30,1).fa_b_c_e(b1,m14,e4,30,1).
fa_b_c_e(b1,m15,e4,30,1).fa_b_c_e(b2,c1,e4,30,1).
fa_b_c_e(b2,c2,e4,30,1). fa_b_c_e(b2,c3,e4,30,1).
fa_b_c_e(b2,c4,e4,30,1). fa_b_c_e(b2,c5,e4,30,1).
fa_b_c_e(b2,c6,e4,30,1). fa_b_c_e(b2,c7,e4,30,1).
fa_b_c_e(b2,c8,e4,30,1). fa_b_c_e(b2,c9,e4,30,1).
fa_b_c_e(b2,m10,e4,30,1).fa_b_c_e(b2,m11,e4,30,1).
fa_b_c_e(b2,m12,e4,30,1).fa_b_c_e(b2,m13,e4,30,1).
fa_b_c_e(b2,m14,e4,30,1).fa_b_c_e(b2,m15,e4,30,1).
fa_b_c_e(b3,c1,e4,30,1). fa_b_c_e(b3,c2,e4,30,1).
fa_b_c_e(b3,c3,e4,30,1). fa_b_c_e(b3,c4,e4,30,1).
fa_b_c_e(b3,c5,e4,30,1). fa_b_c_e(b3,c6,e4,30,1).
fa_b_c_e(b3,c7,e4,30,1). fa_b_c_e(b3,c8,e4,30,1).
fa_b_c_e(b3,c9,e4,30,1). fa_b_c_e(b3,m10,e4,30,1).
fa_b_c_e(b3,m11,e4,30,1).fa_b_c_e(b3,m12,e4,30,1).
fa_b_c_e(b3,m13,e4,30,1).fa_b_c_e(b3,m14,e4,30,1).
fa_b_c_e(b3,m15,e4,30,1).fa_b_c_e(b4,c1,e4,30,1).
fa_b_c_e(b4,c2,e4,30,1). fa_b_c_e(b4,c3,e4,30,1).
fa_b_c_e(b4,c4,e4,30,1). fa_b_c_e(b4,c5,e4,30,1).
fa_b_c_e(b4,c6,e4,30,1). fa_b_c_e(b4,c7,e4,30,1).
fa_b_c_e(b4,c8,e4,30,1). fa_b_c_e(b4,c9,e4,30,1).
fa_b_c_e(b4,m10,e4,30,1).fa_b_c_e(b4,m11,e4,30,1).
fa_b_c_e(b4,m12,e4,30,1).fa_b_c_e(b4,m13,e4,30,1).
fa_b_c_e(b4,m14,e4,30,1).fa_b_c_e(b4,m15,e4,30,1).
```

```
fa_b_c_e(b1,c1,e3,50,1). fa_b_c_e(b1,c6,e3,50,1).
fa_b_c_e(b1,c7,e3,50,1). fa_b_c_e(b1,c8,e3,50,1).
fa_b_c_e(b1,c9,e3,50,1). fa_b_c_e(b2,c3,e3,50,1).
fa_b_c_e(b2,c4,e3,50,1). fa_b_c_e(b2,m13,e3,50,1).
fa_b_c_e(b2,m14,e3,50,1).fa_b_c_e(b3,c7,e3,50,1).
fa_b_c_e(b3,c8,e3,50,1). fa_b_c_e(b3,m13,e3,50,1).
fa_b_c_e(b3,m14,e3,50,1).fa_b_c_e(b3,m15,e3,50,1).
fa_b_c_e(b4,c1,e3,50,1). fa_b_c_e(b4,c2,e3,50,1).
fa_b_c_e(b4,c3,e3,50,1). fa_b_c_e(b4,c4,e3,50,1).
fa_b_c_e(b4,c5,e3,50,1). fa_b_c_e(b4,c9,e3,50,1).
fa_b_c_e(b4,m10,e3,50,1).fa_b_c_e(b4,c3,e1,50,1).
fa_b_c_e(b4,c4,e1,50,1). fa_b_c_e(b4,c5,e1,50,1).
fa_b_c_e(b4,c9,e1,50,1). fa_b_c_e(b4,m10,e1,50,1).
%Orders
orders(1,d1,c1,30,2). orders(2,d1,c2,30,2).
orders(3,d2,c3,30,2). orders(4,d1,c4,30,2).
orders(5,d3,c5,30,2). orders(6,d1,c6,30,2).
orders(7,d4,c7,30,2).  orders(8,d1,c8,30,2).
orders(9,d5,c9,30,2).  orders(10,d1,c10,30,2).
orders(11,d6,c1,30,2). orders(12,d2,c2,30,2).
orders(13,d8,c3,30,2). orders(14,d2,c4,30,2).
orders(15,d9,c5,30,2). orders(16,d2,c6,30,2).
orders(17,d10,c7,30,2).orders(18,d2,c8,30,2).
orders(19,d2,c9,30,2). orders(20,d2,c10,30,2).
orders(21,d3,c1,30,2). orders(22,d3,c2,30,2).
orders(23,d3,c3,30,2). orders(24,d3,c4,30,2).
orders(25,d3,c5,30,2). orders(26,d3,c6,30,2).
orders(27,d3,c7,30,2). orders(28,d3,c8,30,2).
orders(29,d3,c9,30,2). orders(30,d3,c10,30,2).
orders(41,d4,c1,30,2).orders(42,d4,c2,30,2).
orders(43,d4,c3,30,2).orders(44,d4,c4,30,2).
orders(45,d4,c5,30,2).orders(46,d4,c6,30,2).
orders(47,d4,c7,30,2).orders(48,d4,c8,30,2).
orders(49,d4,c9,30,2).orders(50,d10,m10,30,2).
orders(51,d5,c1,30,2).orders(52,d11,c2,30,2).
orders(53,d5,c3,30,2).orders(54,d12,c4,30,2).
orders(55,d5,c5,30,2).orders(56,d13,c6,30,2).
orders(57,d5,c7,30,2).orders(58,d14,c8,30,2).
orders(59,d5,c9,30,2).orders(60,d15,c10,30,2).
```

REFERENCES

[1] K. Burgess, P. J. Singh, R. Koroglu, "Supply chain management: a structured literature review and implications for future research", in: *International Journal of Operations & Production Management*, Vol. 26 Issue: 7, pp.703 – 729, 2006.

[2] K. C. Tan, "A framework of supply chain management literature", in: *European Journal of Purchasing & Supply Management,* 7, pp 39-48, 2001.

[3] G.Q. Huang, J.S.K. Lau, K.L. Mak, "The impacts of sharing production information on supply chain dynamics: a review of the literature", in: *International Journal of Production Research,* 41, pp. 1483–1517, 2003.

[4] J. Mula, D. Peidro, M. Diaz-Madronero, E. Vicens, "Mathematical programming models for supply chain production and transport planning", in: *European Journal of Operational Research,* 204, pp. 377–390, 2010.

[5] K. Apt, M. Wallace, "Constraint Logic Programming using Eclipse". Cambridge: Cambridge University Press, 2006.

[6] F. Rossi, P. Van Beek, T. Walsh, "*Handbook of Constraint Programming"*, New York: Elsevier Sc. Inc, 2006.

[7] P. Sitek J. Wikarek, "A hybrid method for modeling and solving constrained search problems", in: *Federated Conference on Computer Science and Information Systems (FedCSIS 2013)*, pp. 385-392, 2013.

[8] P. Sitek, J. Wikarek, "A Hybrid Programming Framework for Modeling and Solving Constraint Satisfaction and Optimization Problems", in: *Scientific Programming*, vol. 2016, Article ID 5102616, 2016. doi:10.1155/2016/5102616.

[9] P. Sitek, P., J. Wikarek, "A Hybrid Approach to the Optimization of Multiechelon Systems", in: *Mathematical Problems in Engineering* vol. 2015, Article ID 925675, 2015. doi:10.1155/2015/925675.

[10] P. Sitek, "A hybrid approach to the two-echelon capacitated vehicle routing problem (2E-CVRP) *",* in: *Advances in Intelligent Systems and Computing,* 267**,** pp.251–263, 2014. doi:10.1007/978-3-319-05353-0_25.

[11] A. Schrijver, "Theory of Linear and Integer Programming*",* John Wiley & Sons, New York, NY, USA, 1998.

[12] Eclipse, 2015, Eclipse - The Eclipse Foundation open source community website, Accessed August 12, www.eclipse.org.

[13] G.M. Thompson, "Optimizing restaurant table configuration: Specifying combinable tables", in: *Cornell Hotel and Restaurant Administration Quartely,* 44, pp. 53–60, 2003.

[14] M. Milano, M. Wallace, "Integrating Operations Research in Constraint Programming", in: *Annals of Operations Research,* 175(1), pp. 37 – 76, 2010.

[15] T. Achterberg, T. Berthold, T. Koch, K. Wolter, "Constraint Integer Programming. A New Approach to Integrate CP and MIP", in: *Lecture Notes in Computer Science,* 5015, pp. 6-20, 2008.

[16] A. Bockmayr, T. Kasper, "Branch-and-Infer, A Framework for Combining CP and IP", in: *Constraint and Integer Programming Operations Research/Computer Science Interfaces Series*, 27, pp. 59-87, 2004.

[17] J. Wikarek, "A Novel Approach to Optimization of Jobs in Groups", in: *Progress in Automation, Robotics and Measuring Techniques,* pp. 313-322, 2015. doi:10.1007/978-3-319-15796-2_32.

[18] S. Bak, R. Czarnecki, S. Deniziak "Synthesis of Real-Time Cloud Applications for Internet of Things", in: *Turkish Journal of Electrical Engineering &Computer Sciences,* 2013. doi: 10.3906/elk-1302-178.

[19] M. Relich, A. Swic, A. Gola, "A Knowledge-Based Approach to Product Concept Screening", in: *Omatu, S. (eds.) Advances in Intelligent Systems and Computing*, vol. 373, pp. 341–348, 2015.

[20] K. Grzybowska, "Selected Activity Coordination Mechanisms in Complex Systems, Highlights of Practical Applications of Agents, Multi-Agent Systems, and Sustainability", in: *The PAAMS Collection Communications in Computer and Information Science*, Volume 524, J. Bajo et al. (Eds.), Springer International Publishing Switzerland, pp. 69-79, 2015. DOI: 10.1007/978-3-319-19033-4_6.

[21] P. Nielsen, I. Nielsen, K. Steger-Jensen, "Analyzing and evaluating product demand interdependencies", in: *Computers in Industry*, 61 (9), pp. 869-876, 2010. doi:10.1016/j.compind.2010.07.012.

[22] D. Krenczyk, J. Jagodzinski, "ERP, APS and Simulation Systems Integration to Support Production Planning and Scheduling", in: *Advances in Intelligent Systems and Computing,* Vol. 368, Springer International Publishing, pp 451-46, 2015.

[23] S. Deniziak, T. Michno, "Query by Shape for Image Retrieval from Multimedia Databases", in: *Communications in Computer and Information Science*, Springer, 521, pp. 377–386, 2015. doi: 10.1007/978-3-319-18422-7_33

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1223–1228
DOI: 10.15439/2016F300
ACSIS, Vol. 8. ISSN 2300-5963

Solving the *k*-Centre Problem as a method for supporting the Park and Ride facilities location decision

B. Prokop
Faculty of Computer Science
Warsaw School
of Information Technology
ul. Newelska 6
01-447 Warsaw, Poland
Email: prokop@wit.edu.pl

J. W. Owsiński, K. Sęp
Systems Research Institute
Polish Academy of Sciences
ul. Newelska 6
01-447 Warsaw, Poland
Email: {owsinski, sep}@ibspan.waw.pl

P. Sapiecha
The Faculty of Electronics
and Information Technology
Warsaw University of Technology
ul. Nowowiejska 15/19
00-665 Warsaw, Poland
Email: sapiecha@tele.pw.edu.pl

***Abstract*—In this article we analyze the problem of optimal location of transportation hubs in Warsaw, namely the Park and Ride location problem (*P&RP*). We take into account the expected travel time using public transport between particular points of the trip. In the currently existing *P&R* system we have 14 hub locations, and in this case the maximum travel time exceeds 50 minutes. The *P&R* problem can be reduced to the centers location problem (in our particular approach - the dominating set problem, *DS*), which is an *NP* hard problem. In order to determine the optimal locations for *P&R* two methods: the greedy and the tabu search algorithms were chosen and implemented. According to the computational experiments for the travel time restriction to 50 minutes, we obtain the *DS* composed of 3 hubs, in contrast to the existing 14 elements. The analysis of the *P&R* location in time domain is presented in this article in the context of further development of the Warsaw public transportation network, which seems to be interesting.**

I. Introduction

THIS article is devoted to the **Park & Ride** facilities location problem (*P&RP*) for the case of public transportation network of Warsaw [1], [9], [19]. Data used in the analysis of this problem were obtained from the official website of the **Public Transport Authority** in Warsaw [22].

In the previous approach to solve *P&RP* [18], the transport network graph was modeled as follows. The set of vertices represented the collection of bus stops and there was an edge between a pair of vertices if and only if there was a possibility of getting from one stop to another by bus, without any transfers. This model was, obviously, too simple and impractical: only one, and the real transportation mode was taken into consideration (buses), real travel times were irrelevant (e.g there existed some edges that represented travel times exceeding 90 minutes).

In this article we propose a much more precise model. The graph is modeled with application of real-world information about expected travel times between pairs of stops (including all modes of transit as well as transfers). This model takes also into consideration a rather common situation in which a pair of stops does not share any line but distance between them is very short (for example less then 50 meters). In the new model such vertices should be merged together. Using the **Open Trip Planner** (*OTP*) open-source software package the estimated travel time distances were computed.

Our research consists in applying the vertex domination methods in graphs to a real-life public transportation network.

II. Basic mathematical definitions

This section provides some basic notation, following [8], [6]. A **graph** is a representation of a set of objects, where some pairs of objects are connected by links. The interconnected objects are represented by mathematical abstractions called vertices, and the links that connect pairs of vertices are called edges. More formally, a **graph** is an ordered pair $G = (V, E)$ comprising a set V of **vertices** or **nodes** together with a set E of **edges**, which are 2-element subsets of V (E is a subset of $V \times V$). An **undirected** graph is the one in which edges have no orientation. The edge (a, b) is then identical to the edge (b, a). A vertex v is adjacent to u if and only if $(v, u) \in E$. Let $N(v) = \{u \in V : (v, u) \in E\}$ be an open **neighborhood** for a given vertex v.

A **dominating set** for a graph G is a subset D of V such that every vertex not in D is adjacent to at least one member of D [7], [11], [12]. This problem is strongly related to a problem well known in computational geometry, the **art gallery problem**. The domination number $\gamma(G)$ is the number of vertices in a smallest dominating set for G. The ***k*-dominating set problem** concerns testing whether $\gamma(G) = k$ for a given graph G and a natural number k; it is a classical ***NP*-complete** decision problem in computational complexity theory [15]. According to the theorem of Ore [5], if $G = (V, E)$ is a graph without isolated vertices, then the complement of a minimal dominating set of G is also a dominating set of G. This implies that every such graph has two disjoint dominating sets and hence, $\gamma(G) \leq \frac{1}{2} Card(V)$.

978-83-60810-90-3/$25.00

A **hypergraph** is a generalization of a graph in which an edge can connect any number of vertices [3]. Formally, a hypergraph *H* is a pair $H = (X, F)$, where *X* is a set of elements called vertices, and *F* is a set of non-empty subsets of *X* called **hyperedges**. Let *F* be a subset of $P(X) \setminus \{\emptyset\}$, where $P(X)$ is the power set of *X* and $F(x) = \{f \in F : x \in f\}$ for $x \in X$. A hypergraph is also called a **set system** or a family of sets drawn from the universal set *X*. The **rank** of hypergraph *H* is the size of the largest hyperedge in *H*. A **set covering** of a hypergraph $H = (X, F)$ is a subfamily *C* of *F*, such that the union of hyperedges from *C* equals the universe of vertices. A **transversal** (or **hitting set**) of a hypergraph $H = (X, F)$ is a subset *T* of *X* that has a nonempty intersection with every edge. The notions of hitting set and set covering are equivalent. The **decision versions** of the **hitting set** and **set covering** problems are ***NP*-complete**. The **greedy algorithm** for set covering chooses the sets according to one rule: at each stage, choose the hyperedge that contains the largest number of uncovered elements. This algorithm actually achieves an **approximation ratio** $\frac{Card(C)}{Card(Opt)}$ (*Opt* is an optimal set covering) of *h(rank)*, where $h(n)$ is the n^{th} harmonic number. This value is approximately given by: $\mathcal{O}((1 + \log(Card(V)))$. We can construct a dual algorithm for a hitting set problem, for which the performance ratio is: $\mathcal{O}((1 + \log(Card(F)))$.

Algorithm 1 The Greedy set covering method

input: hypergraph $H = (X, F)$;
output: set covering *C*;
$U := X$; $C := \emptyset$;

while($C \neq X$)**do** {
  select *S* from *F* such that
    maximizes $Card(S \cap U)$;
  $U := U \setminus S$;
  $C := C \cup \{S\}$;
  }
return: *C*;

It is interesting that there exists a pair of **polynomial-time reductions** between the **minimum dominating set problem** and the **minimum set covering problem**. These reductions show that an efficient algorithm for the minimum dominating set problem would provide an efficient algorithm for the set covering problem and vice versa. According to the above presented facts, the greedy algorithm provides a factor $1 + \log(Card(V))$ approximation of a minimum dominating set. Let us consider a reduction from the dominating set problem to the set covering problem. For any given graph $G = (V, E)$ with $V = \{1, 2, ..., n\}$, construct a hypergraph $H = (X, F)$ as follows: the universe *X* is *V*, and the family of hyperedges *F* is $\{F_1, F_2, ..., F_n\}$ such that F_v consists of the vertex *v* and all vertices adjacent to *v* in *G*. Hence, if *D* is a dominating set for *G*, then $S = \{S_v : v \in D\}$ is a feasible solution of the set covering problem, with $Card(C) = Card(D)$. Conversely, if $S = \{S_v : v \in D\}$ is a feasible solution of the set covering problem, then *D* is a dominating set for *G*, with $Card(D) = Card(C)$. Hence, the greedy algorithm provides a factor $1 + \log(Card(V))$ approximation of a minimum dominating set.

Problems of finding the best location of facilities in networks or graphs abound in practical situations, such as determining locations for factories, assembly plants or warehouses, as well as in airline crew scheduling. One of the well known facility location problems is the **vertex *k*-center problem**, where given *n* cities and distances between all pairs of cities, the aim is to choose *k* cities called **centers** so that the largest distance of any city to its nearest center is minimal. Let $G = (V, E)$ be a complete undirected graph with edge costs satisfying the triangle inequality (for a given **metric** $d : E \to R$), and *k* be a positive integer not greater than $Card(V)$. For any subset *S* of *V* and a vertex $v \in V$, define: $d(v, S)$ to be the length of the shortest edge from *v* to any vertex in *S*. The **vertex *k*-center problem** is to find such a subset *S* of *V*, where $Card(S) \leq k$, which minimizes: $max(d(v, S))$ for $v \in V$. The vertex *k*-center problem is *NP*-hard. In this paper we solve the *k*-center problem as a series of a minimum dominating set problems.

A set of different approaches (like: tabu search, variable neighborhood search) to solve the *k*-center problem was given by authors [17]. Parallely, the various greedy methods were proposed in the following publications [13], [14].

III. Problem definition

This article is devoted the *P&R* facilities location problem (*P&R*) in public transportation network of Warsaw. The problem is to find a set of stops such that in the worst case scenario using the public transport each trip will take no more than assumed *k* minutes. In this paper we propose the application of two methods: the *greedy algorithm* and the *Tabu Search*.

The *k*-Park&Ride Problem

Input: Given $G = (V, E)$—transport network,
  $d : E \to R^+$—an average distance
    between two vertices,
  $k \in R^+$ - time limit;
Output: Find $S \subseteq V$ such that:
  (*) for each $v \in V$: $\min_{s \in S} d(v, s) \leq k$,
  (**) $minCard(S)$.

IV. Our approach

In this chapter the data characteristics and the solution construction method are presented.

A. Data

The dataset that we are working on consists of coordinates of all public transportation stops in the city of Warsaw and complete schedules for different mass transit modes (buses, trams, metro and passenger trains). There are about **4000 stops** within the city limits and **317 different routes** each with 1-4 unique trips.

The data containing stops, routes and schedules is made available by the Public Transport Authority in Warsaw. However, the files are not shared in any standarized format and thus require parsing to be suitable for processing. Several software tools have been developed by the authors in order to work with this data (parsers, graph builders). They usually consist of over 4 million connections divided in multiple sections. The first step is to parse the file into *JSON* format. The second one is to convert it to the *GTFS* using our tools. *GTFS* is the required format to use *OTP* implementation of the **Raptor** algorithm [2] to find the shortest paths in multi-modal transportation network with a schedule. In our case, over 16 million requests had to be sent to instances of *OTP* servers in order to fill the adjacency matrix of the graph.

B. Methods

We model the network in a time oriented manner, using average travel times between each pair of stops as the weights of the edges in the **graph *G***. In order to achieve this, we have converted the data to *Google Transit Feed Specification* and used *Open Trip Planner* (*OTP*) to calculate the shortest trips between all stops in a given time lapse. The *OTP* application is an open-source tool for journey planning in multi-modal transit networks.

Given a complete directed graph G, in which the weight w of the edges represents the average travel time between the stops, we choose the parameter k measured in minutes as a cut-off value and construct a new **graph *H*** such that:

$$V(H) = V(G),$$
$$E(H) = \{e \mid e \in E(G),\, w(e) \leq k\}.$$

We then use H as an input to the *minimal Dominating Set* algorithms in order to approximate the smallest subset of stops that allows us to reach all other stops within the given time bounds. Information about the density of Warsaw's mass transit network with respect to different values of k is presented in Table 1.

TABLE I
DENSITY AND NUMBER OF EDGES WITH RESPECT TO PARAMETER k

k [min]	**Edges**	**Density**
15	679296	≈ 0.04
30	3863959	≈ 0.23
45	9121623	≈ 0.55
53	11916338	≈ 0.72

Our Approach to The *k*-Park&Ride Problem

1) Construct a complete directed graph *G=(V, E)*, where $d(v, u)$ is the average travel time for $(v, u) \in E$ (obtained from the application of the *OTP* package),
2) Construct a graph *H* based on graph *G* with respect to the time limit *k*,
3) Find a *minimal Dominating Set* for a given graph *H* using the Tabu Search or a greedy method.

Two *minimal Dominating Set* approximation methods were used in our computations. We applied the greedy algorithm and the metaheuristic tabu search method. Both approaches were implemented by us in *Python* programming language with extensive use of *NumPy* library for fast matrix operations.

V. RESULTS

We prepared computation experiments based on typical PC (Intel i5 2.7 Ghz 8GB Ram) using the followin software: OSX El Capitan 10.11.14 (clients and computations), Ubuntu 14.04 (server OTP), Python 3.5.0, iPython 4.0.0, NumPy 1.10.1, python-geojson, environment PyCharm, iPython Notebook, Vizualization: MapBox and Ruby on Rails,

We achieved the following results in terms of the time complexity:

1) The Greedy Set Covering in dependency of *k*:

k [min]	15	30	45	53
Comp.time	≈ 16.4	≈ 5.3	≈ 4.61	≈ 3.52

2) The OTP - 16 000 000 queries ($\approx$ 111 hours).
3) The Tabu Search in dependency of k (time for 100 iterations):

k [min]	15	30	45	53
Comp.time	≈ 4.1	≈ 4.2	≈ 4.5	≈ 5.0

The Park and Ride system in Warsaw consists currently of 14 parkings and they form a dominating set under our model when the time trip is $k \geq 53$ minutes. According to the computational results based on the new model these locations ensure that the longest journey from them to any stop will not exceed 53 minutes (they form a dominating set when $k \geq 53$). Under the same constraint (maximum 53 minutes travel time) we have found dominating sets of sizes 5 (greedy) and 3 (tabu search). It is interesting to note that when the maximum of travel time is limited to 30 minutes, the cardinality of the calculated dominating sets, as well as the locations of parkings are similar to those of the already existing facilities in Warsaw. This might suggest that small improvements/changes to the existing park and ride system might be very beneficial in terms of maximum journey distance (cutting it down from 53 to 30 minutes).

TABLE II
CARDINALITY OF DOMINATING SETS FOUND BY DIFFERENT METHODS

	***k* (min)**			
Algorithm	15	30	45	53
Greedy	69	15	8	5
Tabu search	62	12	5	3

However, we can achieve domination in this network by selecting only 3 nodes (see Figure 2).

VI. CONCLUSIONS

In this article we describe the analysis of the optimal location of transport hubs in Warsaw, in the context of the Park

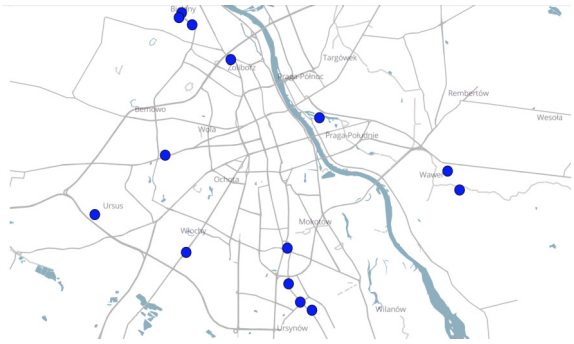


Fig. 1. Existing Park and Ride facilities

Fig. 2. Dominating set of cardinality 3 for $k = 53$ (tabu search)

Fig. 3. Dominating set of cardinality 12 for $k = 30$ (tabu search)

Fig. 4. Dominating set of cardinality 5 for $k = 53$ (greedy algorithm)

Fig. 5. Dominating set of cardinality 15 for $k = 30$ (greedy algorithm)

and Ride location problem. We took into account the expected travel time between two particular points. This problem was reduced to the problem of determining the centers in the sense of the dominating set. Data for our research were obtained and collected from the website of the public transport authority. In order to establish the format, the data was converted to *JSON* a format, and then to the *GTFS* one. Using the *OTP* package we computed the expected journey times between all pairs of stops in Warsaw public transport. In the currently existing *P&R* system we have 14 locations, and in this case the maximum travel time exceeds 50 minutes. In order to determine the optimal locations for *P&R* two methods: the greedy and the tabu search algorithms have been implemented (in *Python*). The result obtained imply the dominating set consisting of 3 stops, in the contrast to the existing 14 *P&R* parks. For the

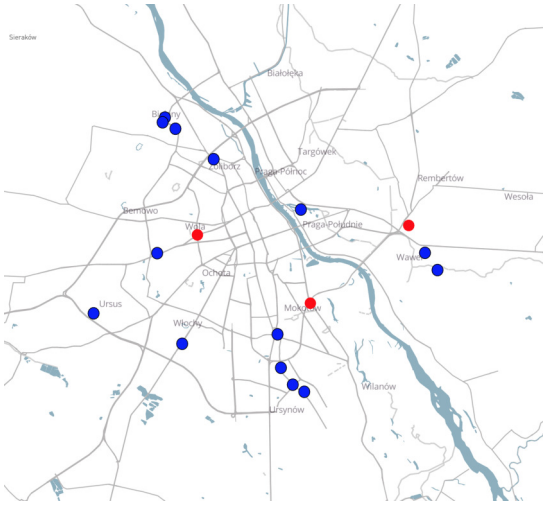


Fig. 6. Dominating set of cardinality 3 for $k = 53$ (red dots) and existing $P\&R$ (blue dots) (tabu search)

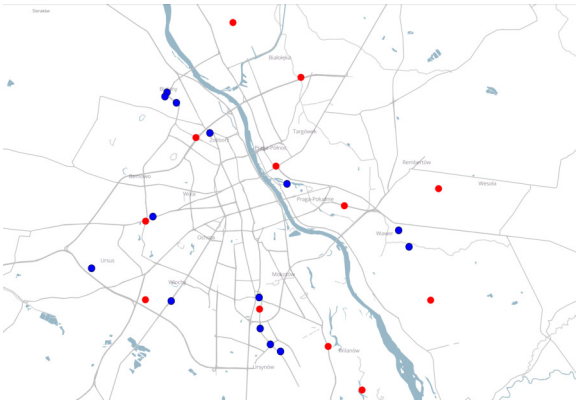


Fig. 7. Dominating set of cardinality 12 for $k = 30$ (red dots) and existing $P\&R$ (blue dots) (tabu search)

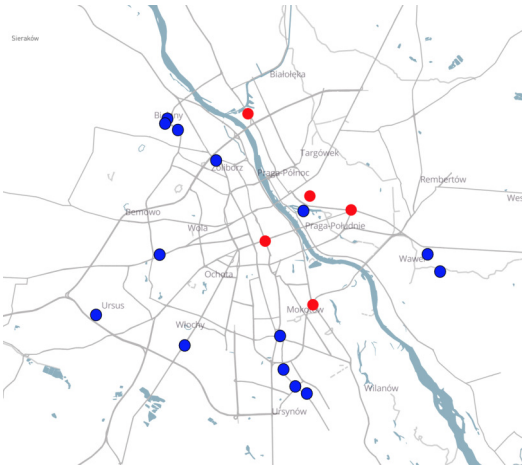


Fig. 8. Dominating set of cardinality 5 for $k = 53$ (red dots) and existing $P\&R$ (blue dots)(greedy algorithm)

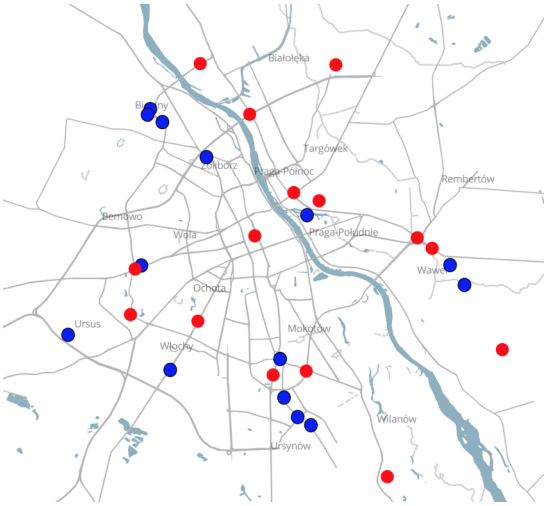


Fig. 9. Dominating set of cardinality 15 for $k = 30$ (red dots) and existing $P\&R$ (blue dots)(greedy algorithm)

assumed 30 minutes time limit, the result consists of 12 stops. This analysis seems to be interesting in the context of further development of the *P&R* system. To summarize, the aim of the *P&R* system is not only to discharge the traffic directed to the city center, but also to enable passengers to conveniently travel to different locations. Unfortunately, we can not say this about the existing solution.

REFERENCES

[1] Aros-Vera F., Marianov V., Mitchell J., *p-Hub approach for the optimal park-and-ride facility location problem*, European Journal of Operational Research 226, 2013

[2] Bast H., Delling D., Goldberg A., Müller-Hannemann M., Pajor T., Sanders P., Wagner D., Werneck R. F., *Route Planning in Transportation Networks*, Cornell University Library, arXiv:1504.05140, 2015

[3] Berge C., *Hypergraphs: Combinatorics of Finite Sets*, Elsevier, 254 pp., 1984

[4] Bunke H., Wang P. S. P., *Graph classification and clustering based on vector space embedding*, World Scientific Publishing Co. Pte. Ltd., Singapore, Series in Machine Perception and Artificial Intelligence, Vol. 77, 2010

[5] Chartrand G., Lesniak L., Zhank P. *Graphs & Digraphs*, CRC Press, 2010

[6] Cohen R., Havin S., *Complex Networks: Structure, Robustness and Function*, Cambridge University Press, Cambridge, 2010

[7] Ding-Zhu Du, Peng-Jun Wan, *Connected Dominating Set: Theory and Applications*, Springer Optimization and Its Applications, 2012

[8] Erciyes K., *Complex Networks: An Algorithmic Perspective*, CRC Press, 320 pp., 2014

[9] Farhan B., Murray A., *Siting Park and Ride facilities using a multi-objective spatial optimization model*, Computers & Operations Research 35, 2008

[10] Fortunato S., *Community detection in graphs*, Complex Networks and Systems Lagrange Laboratory, ISI Foundation, Viale S. Severo 65, 10133, Torino, I-ITALY, 2010

[11] Haynes T., Hedetniemi S., Slater P., *Fundamentals of Domination in Graphs*, Marcel Dekker Inc., 1998

[12] Henning M., Yeo A., *Total Domination in Graph*, Springer Monographs in Mathematics, 2013

[13] Hochbaum D. S., Shmoys D. B., *A best possible heuristic for thek-center problem*, Mathematics of Operations Research, 10:180-184, 1985.

[14] Hochbaum D. S., *Approximation Algorithms for NP-hard Problems*, PWS Publishing Company, 596 pp., 1995

[15] Garey M. R., Johnson D. S. *Computers and Intractability: A Guide to the Theory of NP-Completeness*, W.H. Freeman and Company, 344 pp., New York 1979
[16] Glover F., Laguna M., *Tabu search*, Springer Verlag, Handbook of Combinatorial Optimization, pp. 3261-3362, 2013
[17] Mladenovižc N., Labbe M., Hansen P., *Solving the p-center problem with tabu search and variable neighborhood search*, Networks 42, 1: 48-64, 2003.
[18] Owsiński J.W., Stańczak J., Barski A., Sęp K., Sapiecha P., *Graph based approach to the minimum hub problem in transportation network*, porcedings of FEDCSiS 2015, pp. 1641 1648, 2015
[19] Schöbel A., *Optimization in Public Transportation*, Springer Verlag, Optimization and Its Applications, 2006
[20] Takes F. W., *Algorithms for Analyzing and Mining Real-World Graphs*, PhD thesis was employed at Leiden University, 2014
[21] Vazirani V. V., *Approximation Algorithms*, Springer Verlag, Berlin Heidelberg, 2003
[22] The source of data, *Warsaw public transport timetable*, http://www.ztm.waw.pl, 2016

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1229–1233
DOI: 10.15439/2016F460
ACSIS, Vol. 8. ISSN 2300-5963

Gamification in Enterprise Information Systems: What, Why and How

Jakub Swacha
University of Szczecin, Faculty of Economics and Management
Institute of Information Technology in Management
ul. Mickiewicza 64, 71-101 Szczecin, Poland
Email: jakubs@uoo.univ.szczecin.pl

***Abstract*—With gamification, design elements known from games can be used to increase employees' engagement and improve users' experience. This paper points to Enterprise Information Systems as a viable point of implementation of gamification. There are three relevant questions: what gamification actually consists of, why it is worthwhile to apply it in Enterprise Information Systems, and how to implement it properly. The paper aims to answer them.**

I. Introduction

FOURTEEN years have passed since Nick Pelling coined the term of *gamification* [1, p. 5]. The "use of game elements to increase engagement and make life and work more fun" (as it was neatly defined by Mark Schreiber [2]), despite the criticism (see e.g. [3, p. 18] and works cited therein), has already found its way into such diverse areas as, among others, marketing [4] and knowledge management [5], banking [6] and education [7], production environments [8] and tourism [9], and even scientific research [10].

This paper investigates the application of gamification in Enterprise Information Systems (EIS), understood as "software systems that integrate the business processes of organizations to improve their functioning" [11]. Gamification does not aim at redefining the business processes, but on affecting how they are experienced [12]. The underlying software plays an important role as it allows to automate tracking participants' actions, register their achievements, and pass relevant feedback [13].

The paper aims to provide answers to the three basic questions: what actually forms the applied gamification, what can be the reasons for making use of it in Enterprise Information Systems, and how to implement it there properly. The structure of the paper has been modeled appropriately into three sections, ending with the conclusion.

II. What the applied gamification consists of

Although a definition of gamification has already been provided, even the reading of thirty definitions gathered by Andrzej Marczewski [2] does not tell what a gamified system actually consists of. In order to explain that, it is necessary to look at the actual components of gamification.

One of the most widely known lists of gamification elements is the one devised by Kevin Werbach and Dan Hunter [14]. They classify them into dynamics ("the big-picture aspects of the gamified system"), mechanics ("the basic processes that drive the action forward and generate player engagement"), and components (the "more-specific forms that mechanics or dynamics can take").

The dynamics include:

- *constraints* – limitations or forced trade-offs,
- *emotions* driving players, such as: curiosity, competitiveness, frustration, happiness etc.,
- *narrative* – the storyline of the game,
- *progression*, measuring players' development,
- *relationships* – players' social interactions.

The mechanics include:

- *challenges* – tasks that require effort to solve,
- *chance*, bringing in randomness to the game,
- *competition* between players or groups of them,
- *cooperation*, requiring players to work together to achieve a shared goal,
- *feedback* providing players with information about how they are doing,
- *resource acquisition* – allowing players to gather useful or collectible items,
- *rewards* given for some action or achievement,
- *transactions* – allowing item trading between players,
- *turns* – sequential participation by players,
- *win/loss/draw states* – objectives that make one player/team the winner and the others the losers.

What Werbach and Hunter call components includes:

- *achievements* – defined objectives,
- *avatars* – player's character visual representation,
- *badges* – visual representations of achievements,
- *boss fights* – rare, extremely hard challenges,
- *collections* – sets of items/badges to accumulate,
- *combat* – a defined battle, typically short-lived,
- *content unlocking* – additional game content available after players reach certain objectives,

The publication was financed from the funds of the Department of Engineering of Information Systems at the Faculty of Economics and Management of University of Szczecin for maintaining research potential.

- *gifting* – players' ability to share their resources,
- *leaderboards* that visually present player's achievements in relation to other players,
- *levels* that define steps in players' progression,
- *points* which measure players' progression,
- *quests* – sets of tasks with objectives and rewards,
- *social graphs* that represent players' network of contacts within the game,
- *teams* – groups of cooperating players,
- *virtual goods*, which have perceived value.

A much more comprehensive list of gamification components can be found in the Octalysis framework developed by Yu-Kai Chou [15]. He clusters them into eight categories, labeled as "core drives" of gamification.

The first drive is epic meaning & calling. Its components help players justify devoting their time to the game. They are supposed to make players believe that they are doing something great (*narrative*, *higher meaning*) or were "chosen" to do something (*elitism*, *humanity hero*, *destiny child*), or that they have got some gift that others have not (*beginners luck*, *free lunch*). Players may also feel attached to the game world elements they created (*co-creator*).

The development & accomplishment components exploit the players' internal drive to make progress in absolute (*points*, *progress bar*) or relative terms (*leaderboards*), acquire (*step-by-step tutorial*) and develop skills (with *badges* as achievement symbols), overcome challenges (chosen from *quest list*, or the final *boss fights*), and receive due appreciation for it (*fixed action rewards*, *win prize*, *high-five*, *crowning*, *level-up symphony*, *aura effect*).

The empowerment of creativity & feedback components give players the decisive power (*general's carrot*) and engage them in a creative process, where they have to repeatedly figure things out (*evergreen mechanics*, *blank fills*) and try different combinations (*real-time control*, *chain controls*) possibly using approaches unavailable earlier (*milestone unlock*, *boosters*) after receiving *instant feedback* or hints (*choice reception*). They may also have a chance to opt out of a given challenge (*voluntary autonomy*).

The ownership & possession components give players the feeling of owning something (*virtual good*, *avatar*), so that they want to make what they own better (*build from scratch*) while progressing in the game (*learning curve*) and own even more (*earned lunch*, *collection set*). The players' attachment to their belongings can be augmented by letting them constantly observe what they have (*monitoring*) and have them protect it from dangers (*protection*). The feeling of possession can also be extended to other players one has invited to the game (*recruitment*).

The social influence & relatedness components refer to activities inspired by what other people think, do, or say as well as the desire to draw closer to people, places, or events that players can relate to. It includes making personal relations publicly observable (*friending*), having veteran players as guides (*mentorship*, which also makes the beginner players more attached to the specific culture, as well as helps veteran players stay engaged in the game), cooperating with other players to solve difficult challenges (*group quest*), showing other players one's accomplishments, either explicitly (*bragging*) or implicitly (*touting*), having a place to chat about a variety of topics (*water cooler*), gifts or rewards that players can only receive from other players (*social treasure*), encouraging players to generously give, expecting the recipients to give back somehow (*thank-you economy*), and making social interactions technically very easy to perform (*social prod*).

The scarcity & impatience components address the human tendency to want things they cannot have. They include constantly showing an item that a player cannot easily get (*dangling*), having something accessible only at specified time (*appointment dynamics*), or in *fixed intervals*, or after meeting specified conditions (*moats*), showing time after which something becomes inaccessible (*countdown*), changing the pace at which progress can be made (*throttles*), and requiring players to collect multiple pieces to earn the actual reward (*prize pacing*).

The unpredictability & curiosity components exploit the human infatuation with experiences that are uncertain and involve chance and the natural curiosity to explore. They include surprises (*Easter eggs*), also in form of unexpected rewards (*sudden rewards*), *oracle effect* that makes player expect an event to happen in the future (and wonder whether it will actually happen), quests within quests (*mini quests*), *glowing choice* which leads players in the right direction by appealing to their curiosity, lotteries that some player has to win each period, with actions available that increase one's chance (*rolling rewards*), *random rewards* that recreate the excitement that children have opening gift boxes, and playing small pranks on players (*mischief*).

The loss & avoidance components motivate players with a threat of losing something they have attained. They include *sunk-cost tragedy* (players continue the game, because they spent a lot of time playing so far), *progress loss* (if players stop playing, they lose what they earned), *fear of missing out* (players are aware that when they do not participate, things happen that could benefit them), *evanescence opportunity* (which will disappear if a player does not perform certain action), *scarlet letter* (a shame of not having something all the good players have), *status quo sloth* (wanting to continue the game with the same behavior), and marking the loss in a special way (*weep tune* and *visual grave*).

III. Why apply gamification to EIS

There are various reasons given why gamification can be helpful for an enterprise. The primary argument is the link between games and intrinsic motivation, which can be exploited with gamification. All the seven main intrinsic motivators identified by Thomas W. Malone and Mark R. Lepper (i.e., challenge, curiosity, control, fantasy, cooperation, competition and recognition [16, p. 230 and

242]) can be effectively addressed in a gamified system (see section II). And that the lack of motivation is a serious issue in enterprises can be seen, e.g., from the results of the long-term Gallup poll showing that 70% of American workers are not engaged at work [17].

There are also other reasons provided in the literature (see e.g. [18] and works cited therein). One is the lack of goal prioritization making employees overladen with both present activities and development opportunities losing interest in taking actions that are not needed at the moment, but will be crucial in the future. A gamified system can both rank the possible actions in terms of their relative value to the enterprise, and provide a path of development to follow.

The second reason is the coming of the new generation of employees: Generation Z, *digital natives*, who have different expectations from work and communication habits than previous generations. A gamified system can use the type of communication they are familiar with, and work as a bridge between them and the older employees.

Another one is the omnipresence of stress in many corporate environments, which leads to lower productivity, problems with interpersonal communication, and even physical and psychic health issues. A well-designed gamified system can both address some causes of stress (with its informal communication and a clear system of priorities) and aid in stress relief (with relaxing side activities and mood-improving feedback it offers).

Regardless of what the reason is, there are numerous examples of successful enterprise gamification [19].

Moving on to the details of gamification in Enterprise Information Systems, Table I lists its possible uses grouped in four categories, consisting of those related to the improvement of: work performance, work attitude, social relations, and on-boarding and training processes. Note that the list is not exhaustive, as it is up to the creativity of a designer to find a combination of gamification components that would address particular goals that the enterprise management may consider important.

TABLE I.
POSSIBLE USE OF GAMIFICATION IN EIS

Practice	Expected benefit	Relevant component
Performance		
Bundle tasks and split rewards	Users are motivated to work consistently, and complete related tasks together	Collection set, Prize pacing
Define rush hours when productivity is most needed	Users should increase their effort at the right time	Appointment dynamics, Fixed intervals
Differentiate the rewards for completing various tasks	Users are directed to the most important tasks at a given moment	Fixed action rewards, Virtual good
Mark very hard tasks	Users can prepare better for a big challenge	Boss fights
Reward what every user should do	Users are motivated to catch up with others	Scarlet letter
Visualize relative employees' performance	Users are motivated to rise over others	Leaderboards
Visualize the distance to the goal	Users are motivated to finish the current task	Progression bar
Visualize the time left	Users are motivated to hasten their work	Countdown
Work attitude		
Announce crucial events that will come later	Users are curious of what will actually happen	Oracle effect
Define penalties for failing to complete a task	Users appreciate what they have attained and could lose	Points, Protection, Progress loss
Leave users some degree of freedom	Users feel they have control of what they do	General's carrot, Voluntary autonomy
Let users improve or customize the system	Users feel attached to what they created	Co-create
Provide a chance of surprise	The monotony of repetitive tasks is shunned	Easter eggs, Random rewards, Mischief
Provide a chance to shine for everyone	Even lagging users can have their moment of glory	Rolling rewards
Remind users the importance of their role	Users are aware of the value of their contribution	Elitism, Humanity hero
Remind users the enterprise mission	Users feel they are part of something big and good	Higher meaning
Reward completion even of the simplest tasks	Users feel their effort is recognized	Fixed action rewards, Virtual good, High-five

TABLE I (CONTINUED).
POSSIBLE USE OF GAMIFICATION IN EIS

Practice	Expected benefit	Relevant component
Visualize how the results of repetitive tasks accumulate	Users comprehend the magnitude of their everyday work	Points, Progression bar
Visualize possible big rewards	Users are aware of what they can gain if they stay engaged	Dangling
Visualize the progress a user has made so far	Users comprehend the progress they made which should increase their self-esteem	Levels, Badges, Monitoring
Social relations		
Let employees collaborate on tasks	Arduous tasks can be finished on time and knowledge is transferred from the more to the less experienced users	Group quest
Let users boast their accomplishments	Users can share their gladness with others	Bragging, Touting
Let employees visualize their relations	Social relations are improved	Friending
Let users discuss freely	Users get to know each other, share experiences and ideas	Water cooler
Let users help each other	Knowledge is transferred from the more to the less experienced users and their social relations are improved	Thank-you economy
Let users reward each other	Users feel their effort is recognized and their social relations are improved	Social treasure, Virtual good
Make it fast and easy to interact with others	Social interactions do not hurt productivity much	Social prod
On-boarding and training		
Appoint more difficult tasks to users as they make progress	Users do not get bored or frustrated with tasks that are too easy or too hard for them	Milestone unlock, Learning curve
Celebrate passing important steps of development	Users feel the progress they make matters	Level-up symphony, Aura effect
Communicate action results immediately	Users know when they do well	Instant feedback
Guide users in their steps	Users know what to do to progress	Step-by-step tutorial, Choice reception
Introduce users to the organizational culture	Users adapt faster to the organization	Narrative
Let experienced users guide new ones	Knowledge is transferred from experienced to novice users and social relations are created	Mentorship
Make the first steps easier for beginner users	Users get a positive first impression	Beginners luck, Free lunch, Boosters
Make users improve their weak sides before they can move on	Users' skill development is more balanced	Moats

IV. HOW TO PROPERLY IMPLEMENT GAMIFICATION IN EIS

The successful implementation of gamification in EIS is a matter of primary importance, as a failed attempt will bring costs in morale and productivity, notwithstanding the cost of the implementation itself.

It is therefore strongly advised to carry it out in a carefully planned manner. The first thing is to follow a proven procedure, such as the player experience design process proposed by Brian Burke [1], which includes seven steps respectively devoted to: (1) business outcomes and success metrics, (2) target audience, (3) player goals, (4) engagement model, (5) play space and journey, (6) game economy, and (7) repetitive playing and testing (highly possibly leading back to one of the previous steps).

Mario Herger convincingly argues that gamification should be focused on value creation for the players, and only through it the value for the organization should be created [20]. While the player's value creation does not automatically translate into an organization's value creation, but also the value created for an organization may be larger than the value for an individual player. He also lists three types of reactions that should be achieved with gamification: Aaah-effect (the act of delight), Aha-effect (the act of revelation) and the Haha-effect (the act of amusement).

Fabian Groh provides eleven design principles for implementing gamification [21]: (1) connect to personal goals, (2) connect to a meaningful community of interest, (3) create a meaningful story, (4) beware of social context

meanings, (5) provide interesting challenges, (6) provide clear, visual, varying, and well-structured goals, (7) provide juicy feedback, (8) beware of unintended behaviors, (9) play is voluntary, (10) beware of losing autonomy, and (11) beware of devaluating activities.

Ethan Mollick and Nancy Rothbard highlight the role of consent, understood as the active cooperation of workers with managerial goals, owing to the fact that gamification is not driven organically by employees, but instead imposed from the top by managers [22, p. 14]. They therefore point to the importance of three indicators of consent, i.e. clearly understanding the rules of the game, perceived sense of justice and fairness, and active engagement.

The last aspect that needs to be addressed is the technology of implementation. The gamification subsystem can be developed as a module of an EIS or a separate gamification software that has to be integrated with the main system. A generic platform for enterprise gamification, based on service-oriented and event-driven principles, as well as best practices, and targeted for both modern and legacy systems, was proposed by Philipp Herzig *et al*. [23].

V. Conclusion

Gamification is able to make the employees' experience of performing tedious and repetitive tasks more enjoyable, rise their engagement, improve their attitude towards work, and, consequently, increase their productivity.

The rich portfolio of successful gamification projects makes it something too promising to ignore for enterprises in the world of scarce opportunities for gaining competitive advantage (*cf*. the *scarlet letter* gamification component).

The Enterprise Information Systems, due to their prevalence in today's organizations, are viable points of implementation of gamification spanning over different business processes and addressing various types of employees' activity.

This paper provided answers what the gamification amounts to in practice (with a catalog of gamification components in section II), why it can be useful to implement it in an EIS (with explanations for why it can work and a list of possible uses in section III), and how to do it right (with implementation guidelines in section IV).

References

[1] B. Burke, *Gamify: how gamification motivates people to do extraordinary things*. Brookline: Gartner, 2014.

[2] A. Marczewski, *Defining gamification – what do people really think?*, 2014, http://www.gamified.uk/2014/04/16/defining-gamification-people-really-think. Retrieved 5 May 2016.

[3] K. Seaborn, D.I. Fels, „Gamification in theory and action: A survey," *International Journal of Human-Computer Studies*, vol. 74, pp. 14–31, 2015, http://dx.doi.org/10.1016/j.ijhcs.2014.09.006.

[4] B. Borowski, *Gamification – Engage customers in your business: the hottest marketing trend in 2014*. Luxembourg: CreateSpace, 2014.

[5] J. Swacha, "Gamification in Knowledge Management: Motivating for Knowledge Sharing," *Polish Journal of Management Studies*, vol. 12, no. 2, pp. 150–160, 2015.

[6] L.F. Rodrigues, C.J. Costa, A. Oliveira, "The adoption of gamification in e-banking," in *Proceedings of the 2013 International Conference on Information Systems and Design of Communication*. New York: ACM, 2013, pp. 47–55, http://dx.doi.org/10.1145/2503859.2503867.

[7] K.M. Kapp, *The gamification of learning and instruction: Game-based methods and strategies for training and education*. San Francisco: Pfeiffer, 2012.

[8] O. Korn, M. Funk, A. Schmidt, "Design approaches for the gamification of production environments: a study focusing on acceptance," in *Proceedings of the 8th ACM International Conference on PErvasive Technologies Related to Assistive Environments*. ACM, New York, 2015, http://dx.doi.org/10.1145/2769493.2769549.

[9] A. Negruşa, V. Toader, A. Sofică, M. Tutunea, R. Rus, "Exploring gamification techniques and applications for sustainable tourism," *Sustainability*, vol. 7, no. 8, pp. 11160–11189, 2015, http://doi.org/10.3390/su70811160.

[10] F. Khatib, F. DiMaio, Foldit Contenders Group, Foldit Void Crushers Group, S. Cooper, M. Kazmierczyk, M. Gilski, Sz. Krzywda, H. Zabranska, I. Pichova, J. Thompson, Z. Popović, M. Jaskolski, D. Baker, "Crystal structure of a monomeric retroviral protease solved by protein folding game players," *Nature Structural & Molecular Biology*, vol. 18, no. 10, pp. 1175–1177, 2011.

[11] M. Tabatabaie, R. Paige, C. Kimble, „Exploring Enterprise Information Systems," in M. Khosrow-Pour (ed.), *Enterprise Information Systems: Concepts, methodologies, tools and applications*. Hershey: IGI Global, 2011, pp. 35–52.

[12] N. Teh, D. Schuff, S. Johnson, D. Geddes, "Can work be fun? Improving task motivation and help-seeking through game mechanics," in R. Baskerville, M. Chau (eds.), *Proceedings of the International Conference on Information Systems*. Milan: AIS, 2013.

[13] J. Swacha, "An architecture of a gamified learning management system," in Y. Cao, T. Väljataga, J.K.T. Tang, H. Leung, M. Laanpere (eds.), *New horizons in web based learning*, Cham: Springer, 2014, pp. 195–203, http://dx.doi.org/10.1007/978-3-319-04954-0_24.

[14] K. Werbach, D. Hunter, *For the win: How game thinking can revolutionize your business*. Philadelphia: Wharton Digital Press, 2012.

[15] Y.-K. Chou, *Actionable gamification: Beyond points, badges, and leaderboards*. Fremont: Octalysis Media, 2016.

[16] T.W. Malone, M. R. Lepper, "Making learning fun: A taxonomy of intrinsic motivations for learning," in R. Snow, M.J. Farr (eds.), *Aptitude, learning, and instruction. Volume 3: Conative and affective process analyses*. Hillsdale: Lawrence Erlbaum, 1987, pp. 223–253.

[17] Gallup, *State of the American Workplace*, http://www.gallup.com/services/178514/state-american-workplace.aspx, 2013. Retrieved 5 May 2016.

[18] K. Gruszecka, "Management through entertainment," in J. Swacha, K. Muszyńska (eds.), *Information Technology Meets Management in Knowledge Economy*. Warsaw: Polish Information Processing Society, 2015, pp. 139–158.

[19] Y.-K., Chou, *A Comprehensive List of 90+ Gamification Cases with ROI Stats*, 2015, http://yukaichou.com/gamification-examples/gamification-stats-figures/#.VzCedYSLSUk. Retrieved 5 May 2016.

[20] M. Herger, *Enterprise Gamification. Engaging People by Letting Them Have Fun. Book 1. The Basics*. Luxembourg: CreateSpace, 2014.

[21] F. Groh, "Gamification: State of the art definition and utilization", in N. Asaj, B. Könings, M. Poguntke, F. Schaub, M. Weber, B. Wiederscheim (eds.), *Proceedings of the 4th Seminar on Research Trends in Media Informatics*. Ulm: Institute of Media Informatics, 2012, pp. 39–46.

[22] E.R. Mollick, N. Rothbard, "Mandatory Fun: Consent, Gamification and the Impact of Games at Work," *The Wharton School Research Paper Series*, 2014, http://dx.doi.org/10.2139/ssrn.2277103.

[23] P. Herzig, M. Ameling, A. Schill, „A Generic Platform for Enterprise Gamification," in *2012 Joint Working IEEE/IFIP Conference on Software Architecture and European Conference on Software Architecture*. Helsinki: IEEE, 2012, pp. 219–223.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1235–1239
DOI: 10.15439/2016F489
ACSIS, Vol. 8. ISSN 2300-5963

MCDA-based Decision Support System for Sustainable Management – RES Case Study

Jarosław Wątróbski
West Pomeranian University of Technology in Szczecin,
Żołnierska 49, 71-210 Szczecin, Poland
Email: jwatrobski@wi.zut.edu.pl

Paweł Ziemba
The Jacob of Paradyż University of Applied Sciences in Gorzów Wielkopolski,
Chopina 52, 66-400 Gorzów Wielkopolski, Poland
Email: pziemba@pwsz.pl

Waldemar Wolski
University of Szczecin,
Mickiewicza 64, 71-101 Szczecin, Poland
Email: wwolski@wneiz.pl

***Abstract*—The MCDA methods are used in order to solve complex decision-making problems which require considering many interrelated criteria. They are also the basis of DSS. Nevertheless, these dependencies between criteria can have an influence on the obtained solution. A class of decision-making problems, in which there are intercriteria dependencies, are decisions in the sustainability area, e.g. selection of a location and a design of an RES-based power station. The article presents a complex model, taking into consideration dependencies between criteria.**

I. Introduction

ONE of the greatest challenges of energy-saving in Poland and other countries of the world is its adaptation to the demands of low carbon economy characterized mostly by the use of renewable energy sources (RES) [1].

The Polish Energy Law Act, which is a source of renewable energy, defines the following: wind energy, solar energy, geothermal energy, sea wave and tidal energy, river fall energy, biomass energy, energy from landfilled biogas and biogas produced in the process of sewage disposal and treatment or decomposition of plants and animal remains [4]. Among the above-mentioned RES, the greatest potential for energy production, which can be found in Poland and the EU have wind farms [2], [3].

The selection of a location [5] and of a project design [6] will lead to the successful implementation of a wind farm project. These choices determine the efficiency of wind power plants and also have an effect on the environment, benefits and costs [7]. The problems of location selection and project design selection, as well as other decision factors related to RES management, are multi-criteria decision-making problems which require the examination of many contradictory and mutually correlated criteria which encompass technological, economic, environmental and social issues [5], [6], [8], [9]. Decision-making methods which consist of a sole criterion are unable to cope efficiently with such decision problems [10]. While solving such problems, multi-criteria decision analysis (MCDA) methods can be applied, as they can handle complex decision processes, multiple and conflicting evaluation criteria, different scenarios, preferences of decision-makers, several sources of uncertainty and specific time frames [11] [12]. Many Decision Support Systems (DSS) are based on MCDA methods and algorithms and used for solving environmental problems and those relating to the power industry [13]. DSS provide knowledge indispensable for making decisions and maximize the results of processes of decisions, by lifting cognitive, special and economic restrictions of the decision-maker [14].

The aim of this article is to establish a multi-criteria decision model, based on MCDA methods, which solve a decision problem comprising of the selection of a location and a project design of an onshore wind farm. The model should not bypass the complexity of the decision problem, but should take into account mutual dependencies between criteria and the influence of some criteria on other ones. This model could become a DSS engine for RES management, while paying particular attention to wind farms.

Section II displays the analysis of the explanations on decision support related to the design and construction of wind farms. The pre-selection of evaluation criteria of wind farm location and design was made according to the implementation of the analysis. The evaluation criteria were incorporated into a decision model, which was presented in Section III. Section IV contains a summary of research results, and further research directions are also pointed out.

II. Literature review

Publications regarding the decision for support in wind energy mostly include the construction of decision models as well as that of DSS and GIS (Geographical Information System) systems.

An example of constructing a decision model for the selection of a wind farm location is [15]. In this paper, evaluation criteria and their importance were presented in a decision model for the sake of selecting a wind farm location. On the other hand, in [16] a decision model was devised in order to compare different RES technologies (wind energy received the highest rank), and it subsequently modified in order to select a location of an onshore wind farm. The issue of devising a decision model for the

selection of an onshore wind farm location is also dealt with in [17], and an offshore one received a similar treatment in [5]. GIS decision systems were suggested, amongst other things in [18], [19], [20], [21]. These systems evaluate the potential of onshore areas regarding the situation of wind farms nearby. Similarly, in [22] a GIS was presented; itallowed the evaluation of a location of hybrid power stations based on wind and solar energies. In [23], a GIS-based DSS system analyzing the potential of onshore wind farm locations was discussed. The problem of the construction of a DSS for selecting an offshore wind location was attempted in [24], [7]. As far as decision problems relating to the project design of a wind farms are concerned, these were discussed in [6], [25]. The technical aspects of wind turbines are also related to farm designs[8] as well as wind farm development evaluation, in the bigger picture [26].

The AHP (Analytic Hierarchy Process) [27] method is used in selecting a location or a design of a wind farm, both in its crisp and fuzzy [28] [29] versions. It is primarily employed to determine the importance of criteria. A generalization of AHP, namely ANP (Analytic Network Process) [30], and different variants of the ELECTRE method are rarely used. Other MCDA methods, such as DEMATEL, OWA, SAW, PROMETHEE, NAIADE, TOPSIS, VIKOR, Lexicographic method and the conjunctive method, etc are also incidentally used. [31]. It should be noted that, in order to solve decision-making problems related to wind energy, decision models are used, and these are characterized by various complexities. The amount of criteria considered while selecting a location or a design of a wind firm ranges from 6 [19] to 35 [6]. These criteria are often interrelated and interdependent. For example, in this publication [6], the following criteria were used: generating cost, generating profit, and payback period. It can be easily seen that the payback period results from, among other things, a calculation of costs and profits.

III. The Proposed Decision Model for DSS

The construction of the decision model was done in the following manner:

1) The preparation of a set of criteria and sub-criteria for the evaluation of locations and designs of onshore wind farms,
2) The analysis of sub-criteria and indication of relationships and dependencies which occur place between them,
3) The presentation of sub-criteria dependencies in a networking decision model.

On the basis of the literature analysis presented in Section II, a set of criteria for the evaluations of locations and designs of onshore wind farms was prepared and displayed in Table I. The sub-criteria section was shown as well.

Next, the sub-criteria were analyzed and their interdependencies were explained.

C1.1The wind conditions influence the output power which is obtained from a wind turbine, at a specific wind speed and consequently, the desired amount of energy is generated. Generally, a stronger wind generates more energy; however, the wind should not be too strong, as most wind turbines switch off when the wind reaches the speed of about 25-30 m/s [8]. The wind speed is essential, depending on the height at ground level. Most towers which have turbines mounted onto them are 50-100m high [32], therefore the wind speed at the height of ca. 100m is crucial.

TABLE I.
CRITERIA AND SUB-CRITERIA FOR EVALUATING LOCATIONS AND DESIGNS OF ONSHORE WIND FARMS

Criteria		Sub-criteria		References
C1	Technical	C1.1	Average wind speed at the height of 100m	[5], [6], [15], [17], [19]-[21], [23], [24], [26]
		C1.2	Output power of wind turbine	[8]
		C1.3	Power grid voltage on the site of connection	[26]
C2	Economic	C2.1	Yearly amount of energy generated	[6], [7], [25]
		C2.2	Investment cost	[5]-[7], [16], [25]
		C2.3	Operational costs per year	[5], [6], [16], [25]
		C2.4	Incomes from generated energy per year	[26]
		C2.5	Profits from generated energy per year	[5], [6], [24]
		C2.6	Payback period	[5]-[7]
C3	Social	C3.1	Number of generated workplaces	[5], [6], [15]
		C3.2	Social acceptance	[16], [25]
C4	Spatial and environmental	C4.1	Distance from power grid connection	[20]-[24]
		C4.2	Distance from the road network	[15], [18]-[23]
		C4.3	Location in Natura 2000 protected area	[7], [19], [22], [23]

C1.2 The maximum output power of a turbine is achieved at a specific wind speed, which is critical. If the wind speed is lower, then the output power is equally lower.

C1.3 Voltages of the national power grid, which are used in Poland, amount to 110kV, 220kV, 400kV and 750kV [33]. The voltage of a power grid is an essential sub-criterion, since changes in the wind speed cause frequency, voltage and power fluctuations in the grid connected to a wind turbine [34]. Such fluctuations may, in turn, cause damage to transmission lines of the grid or transformers. Such a hazard is more likely in the case of high-voltage wind farms (from several dozens of MW) to a low-voltage grid (110kV) [35].

C2.1 We have mentioned above that the amount of energy generated is directly affected by the output power of wind turbines installed in the wind farm. The yearly amount of energy generated by a wind farm is presented, in its simplest form, in the formula (1):

$$E = \sum_i W_{out}(t_i) * 8760 \qquad (1)$$

where: E – yearly amount of energy generated [MWh], $W_{out}(t_i)$ – output power of an i-th turbine [AMW], 8760 – the number of hours in a year [Ah].

C2.2 The overall investment costs consist mostly of the costs of purchasing and installing turbines, towers and their foundations, the costs of preparing a wind farm design, as well as the costs of connecting a wind farm to the power grid [36]. The amount of capital investment for an onshore wind farm in Poland, depending on the technology applied, varies from 4.5 million to 7.5 million PLN/MW [37].

C2.3 The operational costs also include the operation, repairs and servicing of the devices, leasing costs of the land, management, insurance, taxes and charges, the energy consumption of the wind farm, as well as balance costs. It is estimated that the operation costs of a wind farm in Poland in 2011, in total, amounted to 83 PLN per one MWh of the energy generated by the wind farm [36].

C2.4 The income from the energy production is a product of generated (and sold) amount of energy and its price. An average sales price of the electric energy on the competitive market in Poland in the third quarter of 2015 amounted to ca. 173 PLN/MWh [38]. However, a new law defining RES auctions has been recently introduced. RES-based power stations can participate in such auctions. When an auction is won, the power station has the certainty that the Polish state will purchase from them the energy at a set price for a period of 15 years [39]. The reference price of the onshore wind energy for an RES auction, which is generated in a wind farm of combined power greater than 1MW in 2016, amounts to 385PLN/MWh [40].

C2.5 In simple terms, a yearly profit from selling the energy can be determined as being the difference between incomes from the energy sales and operational costs incurred to generate the energy sold.

C2.6 The payback period determines a period of time after which capital expenditure incurred through the construction of a wind farm will pay for itself. The payback period is, in fact, a ratio of investment costs to a yearly profit generated by the wind farm.

C3.1 The construction of new wind farms and maintenance of existing ones generate workplaces related to the preparation of the investment, its operation, maintenance, repairs and equipment servicing. Estimates for Poland suggest that the installation of 10MW in a given year generates 39 direct and 75 indirect workplaces. The maintenance of 10MW, in turn, is connected with the employment of 5 workers in subsequent years [41].

C3.2 Social acceptance refers to benefits, threats and inconveniences for a local community. Research results highlight that the high level of acceptance for the wind energy is declared by about 12% of Polish citizens, low – 3%, whereas 85% of Polish citizens accept the wind energy to a certain degree. [42]. It is unlikely that potential workplaces can positively influence the social acceptance of constructing a wind farm.

C4.1 The distance from a power grid connection is related with the facility of connecting the wind farm to the power grid. Such a connection should be as close as possible, since it reduces the possibility of potential problems on a transfer line related to the quality and stability of power supply [43].

C4.2 The distance from the road is important during the construction period. A short distance from main roads enables the comfortable delivery of construction elements, such as masts or rotors, to the site. One needs to understand that, as far as Poland is concerned, the road infrastructure is usually poorly developed in the areas with good wind conditions, [44].

C4.3 The Natura 2000 protected areas are breeding and resting sites for rare and endangered fauna and flora, as well as some rare natural habitats which are crucial to the European Community [45]. Locations which form part of the Natura 2000 protected areas are more likely to encounter difficulties with investment, since some potentially negative acts towards the sites are prohibited. It is possible to obtain permits to carry out actions which negatively impact the sites [45]; however, these are connected with incurring financial outlays and downtimes in the construction of a wind farm. It should be noted that, according to Polish law, wind farms and other buildings cannot be situated within national and landscape parks or nature reserves [45].

On the basis of the individual analysis of sub-criteria one can easily determine their dependencies and relationships. These dependencies of one criterion onto another are illustrated by a graphic outline of the decision model presented in Figure 1.

IV. Conclusions

The complex decision model, prepared by the authors of this article was prepared, in order to deal with the problem of selection of the location and design of a wind farm. Therefore, it takes into consideration complex dependencies between decision-making criteria (sub-criteria) and consequently, it can be more precise than decision models which assume independence between criteria.

The precision of a decision model is particularly important in Decision Support Systems; the latter recommend pareto optimal solutions to decision-makers. In the case of a less accurate decision model which does not take into consideration interdependencies of criteria, recommendations obtained in DSS can be inexact. Therefore, the designed decision model can be used as a DSS decision engine.

The verification of the prepared decision model should be carried out while taking into account the decision problem in the field of wind energy. Unfortunately, most MCDA methods, which form the foundation of DSS, assume that the independency between criteria is to simplify the decision-making process. These methods cannot be easily applied to more challenging decision-making problems [46] because, if redundant criteria are used in the model, one can reach an incorrect solution [47]. Therefore, an essential factor is a proper selection of an MCDA method [48], [31] which would make it possible to build a complex decision-making

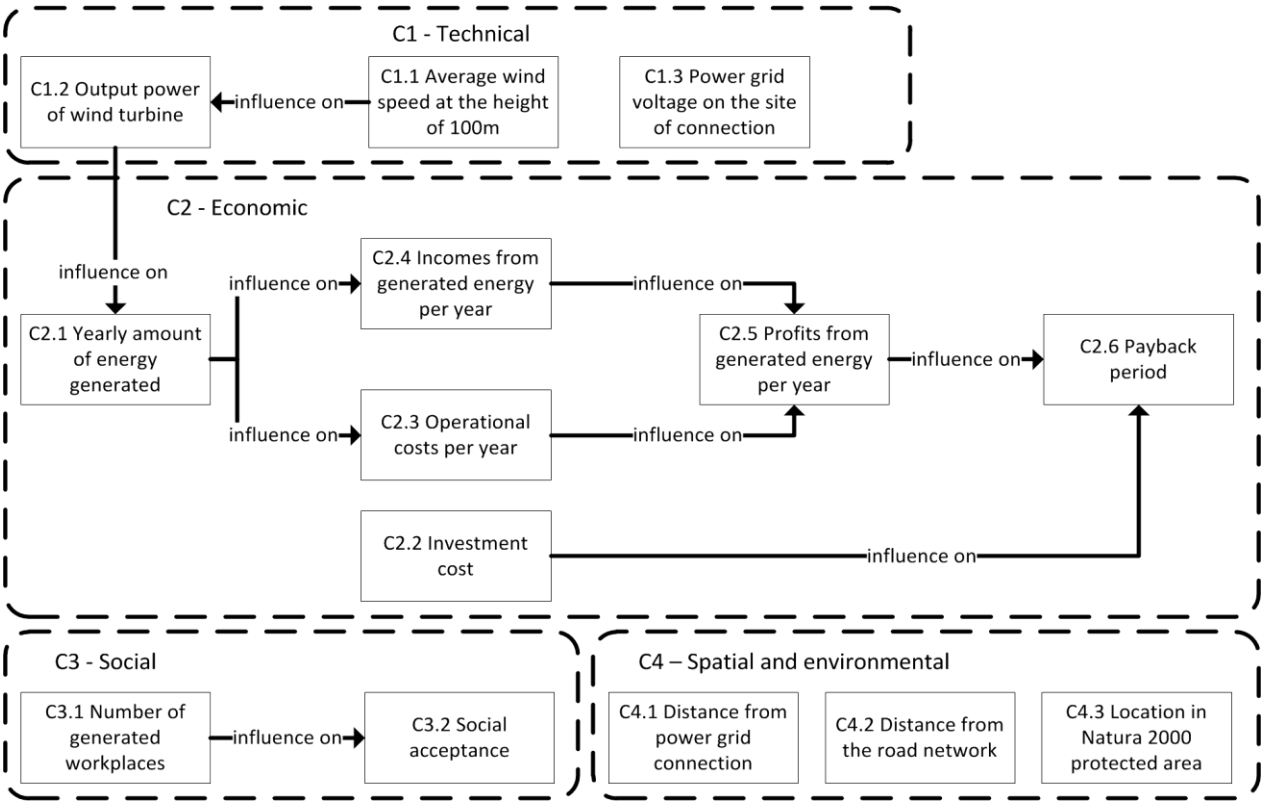


Fig. 1 Dependencies between criteria considered in the decision model

model for DSS, while taking into consideration mutual dependencies between criteria. The methods which allow the observation of the influence of individual criteria on other ones in the model are ANP [49], [30] and DEMATEL [50], [24].

The decision model which has been designed should evidently be used on a continuous basis. The best course of action would be to add criteria and sub-criteria of evaluations of wind farms, as well as other RES factors to the model. It would also be interesting to present the decision model in the form of an ontology [51], which would enable the deduction of new information from the model [52], [53]. The performance of the model would also move in a smoother manner towards the application of DSS in other situations.

REFERENCES

[1] K. Halicka, "Designing routes of development of renewable energy technologies," *Procedia – Social and Behavioral Sciences*, vol. 156, pp. 58-62, 2014.

[2] J. Paska and T. Surma, "Electricity generation from renewable energy sources in Poland," *Renewable Energy*, vol. 71, pp. 286-294, 2014.

[3] N. Scarlat, J.F. Dallemand, F. Monforti-Ferrario, M. Banja, and V. Motola, "Renewable energy policy framework and bioenergy contribution in the European Union – An overview from National Renewable Energy Action Plans and Progress Reports," *Renewable and Sustainable Energy Reviews*, vol. 51, pp. 969-985, 2015.

[4] J. Paska, M. Sałek, and T. Surma, "Current status and perspectives of renewable energy sources in Poland," *Renewable and Sustainable Energy Reviews*, vol. 13, pp. 142-154, 2009.

[5] Y. Wu, J. Zhang, J. Yuan, S. Geng, and H. Zhang, "Study of decision framework of offshore wind power station site selection based on ELECTRE-III under intuitionistic fuzzy environment: A case of China," *Energy Conversion and Management*, vol. 113, pp. 66-81, 2016.

[6] Y. Wu, S. Geng, H. Xu, and H. Zhang, "Study of decision framework of wind farm project plan selection under intuitionistic fuzzy set and fuzzy measure environment," *Energy Conversion and Management*, vol. 87, pp. 274-284, 2014.

[7] J. Wątróbski, P. Ziemba, and W. Wolski, "Methodological Aspects of Decision Support System for the Location of Renewable Energy Sources," *Annals of Computer Science and Information Systems*, vol. 5, pp. 1451-1459, 2015. http://dx.doi.org/ 10.15439/2015F294

[8] A.H.I. Lee, M.C. Hung, H.Y. Kang, and W.L. Pearn, "A wind turbine evaluation model under a multi-criteria decision making environment," *Energy Conversion and Management*, vol. 64, pp. 289-300, 2012.

[9] R.A. Taha and T. Daim, "Multi-Criteria Applications in Renewable Energy Analysis, a Literature Review," in *Research and Technology Management in the Electricity Industry*, T. Daim, T. Oliver, and J. Kim, Ed. London: Springer, 2013, pp. 17-30.

[10] J.R. San Cristobal, "Multi-criteria decision making in the selection of a renewable energy project in spain: The Vikor method," *Renewable Energy*, vol. 36, pp. 498-502, 2011.

[11] C. Henggeler Antunes and C. Oliveira Henriques, "Multi-Objective Optimization and Multi-Criteria Analysis Models and Methods for Problems in the Energy Sector," in *Multiple Criteria Decision Analysis. State of the Art Surveys*, 2nd ed., S. Greco, M. Ehrgott, and J.R. Figueira, Ed. New York: Springer, 2016, pp. 1067-1165.

[12] J. Jankowski, J. Wątróbski, P. Ziemba, "Modelling the impact of visual components on verbal comunication in online advertising," in *Computational Collective Intelligence. ICCCI 2015, Part II. LNAI,* vol. 9330, Heidelberg: Springer, 2015, pp. 44-53.

[13] F. Cavallaro, "Multi-criteria decision aid to assess concentrated solar thermal technologies," *Renewable Energy*, vol. 34, pp. 1678-1685, 2009.

[14] C.W. Holsapple, "DSS Architecture and Types," in *Handbook on Decision Support Systems*, vol. 1, F. Burstein and C.W. Holsapple, Ed. Heidelberg: Springer, 2008, pp. 163-189.

[15] T.M. Yeh and Y.L. Huang, "Factors in determining wind farm location: Integrating GQM, fuzzy DEMATEL and ANP," *Renewable Energy*, vol. 66, pp. 159-169, 2014.
[16] T. Kaya and C. Kahraman, "Multicriteria renewable energy planning using an integrated fuzzy VIKOR & AHP methodology: The case of Istanbul," *Energy*, vol. 35, pp. 2517-2527, 2010.
[17] A.H.I. Lee, H.H. Chen, and H.Y. Kang, "Multi-criteria decision making on strategic selection of wind farms," *Renewable Energy*, vol. 34, pp. 120-126, 2009.
[18] S. Al-Yahyai, Y. Charabi, A. Gastli, and A. Al-Badi, "Wind farm land suitability indexing using multi-criteria analysis," *Renewable Energy*, vol. 44, pp. 80-87, 2012.
[19] D. Latinopoulos and K. Kechagia, "A GIS-based multi-criteria evaluation for wind farm site selection. A regional scale application in Greece," *Renewable Energy*, vol. 78, pp. 550-560, 2015.
[20] J.M. Sanchez-Lozano, M.S. Garcia-Cascales, and M.T. Lamata, "Identification and selection of potential sites for onshore wind farms development in Region of Murcia, Spain," *Energy*, vol. 73, pp. 311-324, 2014.
[21] J.M. Sanchez-Lozano, M.S. Garcia-Cascales, and M.T. Lamata, "GIS-based onshore wind farm site selection using Fuzzy Multi-Criteria Decision Making methods. Evaluating the case of Southeastern Spain," *Applied Energy*, vol. 171, pp. 86-102, 2016.
[22] N.Y. Aydin, E. Kentel, H.S. Duzgun, "GIS-based site selection methodology for hybrid renewable energy systems: A case study from western Turkey," *Energy Conversion and Management*, vol. 70, pp. 90-106, 2013.
[23] Y. Noorollahi, H. Yousefi, and M. Mohammadi, "Multi-criteria decision support system for wind farm site selection using GIS," *Sustainable Energy Technologies and Assessments*, vol. 13, pp. 38-50, 2016.
[24] A. Fetanat and E. Khorasaninejad, "A novel hybrid MCDM approach for offshore wind farm site selection: A case study of Iran," *Ocean & Coastal Management*, vol. 109, pp. 17-28, 2015.
[25] F. Cavallaro and L. Ciraolo, "A multicriteria approach to evaluate wind energy plants on an Italian island," *Energy Policy*, vol. 33, pp. 235-244, 2005.
[26] W. Tian, J. Bai, H. Sun, and Y. Zhao, "Application of the analytic hierarchy process to a sustainability assessment of coastal beach axploitation: A case study of the wind power projects on the coastal beaches of Yancheng, China," *Journal of Environmental Management*, vol. 115, pp. 251-256, 2013.
[27] P. Ziemba, J. Wątróbski, J. Jankowski, and M. Piwowarski, "Research on the Properties of the AHP in the Environment of Inaccurate Expert Evaluations," in *Selected Issues in Experimental Economics*, K. Nermend and M. Łatuszyńska, Ed. Switzerland: Springer, 2016, pp. 227-243.
[28] J. Jankowski, J. Watrobski, and M. Piwowarski, "Fuzzy Modeling of Digital Products Pricing in the Virtual Marketplace," in *Proceedings of 6th International Conference on Hybrid Artificial Intelligent Systems*, LNCS, vol. 6678. Heidelberg: Springer, 2011, pp. 338-346.
[29] J. Jankowski, K. Kolomvatsos, P. Kazienko, J. Wątróbski, "Fuzzy Modeling of User Behaviors and Virtual Goods Purchases in Social Networking Platforms," *Journal of Universal Computer Science*, vol. 22, no. 3, pp. 416-437, 2016.
[30] P. Ziemba and J. Wątróbski, "Selected Issues of Rank Reversal Problem in ANP Method," in *Selected Issues in Experimental Economics*, K. Nermend and M. Łatuszyńska, Ed. Switzerland: Springer, 2016, pp. 203-225.
[31] J. Wątróbski, J. Jankowski, "Guideline for MCDA Method Selection in Production Management Area," in *New Frontiers in Information and Production Systems Modelling and Analysis. Intelligent Systems Reference Library*, vol. 98, Heidelberg: Springer, 2016, pp. 119-138.
[32] Y. Kumar, J. Ringenberg, S.S. Depuru, V.K. Devabhaktuni, J.W. Lee, E. Nikolaidis, B. Andersen, and A. Afjeh, "Wind energy: Trends and enabling technologies," *Renewable and Sustainable Energy Reviews*, vol. 53, pp. 209-224, 2016.
[33] *Plan sieci elektroenergetycznej najwyższych napięć*, PSE, http://www.pse.pl/index.php?dzid=80&did=23
[34] H. Sadeghi, "A novel method for adaptative distance protection of transmission line connected to wind farms," *Electrical Power and Energy Systems*, vol. 43, pp. 1376-1382, 2012.
[35] J. Paska and M. Kłos, "Elektrownie wiatrowe w systemie elektroenergetycznym – przyłączanie, wpływ na system i ekonomika," *Rynek energii*, no. 1/2010, pp. 3-10, 2010.
[36] *Wpływ energetyki wiatrowej na wzrost gospodarczy w Polsce*, Report, Ernst & Young, March 2013.
[37] *Wind energy in Poland*, Report, TPA Horwath, November 2013.
[38] *Informacja Prezesa Urzędu Regulacji Energetyki nr 46/2015 w sprawie średniej ceny sprzedaży energii elektrycznej na rynku konkurencyjnym w III kwartale 2015 roku*, Energy Regulatory Office, 21 December 2015, http://www.ure.gov.pl/pl/stanowiska/ 6361, Informacja-nr-462015.html
[39] *Ustawa o odnawialnych źródłach energii*, Dziennik Ustaw RP, 20 February 2015, http://isap.sejm.gov.pl/ DetailsServlet? id=WDU20150000478
[40] *Rozporządzenie Ministra Gospodarki w sprawie ceny referencyjnej energii elektrycznej z odnawialnych źródeł energii w 2016 roku*, Dziennik Ustaw RP, 13 November 2015, http://dziennikustaw.gov.pl/du/2015/2063/1
[41] M. Bukowski and A. Śniegocki, *Wpływ energetyki wiatrowej na polski rynek pracy*. Warszawa: Warszawski Instytut Studiów Ekonomicznych, 2015.
[42] B. Mroczek, *Akceptacja dorosłych Polaków dla energii wiatrowej i innych odnawialnych źródeł energii (streszczenie raportu)*. Szczecin: Polskie Stowarzyszenie Energetyki Wiatrowej, 21 March 2011.
[43] F. Santier, "Influence of Transmission Lines on Grid Connection," in *Proc. Deutsche Windenergie-Konferenz DEWEK 2006*, Bremen, 22-23 November 2006.
[44] P. Michalak and J. Zimny, "Wind energy development in the world, Europe and Poland from 1995 to 2009; current status and future perspectives," *Renewable and Sustainable Energy Reviews*, vol. 15, pp. 2330-2341, 2011.
[45] *Ustawa o ochronie przyrody*, Dziennik Ustaw RP, 18 April 2016, http://isap.sejm.gov.pl/DetailsServlet?id=WDU20040920880
[46] A. de Montis, P. De Toro, B. Droste-Franke, I. Omann, and S. Stagl, "Assessing the quality of different MCDA methods," in *Alternatives for Environmental Valuation*, M. Getzner, C.L. Spash, and S. Stagl, Ed. New York: Taylor & Francis, 2005, pp. 99-133.
[47] P. Thokala and A. Duenas, "Multiple Criteria Decision Analysis for Health Technology Assessment," *Value in Health*, vol. 15, no. 8, pp. 1172-1181, 2012.
[48] J. Wątróbski, J. Jankowski, "Knowledge Management in MCDA Domain," in *Proceedings of the Federated Conference on Computer Science and Information Systems. Annals of Computer Science and Information Systems*, vol. 5, pp. 1445–1450, 2015.
[49] T. L. Saaty and L.G. Vargas, *Decision Making with the Analytic Network Process. Second Edition*. New York: Springer, 2013.
[50] H. S. Lee, G.H. Tzeng, W. Yeih, Y.J. Wang, and S.C. Yang, "Revised DEMATEL: Resolving the Infeasibility of DEMATEL," *Applied Mathematical Modelling*, vol. 37, no. 10-11, 2013, pp. 6746-6757.
[51] P. Ziemba, J. Jankowski, J. Wątróbski, J. Becker, „Knowledge Management in Website Quality Evaluation Domain," *Lecture Notes in Artificial Intelligence*, vol. 9330, pp. 75-85, 2015.
[52] P. Ziemba, J. Jankowski, J. Wątróbski, W. Wolski, and J. Becker, "Integration of Domain Ontologies in the Repository of Website Evaluation Methods," *Annals of Computer Science and Information Systems*, vol. 5, pp. 1585-1595, 2015. http://dx.doi.org/ 10.15439/2015F297
[53] P. Ziemba, J. Wątróbski, J. Jankowski, and W. Wolski, "Construction and Restructuring of the Knowledge Repository of Website Evaluation Methods," *Lecture Notes in Business Information Processing*, vol. 243, pp. 29-52, 2016. http://dx.doi.org/ 10.1007/978-3-319-30528-8_3

11th Conference on Information Systems Management

THIS event constitutes a forum for the exchange of ideas for practitioners and theorists working in the broad area of information systems management in organizations. The conference invites papers coming from two complimentary directions: management of information systems in an organization, and uses of information systems to empower managers. The conference is interested in all aspects of planning, organizing, resourcing, coordinating, controlling and leading the management function to ensure a smooth operation of information systems in an organization. Moreover, the papers that discuss the uses of information systems and information technology to automate or otherwise facilitate the management function are specifically welcome.

TOPICS

- Management of Information Systems in an Organization:
  - Modern IT project management methods
  - User-oriented project management methods
  - Business Process Management in project management
  - Managing global systems
  - Influence of Enterprise Architecture on management
  - Effectiveness of information systems
  - Efficiency of information systems
  - Security of information systems
  - Privacy consideration of information systems
  - Mobile digital platforms for information systems management
  - Cloud computing for information systems management
- Uses of Information Systems to Empower Managers
  - Achieving alignment of business and information technology
  - Assessing business value of information systems
  - Risk factors in information systems projects
  - IT governance
  - Sourcing, selecting and delivering information systems
  - Planning and organizing information systems
  - Staffing information systems
  - Coordinating information systems
  - Controlling and monitoring information systems
  - Formation of business policies for information systems
  - Portfolio management,
  - CIO and information systems management roles

EVENT CHAIRS

- **Arogyaswami, Bernard,** Le Moyne University, USA
- **Chmielarz, Witold,** University of Warsaw, Poland
- **Karagiannis, Dimitris,** University of Vienna, Austria
- **Kisielnicki, Jerzy,** University of Warsaw, Poland
- **Ziemba, Ewa,** University of Economics in Katowice, Poland

PROGRAM COMMITTEE

- **Abu-Shanab, Emad,** Yarmouk University, Jordan
- **Alghamdi, Saleh,** University of Sussex, United Kingdom
- **Bialas, Andrzej,** Institute of Innovative Technologies EMAG, Poland
- **Bicevska, Zane,** DIVI Grupa Ltd, Latvia
- **Chung, Tsungting,** Douliou Yunlin Uniwersytet, Taiwan
- **Czarnacka-Chrobot, Beata,** Warsaw School of Economics, Poland
- **DeLorenzo, Gary,** California University of Pennsylvania, United States
- **Dima, Ioan Constantin,** Valahia University of Targoviste, Romania
- **Duan, Yanqing,** University of Bedfordshire, United Kingdom
- **Dudycz, Helena,** Wrocław University of Economics, Poland
- **El Emary, Ibrahim,** King Abdulaziz Univetrsity, Saudi Arabia
- **Espinosa, Susana de Juana,** University of Alicante, Spain
- **Geri, Nitza,** The Open University of Israel, Israel
- **Grublješič, Tanja,** University of Ljubljana, Slovenia
- **Halawi, Leila,** Embry-Riddle Aeronautical University, United States
- **Jankowski, Jarosław,** West Pomeranian University of Technology in Szczecin, Poland
- **Jelonek, Dorota,** Czestochowa University of Technology, Poland
- **Kobyliński, Andrzej,** Warsaw School of Economics, Poland
- **Michalik, Krzysztof,** University of Economics in Katowice, Poland
- **Mullins, Roisin,** University of Wales Trinity Saint David, United Kingdom
- **Niedźwiedziński, Marian,** University of Lodz, Poland
- **Owoc, Mieczyslaw,** Wroclaw University of Economics, Poland

- **Ozkan, Necmettin,** Turkiye Finans Participation Bank, Turkey
- **Pastuszak, Zbigniew,** Maria Curie-SKlodowska University, Poland
- **Ranjan, Jayanthi,** Institute of Management Technology in Ghaziabad, India
- **Ren, Kun,** Yale University, United States
- **Rizun, NIna,** Alfred Nobel University, Dnipropetrovs'k, Ukraine
- **Schroeder, Marcin,** Akita International University, Japan
- **Sikorski, Marcin,** Gdańsk University of Technology, Poland
- **Silber-Varod, Vered,** The Open University of Israel, Israel
- **Skovira, Robert,** Robert Morris University, United States
- **Sobczak, Andrzej,** Warsaw School of Economics, Poland
- **Surma, Jerzy,** Warsaw School of Economics, Poland
- **Świerczyńska-Kaczor, Urszula,** Jan Kochanowski University in Kielce, Poland
- **Symeonidis, Symeon,** Democritus University of Thrace, Greece
- **Szczerbicki, Edward,** University of Newcastle, Australia
- **Tarhini, Ali,** Brunel University London, United Kingdom
- **Travica, Bob,** University of Manitoba, Canada
- **Wachnik, Bartosz,** University of Technology in Warsaw, Poland
- **Wątróbski, Jarosław,** West Pomeranian University of Technology in Szczecin, Poland
- **Wielki, Janusz,** Opole University of Technology, Poland
- **Wolski, Waldemar,** University of Szczecin, Poland
- **Ziemba, Paweł,** The Jacob of Paradyż University of Applied Science, Poland

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1243–1252
DOI: 10.15439/2016F37
ACSIS, Vol. 8. ISSN 2300-5963

Critical success factors for ERP implementation in SMEs

Prodromos Chatzoglou
Democritus University of Thrace,
Vasillisis Sofias 12, 67100, Xanthi, Greece
Email: pchatzog@pme.duth.gr

Dimitrios Chatzoudes
Democritus University of Thrace,
Vasillisis Sofias 12, 67100, Xanthi, Greece
Email: dchatzoudes@yahoo.gr

Leonidas Fragidis
Democritus University of Thrace,
Vasillisis Sofias 12, 67100, Xanthi, Greece
Email: lfrangid@pme.duth.gr

Symeon Symeonidis
Democritus University of Thrace,
Vasillisis Sofias 12, 67100, Xanthi, Greece
Email: ssymeoni@ee.duth.gr

***Abstract*—The highly competitive global environment of the last few decades has urged companies to rely on Information Systems (IS) in order to improve customer service, reduce costs and increase productivity. In that direction, Enterprise Resource Planning (ERP) systems are being used as significant strategic tools that provide competitive advantages and lead to operational excellence. Despite that, ERP implementation projects are complicated, costly and include high failure risks. The present study aims (a) to develop and (b) empirically test a conceptual framework that investigates the factors affecting ERP system effective implementation in Small and Medium Enterprises (SMEs). The examination of the conceptual framework was made with the use of a newly-developed structured questionnaire that was distributed to a group of Greek SMEs. After the completion of the research period, 159 usable questionnaires were returned. The reliability and the validity of the questionnaires were thoroughly examined, while research hypotheses were tested using the "Structural Equation Modeling" (SEM) technique. Results offer interesting empirical observations and managerial implications.**

I. Introduction

THE business world has been hugely transformed during the last few decades [1]. Globalisation, increasing competition, constant change in the external environment, and private sector growth are among the most significant changes in the global business environment [2]. These transformations have urged most companies to adapt in order to survive [1] [3] [4].

More specifically, organisations aim at maintaining or improving the level of their competitiveness by using Information Systems (IS) in order to reduce costs, increase customer satisfaction, and improve business processes [5]. According to various authors [3] [6] [7], this drive for achieving higher levels of productivity, effectiveness, and performance is urging organisations to adopt Enterprise Resource Planning (ERP) systems. Tsai, Li, Lee and Tung [4] argue that, since their introduction in the early 1990s, ERP systems have become the centre of modern business.

ERP systems are Information Systems (IS) that facilitate the integration of business processes across functional units, using a common database and shared information [4] [6] [8]. According to Garg and Garg [8], this "enables the decision-making process to be timely, consistent and reliable across organizational units and geographical locations" (p. 424).

ERP implementation has various benefits throughout the organisation: elimination of redundant information, drastic declines in inventory, reduction of production cost, better understanding of the changing customer needs, more efficient management of the extended network of suppliers and customers, increased productivity, improved response time, and decreased production cycle [5] [8] [9] [10] [11]. Considering these benefits, it is not surprising that ERP systems are being treated as a major development in the world of business, and have been accepted as a standard business software over the last fifteen years [8] [12].

However, ERP implementation requires considerable financial resources, while the whole implementation project is considered complex, lengthy, and quite challenging [9] [13]. As a result, the success rate of such projects is considered to be quite disappointing [14] [15] [16]. More specifically, Samuel and Kumar [17], argue that the success rate is, only, around 50%, while approximately 90% of ERP implementation projects are late, or over budget. On the same vein, Umble and Umble [18] reported failure rates between 50% and 75%. Therefore, additional empirical studies are necessary in order to assist companies in increasing the success rates of ERP implementation projects.

Under that context, the aim of the present study is twofold: (a) develop an original conceptual framework (research model) examining the impact of various research factors on ERP implementation success, (b) empirically test that framework, using data from Small and Medium Enterprises (SMEs) located in Greece (empirical research).

(a) The development of the conceptual framework was based on two methodological steps: firstly, a review of the literature identified the factors that were used by previous studies as antecedents of ERP implementation success; secondly, a panel of experts was used in order to discuss these factors and provide a list of the most significant ones. That approach was selected due to the significant number of factors that have been proposed in the relevant literature. More specifically, the members of the research team used the opinions of experienced practitioners as a criterion for selecting a specific set of factors from the extensive list that

This work was not supported by any organization.

was provided from the literature review analysis. It is strongly argued that randomly selecting the research factors of the proposed conceptual framework would have resulted in the limited reliability of the present research.

(b) The empirical examination of the conceptual framework (that was crystallised after the literature review analysis and the completion of the qualitative research) was conducted on a sample of Greek SMEs. More specifically, a newly-developed structured questionnaire was used in order to collect the appropriate primary data. The questionnaire was distributed to 421 companies, while 159 usable questionnaires were, finally, returned. Advanced statistical techniques (EFA, CFA) were used in order to enhance the validity and reliability of the results, while research hypotheses were tested using the "Structural Equation Modeling" (SEM) technique.

The present study makes an effort to point out areas that companies should emphasize in order to successfully adopt ERP systems and, therefore, harvest their potential benefits. Its contribution lies in this enhanced approach. In synopsis, the study contributes in the following areas:

- It focuses on Small and Medium enterprises (SMEs), an approach that has found limited empirical investigation in the international literature. The literature review analysis underlined that the contemporary research mostly examines the implementation of ERP systems in large organizations.
- It examines the antecedents of ERP implementation success in SMEs of a European country. The literature review analysis that was conducted failed to recognise enough similar studies.
- It uses a qualitative research in order to recognise the most important antecedents of ERP implementation success and, then, develops a conceptual framework based on these factors. According to the best of the researchers' knowledge, such an approach is unique in the relevant literature. Moreover, it is significant, since previous studies used factors that were randomly selected from the literature, without a solid empirical basis [8] [19] [20] [21].
- It can be perceived as a reference point for future studies, since it offers a critique concerning the multitude of ERP implementation antecedents that have been examined in the international literature.
- Its results may be generalized in other developed countries with similar characteristics, and produce valuable managerial lessons for practitioners in these countries.

The following section includes a review of the relevant literature, while section three presents the conceptual framework of the study. The fourth section includes the research methodology. Results and conclusions are discussed in sections 5 and 6 respectively.

II. Literature Review

A. Critical success factors

Critical Success Factors (CSFs) have been introduced during the 1960s, as a concept that would assist companies to achieve their goals and enhance their overall competitiveness [22] [23] [24]. According to Ram and Corkindale [22], CSFs constitute a systematic way of identifying key business areas that require constant management attention. On the same vein, Rockart [23] argues that the results obtained in these critical areas, if satisfactory, are able to significantly enhance organisational performance. In plain words, CSFs assist managers to directly affect a specific outcome, by proactively taking necessary actions in certain areas [25].

Not surprisingly, the concept of CSFs gained wide recognition in the Information Systems domain and, consequently, in the context of ERP systems [6]. Since high failure rates of ERP implementation projects have been observed by numerous studies [14] [15], many scientists have attempted to investigate the factors that may enhance the whole implementation process. According to Ram, Corkindale and Wu [19], a large number of CSFs have been identified throughout the international literature.

Indeed, the literature review analysis that was conducted revealed that the relevant literature includes numerous studies that have, mostly, been published during the last 15 years [24]. Among these studies, some are theoretical [26], some others are empirical [8] [9] [27], while just a few have adopted the case-study approach [28].

According to Saade and Nijher [24], despite the growth in the investigation of CSFs regarding ERP implementation, there is a long way before the empirical contribution can be considered to be substantial. Moreover, most of the empirical studies that have been conducted [8] [9] [21], incorporated a limited number of critical factors in their analysis, failing to draw a more complete picture of the phenomenon. Finally, despite the wide range of CSFs proposed in the literature, many organisations continue to experience failures and difficulties in implementing ERP systems [19], thus, calling for additional research.

More significantly, according to Ram and Corkindale [22], there is a lack of an established process for the identification of CSFs. Various authors use subjective criteria in order to select the critical factors utilised in their studies, something that results in a lack of objective approaches. The present study heals that gap in the relevant literature, by developing a conceptual framework that was crystallised after a coherent two-step approach (literature review analysis and consultation with experienced practitioners / focus-group methodology).

B. Previous studies

Numerous empirical studies have investigated the critical success factors for ERP system implementation. The present study conducted an extensive review of the relevant literature, in an effort to grasp a spherical view of the subject and, therefore, better define its scope. The following paragraphs present a brief analysis of a representative sample of previous empirical studies.

Saini, Nigam, and Misra [15] examined the success factors for implementing ERP systems at Indian SMEs. Their sample included 164 companies, while the empirical data were analysed using the statistical z-test. Support was found for all hypotheses, arguing that technological factors

(e.g. system testing, IT infrastructure), people factors (e.g. cross-functional team, morale of the implementation team), and organizational factors (e.g. adaptability to changes, comprehensiveness of the implementation strategy) have a direct impact on the success of ERP implementation [15].

Garg and Chauhan [20] explored the factors affecting the success of ERP implementation in the Indian retail sector. Their conceptual framework, which included various critical success factors, explained 62,7% of the variations of ERP implementation success. As with Saini, Nigam, and Misra [15], organizational, technological, and people-related factors were found to be significant antecedents of ERP implementation success. Additionally, the impact of project management was, also, identified as being significant [20]. Garg and Garg [8] in another similar study that was, also, conducted in the Indian retail sector, found out that strategic, technological, people and project management factors have a positive influence on ERP implementation success.

Chien, Lin, and Shih [10] investigated the impact of centrifugal and centripetal forces on team cohesion and successful ERP implementation. Their empirical results were based on a survey of 305 Taiwanese SMEs. It was found that centripetal forces have a significant impact on ERP implementation, while the same was not verified for centrifugal forces, as well. Finally, team cohesion seemed to moderate the relationship between centripetal forces and ERP implementation performance [10].

Zabjek, Kovacic, and Indihar Stemberger [9] identified business process management as an important antecedent of ERP effective implementation. Their analysis was based on 152 questionnaires collected from Slovenian companies. They concluded that top management support, change management and business process management have a positive impact on successful ERP implementation [9]. The same authors conducted another similar research [16], obtaining identical results. On the same vein, Garg and Agarwal [21], also, underlined the significance of top management commitment, user involvement, business process reengineering, project management and ERP teamwork and composition on the success of ERP implementation [21].

Li, Markowski, Xu, and Markowski [29] used Structural Equation Modeling (SEM) in order to analyze data from 154 manufacturing companies operating in the USA. Their analysis revealed that Total Quality Management (TQM) is an important predecessor of ERP implementation. Chou, Hung, and Chang [30] focused on ERP organizational fit and knowledge transfer. They concluded that ERP success is influenced by organizational fit (data fit, process fit, user fit), ERP knowledge factors (e.g. shared understanding), and ERP communication factors (e.g. communication decoding competence), either directly or indirectly [30].

On a different approach, Amid, Moalagh, and Ravasan [31] focused on Critical Failure Factors (CFFs), rather that Critical Success Factors (CSFs). Firstly, they conducted semi-structured interviews with practitioners, identifying 47 failure factors. Secondly, they collected empirical data with the use of a structured questionnaire. Using Exploratory Factor Analysis (EFA), they classified CFFs in seven large groups (vendor and consultant, human resources, managerial, project management, processes, organizational, technical). Their research was conducted on a developing country, namely Iran [31].

Wee [32] underlined the importance of formulating an overall ERP architecture before the deployment of the system, since only in such a way the need for reconfiguration during, or after, its real-time implementation will significantly diminish. Similar views were supported by other authors, arguing that the use of proper and formal modelling methods, tools and architectures is necessary for ERP implementation success [33]. Ferratt, Ahire, and De [34] argued that implementing organisations need to follow the basics of project management and, simultaneously, adopt the best industry practices in order to successfully implement an ERP system.

Ngai, Law, and Wat [35] focused on the importance of national culture and country-related characteristics on ERP implementation success. Sheu, Chae and Yang [36] underlined the impact of different cultural backgrounds, while Tarafdar and Roy [37] analysed the cultural issues that a typical Indian firm, usually, faces when implementing an ERP system. Finally, Lee, Lee, and Kang [38] argued that implementation success largely depends upon the attitude of the employees towards the whole ERP project.

As it became evident from the previous paragraphs, contemporary research includes a wide range of critical factors predicting ERP implementation success, ranging from vendor selection [39], to project management aspects [14] [40]. However, most of these studies are focused on larger enterprises [41]. On the other hand, ERP adoption by SMEs has traditionally received less attention from the international literature. According to Poba-Nzaou, Raymond and Fabi [41], this represents an area for additional research, especially since SMEs face greater difficulties in adopting ERP systems.

In summary, the literature includes the following gaps: (a) There is a multitude of critical success factors (antecedents) that have been used in order to predict ERP implementation success. Therefore, one is unable to determine which are actually the most important. The need for additional research is imperative; (b) The focus on SMEs has been limited; (c) Very few studies have utilised specific criteria for selecting certain factors, and excluding others, from their analysis. Selecting factors without justification is considered as a significant limitation; (d) Few of the published empirical studies were carried out in European countries; (e) Very few studies built on previous research. The present study was designed so as to cover these limitations (research gaps) found in the relevant literature.

III. Conceptual framework

As mentioned earlier, the present study aims to: (a) built a coherent conceptual framework including the most significant antecedents of ERP implementation success and (b) test that framework gathering quantitative data.

The literature review analysis that was conducted prior to the development of the conceptual framework of the present study revealed that numerous factors have been used in order to predict ERP implementation success. Therefore, an important challenge was to decide upon the factors that were going to be incorporated into the proposed conceptual framework. The main objective was to construct a conceptual framework that incorporates the most significant factors used in the literature. Moreover, the incorporated factors were expected to have a high degree of relevance with the overall context of the study (Greek SMEs).

In order to address that critical issue, a qualitative research was conducted prior to the quantitative research. More analytically, a 'panel of experts' was formed in order to evaluate the factors that have been used in the relevant literature and assist in selecting the most appropriate ones for the proposed conceptual framework of the present study. More specifically, the focus group methodology was used.

This approach offers certain benefits: (a) the selection of the factors that were, finally, incorporated in the proposed conceptual framework was not conducted according to the subjective judgment of the researchers, but was a result of a more coherent and objective procedure, (b) the proposed conceptual framework has a strong basis on the opinions of experienced practitioners (managers of SMEs), (c) the selection of factors with low significance was avoided. It is believed that the random selection of the research factors, without any theoretical or empirical justification, would have resulted in the limited reliability of the present study.

In order to enhance the validity of the qualitative research, two sessions held in different geographical areas were conducted. All companies were selected in random, using data from the Chamber of Commerce. Each focus group included five managers of SMEs. This approach is in line with the main principles of the focus group methodology [42], since there was an appropriate number of participants for each session, two different sessions with different participants were conducted, while the represented companies were randomly selected.

The participants of each group were given (in paper) an extensive list of factors that have been used in the literature in order to predict ERP successful implementation. Then, a detailed conversation was conducted, with two members of the research team acting as moderators [43]. Each focus group took approximately two hours. Notes were taken during each session by a second moderator, while additional notes were added after reviewing the recorded sessions. After long discussions and deliberations, each focus group unanimously chose the nine most important factors of the provided list. The two focus groups agreed, with minor exceptions, in the same factors.

The conceptual framework of the present study incorporates these nine (independent) factors, resulting from the qualitative research, and one dependent factor, namely ERP implementation success. Additionally, 'organisational impact' was added in the proposed conceptual framework, in order to investigate the effect of ERP implementation on various measures of organisational performance.

The nine independent factors are listed below: Top management support, Organizational culture, External pressure, Vendor support, Project management, Training, User involvement, Business Process Reengineering, Implemented modules.

A. Top management support could be easily defined as the involvement of business executives in the areas related with ERP implementation [44]. It has been highlighted, by several authors, as a critical factor for the successful implementation of ERP systems [14] [45] [46].

Ngai, Law, and Wat [35] argued that senior executives play a significant role in ERP implementation success, since these projects are, usually, time consuming and demand extensive financial support. Senior management has two roles during implementation: supplying funds and offering leadership [14]. Al-Mashari, Al-Mudimigh, and Zairi [47] insisted that senior management support should be offered, without disruption, during the whole implementation period. The tasks of senior executives when implementing ERP projects include: communicating the strategy to all business employees, setting limitations, proving engagement, and setting reasonable goals [46]. Participation, support, and senior-level sponsorship are dimensions that have been found to significantly affect ERP implementation [48] [49].

ERP implementation does not, exclusively, evolve around software reengineering. On the contrary, it includes the extensive restructuring of business processes. Consequently, senior executives must clearly, publicly and truly indicate their support (economic or not), in order to highlight the priority given to implementation [48] [49]. Therefore:

H1: Top management support has a direct positive effect on ERP implementation success.

B. Organizational culture represents the shared ideologies, standards, convictions that have an impact on organizational attitudes and activities [50].

A common culture, shared between various organizational members, has an impact on the willingness to change, e.g. to adopt a new Information System. Research has shown that organizational culture is quite significant for the success of most organizational changes [51] [52]. Jarvenpaa and Staples [52] argue that there should be a fit between the culture of the organization and the nature of the changes that may occur from implementing an ERP system.

Additionally, according to Jones, Cline, and Ryan [51], organizational culture has an effect on employee behavior towards knowledge sharing, while knowledge sharing is crucial for the successful implementation of ERP systems. Ruppel and Harrington [53] argue that organizational culture has an effect on the implementation of intranet and other information systems used inside the organisation. Hence, it can be hypothesised that:

H2: Organizational culture has a direct positive effect on ERP implementation success.

C. Sometimes, the implementation of an ERP system does not have intrinsic motives. On the contrary, companies are being forced to implement an Information System, either by

their supply chain partners or by their competitors [54] [55]. In the first case, implementation becomes a prerequisite for the continuous cooperation with a partner (supplier and/or customer), while in the second case, the adoption decision is based on the need to follow the competitors, and, hence, avoid any possible downturn from not doing so [56].

In the present study, it is hypothesised that when companies find themselves under pressure from the external environment, they tend to try harder to achieve their desired goals. Therefore, the higher the external pressure, the more successful the implementation of the ERP system.

H3: External pressure has a direct positive effect on ERP implementation success.

D. Vendor support is offered from software retailers and/or consulting companies [4]. In most of the cases, the retailer is, also, the consultant during, or after, the implementation.

Vendor support includes user training, extended technical assistance during and after the implementation, maintenance, updates, etc. Additionally, vendors offer analytical advice concerning the selection of the appropriate ERP software [57] [58]. According to Wang, Lin, Jiang, and Klein [57], vendors significantly enhance the effectiveness of the implemented system, via experience sharing and knowledge transfer.

Through continuous collaboration, formal training and knowledge dissemination, consultants assist their costumers in receiving the full benefits of the implemented system [57] [59]. The trustworthiness of the vendor is extremely important in determining the success, or the failure of the whole effort [60]. Koh, Simpson, Padmore, Dimitriadis, and Misopoulos [60] found out that the close relationship with the vendor is a critical success factor for the implementation of an ERP system.

H4: Vendor support has a direct positive effect on ERP implementation success.

E. The implementation of an ERP system is a risky and complex project [21]. As it is evident, such projects acquire excellent management, since numerous stakeholders (different business units, suppliers, customers, vendors/ consultants) are deeply involved [15] [20]. The manager of an ERP project should bear in mind different timetables, various milestones, equipment requirements, workforce availability, and budget needs [49]. Hence, successful implementation is synonymous with the management of a plethora of tasks. All these tasks should be carefully monitored and managed.

More specifically, standard meetings and reports should be provided for all project collaborators. Effective project management is very crucial, since implementation success is, usually, assessed on the basis of budget and time compliance [21]. Executives expect the implementation period to be completed on time, and on budget.

H5: Project management has a direct positive effect on ERP implementation success.

F. Training is considered to be a basic parameter in every ERP implementation project [45] [46]. Hong and Kim [27] argue that training should be provided before, during and after implementation, while both technical and procedural issues should be carefully addressed. Finally, in-house training (on-the-job training) appears to be the most efficient choice, between all available methods [14].

Dezdar and Ainin [61] argued that sufficient training allows employees to efficiently utilize the implemented ERP system. More specifically, training enhances the skills and increases the practical expertise of real-time users [62]. Nah, Zuckweiler, and Lau [63] found out that adequate training enhances implementation success, while lack of training undermines the whole process. Additionally, sufficient training builds a positive climate towards the implemented system, thus, increasing its use and overall acceptance. Moreover, training enhances the ease of use, which in turn increases the probability for system success [62].

H6: Training has a direct positive effect on ERP implementation success.

G. User involvement is one of the most influential factors in ERP implementation projects [14] [27] [47]. Numerous studies argue that users should be actively involved before and during the entire ERP implementation process [8]. This will ensure that the system has a better fit with business processes, since its development will be based on real needs. Moreover, the acceptance of the ERP system will be increased, since users will have participated in its development. Finally, resistance to change will be significantly decreased [8] [47].

According to various authors [20], user involvement increases user satisfaction and user acceptance, by developing realistic expectations about the capabilities of the system. Additionally, user involvement increases the perceived level of control, through user participation in the entire project [20]. When all the above conditions are being successfully met, the implementation of the ERP system will be much more efficient.

H7: User involvement has a direct positive effect on ERP implementation success.

H. Business Process Reengineering (BPR) is the fundamental rethinking and drastic redesign of business processes, in order to achieve improvements in critical measures, such as cost, quality, delivery, and speed [49] [64]. ERP implementation requires such a radical redesign of business processes, since the new ERP system is expected to drastically change several aspects of doing business [40].

It is the ERP system that underlines the necessity for BPR and forces the organization to redefine and redesign work flows in order to fit the new software [40]. Reengineering business processes in a way that makes them compatible with the implemented system appears as an important antecedent of ERP implementation success [64].

H8: Business Process Reengineering (BPR) has a direct positive effect on ERP implementation success.

I. ERP systems may be implemented in modules. A company does not have to conduct a full scale implementation; on the contrary, certain modules could be implemented on the basis of its special needs and requirements [65]. According to Yeh, Yang, and Lin [66], it would be unwise to avoid implementing most of the available modules, since only full implementation really ensures the expected benefits. Some empirical studies have argued that there is a relationship between the number of implemented modules and the functional effectiveness of the ERP system [67]. After all, the more modules a company implements, the higher its benefits from cross-operational cooperation [65].

H9: The number of implemented modules has a direct positive effect on ERP implementation success.

J. The construct of "organizational performance", as it has been captured in the present study, includes measures of multiple dimensions, such as productivity, cycle time, cost reduction, information flow, and customer satisfaction. Its main goal is to include both qualitative and quantitative measures of organisational performance. Law and Ngai [68] followed a similar approach. Many previous studies have investigated the impact of ERP implementation on firm performance [69] [70], while its impact on organisational performance has received less empirical examination. Therefore, it is hypothesised:

H10: ERP implementation success has a direct positive effect on organizational performance.

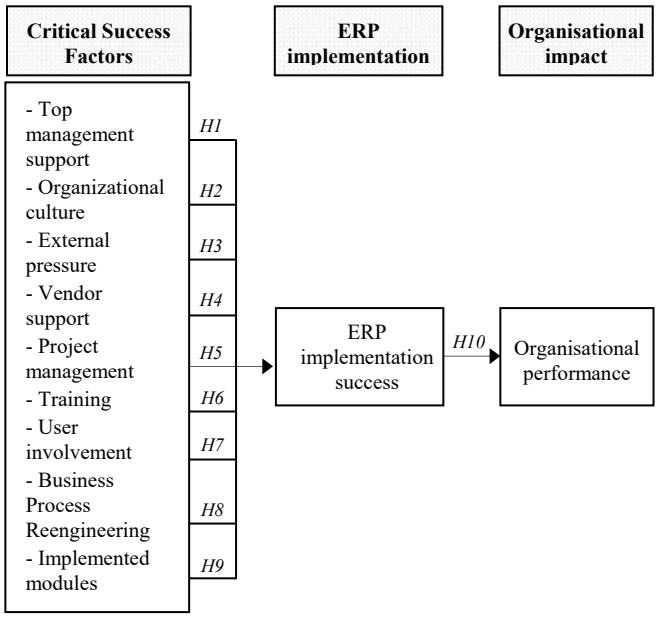


Fig. 1 The proposed conceptual framework

The synthesis of the hypotheses presented above formulates the proposed conceptual framework of the present study (Fig. 1). It should be underlined that, according to the best of the researcher's knowledge, such a conceptual framework (combination of factors) has never been examined before in the literature.

IV. Research Methodology

A. Population of the study

The population of the present study includes Greek Small and Medium Enterprises (SMEs) that have implemented an ERP system. There are no available official data that can define the population of the study on numerical terms.

SMEs are considered to be the heart of the Greek economy, since they represent 99% of the total number of companies. In 2010, there were 742.000 SMEs, with 2.512.493 employees, which represent more than 75% of total employment, well above the EU average. Greece has a very high share of SMEs, particularly micro enterprises, compared to the EU average (Annual Report on EU SMEs 2010/2011, 2011).

B. Measurement

The proposed conceptual framework was tested with the use of a newly-developed structured questionnaire. The measurement of the eleven research factors was conducted with the use of multiple questions (items) that were adopted from the international literature [14] [15] [20] [27] [51] [52] [54] [55] [56] [58] [59] [65] [66] [68]. All questions were translated to Greek and then back to English by another person, in order to detect any discrepancies. The five-point Likert scale was used for the measurement of all factors.

C. Data collection

Data concerning companies that could possibly be included in the sample were obtained via the web sites of the leading ERP system providers operating in Greece. Since no other database including companies using ERP systems exist, the use of the certain method was the only one able to provide usable information. Totally, 678 companies that have implemented an ERP system were identified. The questionnaire and a cover letter including clarifications, was sent to the IT managers of these companies.

Questionnaires were sent only after a telephonic contact with the IT manager in each company has been established. After making all necessary telephone calls, 421 questionnaires were distributed to 421 companies that agreed to participate in the survey. The research period lasted three months (March to May 2015). Totally, 165 questionnaires were returned, and after conducting all necessary controls 159 were used for data analysis. The 159 questionnaires represent a very satisfactory response rate (38%). The majority of the participating companies are small sized (less than 100 employees), something that is in line with the country's average firm size.

D. Reliability and validity

The questionnaire that was used in the present study was rigorously tested for its content and construct validity.

The test for the content validity was conducted via a pilot study. More specifically, a draft of the final questionnaire was sent to four practitioners and two academics, in order to test whether it met all theoretical and practical requirements.

TABLE I. ESTIMATION OF UNIDIMENSIONALITY AND RELIABILITY

Factors	*KMO*	*Bartlett's Test*	*Eigen-value*	*TVE*	*Cronbach Alpha*
Top management support	0,736	139,2[a]	2,546	67,3%	0,789
Organizational culture	0,894	214,9[a]	2,871	71,5%	0,823
External pressure	0,779	77,5[a]	1,371	68,4%	0,801
Vendor support	0,831	145,6[a]	2,874	81,7%	0,745
Project management	0,799	154,2[a]	1,741	84,7%	0,771
Training	0,854	95,5[a]	2,713	71,9%	0,723
User involvement	0,736	214,3[a]	2,124	76,2%	0,755
Business Process Reengineering	0,711	325,3[a]	2,587	74,1%	0,737
Implemented modules	0,857	217,6[a]	1,342	83,4%	0,741
ERP implementation success	0,839	169,7[a]	1,619	84,5%	0,901
Organisational performance	0,759	171,3[a]	2,391	88,6%	0,733

[a.] $p<0,01$

TABLE II. ESTIMATION OF THE GOODNESS OF FIT

Factors	*Normed X^2*	*C.R.*	*V.E.*	*RMSEA*	*CFI / GFI*
Top management support	1,57	0,78	65,6%	0,077	0,94 / 0,96
Organizational culture	2,67	0,74	69,4%	0,053	0,97 / 0,97
External pressure	3,15	0,86	0,81%	0,067	0,99 / 0,97
Vendor support	3,52	0,82	0,76%	0,084	0,91 / 0,93
Project management	2,19	0,76	0,67%	0,075	0,99 / 0,98
Training	1,97	0,77	0,63%	0,063	0,90 / 0,93
User involvement	2,37	0,69	0,57%	0,086	0,95 / 0,99
Business Process Reengineering	2,45	0,73	0,81%	0,059	0,90 / 0,90
Implemented modules	2,65	0,83	0,74%	0,061	0,91 / 0,96
ERP implementation success	2,77	0,77	0,64%	0,074	0,93 / 0,91
Organisational performance	1,61	0,74	0,61%	0,081	0,93 / 0,95

TABLE III. RESULTS OF THE STRUCTURAL MODEL

Causal Paths (hypotheses)		*Estimate*	*p*	*Result*
H1	Top management support → ERP implementation success	0,26	0,000	Accepted
H2	Organizational culture → ERP implementation success	0,23	0,000	Accepted
H3	*External pressure → ERP implementation success*	-	*0,098*	*Rejected*
H4	Vendor support → ERP implementation success	0,36	0,011	Accepted
H5	*Project management → ERP implementation success*	-	*0,267*	*Rejected*
H6	Training → ERP implementation success	0,29	0,000	Accepted
H7	User involvement → ERP implementation success	0,26	0,000	Accepted
H8	Business Process Reengineering → ERP implementation success	0,35	0,000	Accepted
H9	*Implemented modules → ERP implementation success*	-	*0,164*	*Rejected*
H10	ERP implementation success → Organisational performance	0,34	0,003	Accepted

To test the construct validity, each research factor was evaluated: (a) for its unidimensionality and reliability (Table I), (b) for its goodness of fit to the proposed research model (Table II). The examination of the unidimensionality of each factor was conducted using Explanatory Factor Analysis (EFA) [71]. Moreover, 'Cronbach Alpha' was used for estimating each factor's reliability. All tests concluded that the scales used are valid and reliable.

The evaluation of the goodness of fit of each research factor to the proposed model was conducted using Confirmatory Factor Analysis (CFA). All tests produced satisfactory results (see Table II for the main results).

V. EMPIRICAL RESULTS

A. Model valuation

The examination of the proposed conceptual framework was conducted using the "Structural Equation Modeling" (SEM) technique [72] [73] [74]. To evaluate the fit of the overall model the chi-square value ($X^2 = 49,7$) and the p-value ($p = 0,000$) were estimated. These values indicate a satisfactory fit. However, the sensitivity of the X^2 statistic to the sample size enforces to control other supplementary measures of evaluating the overall model, such as the "Normed-X^2" index (3,1), the RSMEA index (0,077) the CFI (0,99) and the GFI (0,97), that all indicate a good fit.

B. Hypothesis testing

Seven hypotheses were found significant (H1, H2, H4, H6, H7, H8, H10), while three hypotheses were rejected by the empirical data (H3, H5, H9). After reviewing the empirical results, the following observations can be made:

A. The successful implementation of an ERP system has its roots on vendor support, training, and user involvement. These three factors were found to have the strongest impact on the main dependent factor of the present study (ERP implementation success). According to these empirical results, the present study proposes a mechanism that will drive implementation success. Various organisations may utilise this mechanism in order to experience a seamless implementation process. It includes three steps, each describing tasks that should be performed before, during and after the implementation of an ERP system.

Firstly, before the implementation, companies should spend their limited time and resources in selecting the appropriate software retailer. A good fit between the two seems to be very crucial for implementation success. Moreover, employees should be involved in the decision to adopt an ERP system. Executives should take employee attitudes and beliefs under serious consideration. In general, employees should feel like an integrated part of the whole process, while the adoption of the new system should not be understood as a decision that has been forced upon them. Only when employees fell like they have contributed to the implementation initiative, will they accept the changes that may occur. On a more practical level, the contribution of employees before the implementation is crucial for ensuring that the system will be designed in order to have a better fit with existing business practices.

Secondly, during the implementation period (that may be quite short, especially in micro-enterprises), vendors should adopt an analytical (linear) approach. Initially, the most

technological-ready employees should be selected in order to test the implemented system. Then, its advantages should be underlined and communicated amongst all personnel. After that, initial training should take place. The main goal is to initiate the system after all employees have been fully involved in the whole process.

Thirdly, after the implementation period, continuous training should be offered by the vendor (or another consultant). After all, the first month following the implementation of the new ERP system is extremely crucial. Employees should feel that the new system enhances their job, while resulting in many other organisational benefits.

B. No matter how important the role of vendor support (r=0,36), training (r=0,29), and user involvement (r=0,26), the support of top management has, also, been underlined as a significant antecedent of ERP implementation success (r=0,26). Without any doubt, executives should demonstrate their belief on the implemented system, mostly by ensuring its funding and setting the example for its use.

C. Moreover, the empirical data revealed that Business Process Reengineering (BPR) is a quite significant factor (r=0,35). This finding adds further support to the previous observations, arguing that BPR should be a priority for vendors, employees and executives.

D. Additionally, organizational culture (with emphasis on knowledge-sharing) affects implementation (r=0,23). This factor cannot be easily enhanced prior or during the implementation period, since its development is, usually, a result of the unique history of the organisation.

Finally, the relationship between ERP implementation success and organisational performance has been verified by the empirical data (r=0,34). Concerning the strength of that relationship (R^2=26%, including direct and indirect effects), it should be noted that when examining complex phenomena, like organisational performance, even a relatively small predictive power seems to be satisfactory.

VI. Conclusions

The present study was motivated by specific gaps that were recognised in the relevant literature of the specific field. In order to cover these gaps, the present study used an extensive literature review and qualitative data (focus group sessions with managers of SMEs) in order to develop a conceptual framework that investigated the antecedents of ERP implementation success. Moreover, this framework was tested with the use of a newly-developed structured questionnaire (quantitative data) on a sample of Greek SEMs that have implemented and ERP system.

That specific approach offered certain advantages: focus groups offered practical knowledge concerning the factors with the most significant impact on ERP implementation, while the quantitative research revealed which of these factors are actually significant. The contribution of the study lies on this enhanced approach. More specifically, it offers the necessary ground for comparison and replication. Its conceptual framework may be replicated from future studies, while other scientists may employ its twofold approach as a basis for their future empirical investigation.

The proposed conceptual framework of the study included nine antecedents of ERP implementation success. These factors are perceived as Critical Success Factors (CSFs) for successful ERP implementation. Empirical data were analysed using the "Structural Equation Modeling" (SEM) technique, while the validity and the reliability of all research factors were evaluated with the use of enhanced statistical methods (EFA, CFA).

According to the results of the statistical analysis, six of the antecedents included in the research model of the present study were found to have a direct (positive) effect on successful ERP implementation. Additionally, the predictive power of the proposed model was found to be very satisfactory. More specifically, the six antecedents can explain the variance of ERP successful implementation by 72% (R^2 = 0,72). On the other hand, three research factors (external pressure, project management, implemented modules) were not found to have an effect on the successful implementation of an ERP system.

Therefore, it is concluded that when implementing an ERP system, organisations should focus on the following six factors: Top management support, Organizational culture, Vendor support, Training, User involvement, Business Process Reengineering. The present study argues that the enhancement of these Critical Success Factors should be conducted before, during and after ERP implementation. Partial focus will only limit their positive effect.

In general, it is concluded that ERP implementation success is a result of intangible factors (organisational culture), people-related factors (vendor support, training, user involvement), and proper leadership (top management support). Moreover, reengineering, a more practical issue of implementation, is also a prerequisite for success.

Previous studies conducted in other geographical regions of the European continent (e.g. Eastern and Central Europe) have found similar results. For example, Ziemba and Kolasa [61] found that top management support, user involvement and process management have an impact on information systems projects, while Bradley [62] concluded that the determinants of enterprise system adoption success are user involvement, user empowerment, system reliability and cooperation with the system supplier (vendor).

The present study is somehow limited by the poor definition of its population. This limitation is inherent to all studies of the field, since a complete list of companies that have implemented an ERP system can not be found in most databases. Further research is suggested with larger samples that would, probably, offer more information and strengthen the results of the present study. Moreover, it would be interesting to examine more factors and gather primary data from all company personnel, so as to achieve a more complete view of the subject under investigation.

References

[1] C. Leyh, "Critical success factors for ERP projects in small and medium-sized enterprises-The perspective of selected German SMEs", IEEE Federated Conference on Computer Science and Information Systems (FedCSIS), pp. 1181-1190, 2014.

[2] C. Spathis, and S. Constantinides, "Enterprise resource planning systems' impact on accounting processes", Business Process Management Journal, Vol. 10, No. 2, pp. 234-247, 2004, http://dx.doi.org/10.1108/14637150410530280.

[3] M. Łobaziewicz, "Integration of B2B system that supports the management of construction processes with ERP systems", IEEE Federated Conference on Computer Science and Information Systems (FedCSIS), pp. 1461-1466, 2015.

[4] M.T. Tsai, E.Y. Li, K.W. Lee, and W.H. Tung, "Beyond ERP implementation: the moderating effect of knowledge management on business performance", Total Quality Management, Vol. 22, No. 2, pp. 131-144, 2011, http://dx.doi.org/10.1080/14783363.2010.529638.

[5] H.R. HassabElnaby, W. Hwang, and M.A. Vonderembse, "The impact of ERP implementation on organizational capabilities and firm performance", Benchmarking: an International Journal, Vol. 19, No. 4/5, pp. 618-633, 2012, http://dx.doi.org/10.1108/14635771211258043.

[6] P. Ifinedo, B. Rapp, A. Ifinedo, and K. Sundberg, "Relationships among ERP post-implementation success constructs: an analysis at the organizational level", Computers in Human Behavior, Vol. 26, No. 5, pp. 1136-1148, 2010, http://dx.doi.org/ 10.1016/j.chb.2010.03.020.

[7] V.A. Mabert, A. Soni, and M.A. Venkatraman, "The impact of organizational size on enterprise resource planning (ERP) implementation in the US manufacturing sector", Omega, Vol. 31, pp. 235-246, 2003, http://dx.doi.org/10.1016/S0305-0483(03)00022-7.

[8] P. Garg, and A. Garg, "Factors influencing ERP implementation in retail sector: an empirical study from India", Journal of Enterprise Information Management, Vol. 27, No. 4, pp. 424-448, 2014, http://dx.doi.org/10.1108/JEIM-06-2012-0028.

[9] D. Zabjek, A. Kovacic, and M. Indihar Stemberger, "Business process management as an important factor for a successful ERP system implementation", Ekonomska istrazivanja, Vol. 21, No. 4, pp. 1-18, 2008.

[10] S.W. Chien, H.C. Lin, and C.T. Shih, "A Moderated Mediation Study: Cohesion Linking Centrifugal and Centripetal Forces to ERP Implementation Performance", International Journal of Production Economics, Vol. 158, pp. 1-8, 2014, http://dx.doi.org/10.1016/j.ijpe.2014.06.001.

[11] A.I. Nicolaou, and S. Bhattacharya, "Organizational performance effects of ERP systems usage: the impact of post-implementation changes", International Journal of Accounting Information Systems, Vol. 7, No. 1, pp. 18-35, 2006, http://dx.doi.org/10.1016/j.accinf.2005.12.002.

[12] J.R. Muscatello, and I.J. Chen, "Enterprise resource planning (ERP) implementations: theory and practice", International Journal of Enterprise Information Systems, Vol. 4 No. 1, pp. 63-78, 2008.

[13] D. Lee, S.M. Lee, D.L. Olson, and S. Hwan Chung, "The effect of organizational support on ERP implementation", Industrial Management & Data Systems, Vol. 110, No. 2, pp. 269-283, 2010, http://dx.doi.org/10.1108/02635571011020340.

[14] Z. Zhang, M.K. Lee, P. Huang, L. Zhang, and X. Huang, "A framework of ERP systems implementation success in China: an empirical study", International Journal of Production Economics, Vol. 98, No. 1, pp. 56-80, 2005, http://dx.doi.org/ 10.1016/j.ijpe.2004.09.004.

[15] S. Saini, S. Nigam, and S.C. Misra, "Identifying success factors for implementation of ERP at Indian SMEs: a comparative study with Indian large organizations and the global trend", Journal of Modelling in Management, Vol. 8, No. 1, pp. 103-122, 2013, http://dx.doi.org/10.1108/17465661311312003.

[16] D. Zabjek, A. Kovacic, and M. Indihar Stemberger, "The influence of business process management and some other CSFs on successful ERP implementation", Business Process Management Journal, Vol. 15, No. 4, pp. 588-608, 2009, http://dx.doi.org/10.1108/14637150910975552.

[17] R.D. Samuel, and S. Kumar, "Prediction of ERP Success before the Implementation", in International Asia Conference on Industrial Engineering and Management Innovation (IEMI2012) Proceedings, pp. 219-227, 2013, http://dx.doi.org/ 10.1007/978-3-642-38445-5_22.

[18] E.J. Umble, and M.M. Umble, "Avoiding ERP implementation failure", Industrial Management, Vol. 44, No. 1, pp. 25-33, 2002.

[19] J. Ram, D. Corkindale, and M.L. Wu, "Implementation critical success factors (CSFs) for ERP: do they contribute to implementation success and post-implementation performance?", International Journal of Production Economics, Vol. 144, No. 1, pp. 157-174, 2013, http://dx.doi.org/10.1016/j.ijpe.2013.01.032.

[20] P. Garg, and A. Chauhan, "Factors affecting the ERP implementation in Indian retail sector", Benchmarking: an International Journal, Vol. 22, No. 7, pp. 1315-1340, 2015, http://dx.doi.org/10.1108/BIJ-11-2013-0104.

[21] P. Garg, and D. Agarwal, "Critical success factors for ERP implementation in a Fortis hospital", Journal of Enterprise Information Management, Vol. 27, No. 4, pp. 402-423, 2014, http://dx.doi.org/10.1108/JEIM-06-2012-0027.

[22] J. Ram, and D. Corkindale, "How "critical" are the critical success factors (CSFs)? Examining the role of CSFs for ERP", Business Process Management Journal, Vol. 20, No. 1, pp. 151-174, 2014, http://dx.doi.org/10.1108/BPMJ-11-2012-0127.

[23] J.F. Rockart "A new approach to defining the chief executive's information needs", MIT Working Paper, CISR 37, No. 1008-78, 1978.

[24] R. Saade, and H. Nijher, "Critical success factors in enterprise resource planning implementation: a review of case studies", Journal of Enterprise Information Management, Vol. 29, No. 1, pp. 72-96 , 2016, http://dx.doi.org/10.1108/JEIM-03-2014-0028.

[25] A. Boynton, and R. Zmud, "An assessment of critical success factors", Sloan Management Review, Vol. 25, No. 4, pp. 17-27, 1984.

[26] P. Bingi, M.K. Sharma, and J.K. Godla, "Critical issues affecting an ERP implementation", IS Management, Vol. 16, No. 3, pp. 7-14, 1999, http://dx.doi.org/10.1201/1078/43197.16.3.19990601/31310.2.

[27] K.K. Hong, and Y.G. Kim, "The critical success factors for ERP implementation: an organizational fit perspective", Information & Management, Vol. 40, No. 1, pp. 25-40, 2002, http://dx.doi.org/10.1016/S0378-7206(01)00134-3.

[28] J. Motwani, R. Subramanian, and P. Gopalakrishna "Critical factors for successful ERP implementation: exploratory findings from four case studies", Computers in Industry, Vol. 56, No. 6, pp. 529-544, 2005, http://dx.doi.org/10.1016/j.compind.2005.02.005.

[29] L. Li, C. Markowski, L. Xu, and E. Markowski, "TQM - A predecessor of ERP implementation", International Journal of Production Economics, Vol. 115, No. 2, pp. 569-580, 2008, http://dx.doi.org/10.1016/j.ijpe.2008.07.004.

[30] S.W. Chou, I.H. Hung, and Y.C. Chang, "Understanding the antecedents of ERP Implementation success - The Perspective of Knowledge Transfer", Asia Pacific Management Review, Vol. 18, No. 3, pp. 301-322, 2013.

[31] A. Amid, M. Moalagh, and A.Z. Ravasan, "Identification and classification of ERP critical failure factors in Iranian industries", Information Systems, Vol. 37, No. 3, pp. 227-237, 2012, http://dx.doi.org/10.1016/j.is.2011.10.010.

[32] S. Wee, "Juggling toward ERP success: keep key success factors high", ERP News, February, available at: www.erpnews.com/erpnews/erp904/ 02get.html, 2000.

[33] A. Scheer, and F. Habermann, "Enterprise resource planning: making ERP a success", Communications of the ACM, Vol. 43, No. 4, pp. 57-61, 2000, http://dx.doi.org/10.1145/332051.332073.

[34] T.W. Ferratt, S. Ahire, and P. De, "Achieving success in large projects: implications from a study of ERP implementations", Interfaces, Vol. 36, No. 5, pp. 458-469, 2006.

[35] E.W.T. Ngai, C.C.H. Law, and F.K.T. Wat, "Examining the critical success factors in the adoption of enterprise resource planning", Computers in Industry, Vol. 59, No. 6, pp. 548-564, 2008, http://dx.doi.org/10.1016/j.compind.2007.12.001.

[36] C. Sheu, B. Chae, and C.-L. Yang, "National differences and ERP implementation: issues and challenges", Omega, Vol. 32, No. 5, pp. 361-371, 2004, http://dx.doi.org/10.1016/j.omega.2004.02.001.

[37] M. Tarafdar, and R.K. Roy, "Analyzing the adoption of enterprise resource planning systems in Indian organizations: a process framework", Journal of Global Information Technology Management, Vol. 6, No. 1, pp. 21-51, 2003, http://dx.doi.org/10.1080/1097198X.2003.10856342.

[38] C.K. Lee, H.H. Lee, and M. Kang, "Successful implementation of ERP systems in small businesses: a case study in Korea", Service Business, Vol. 2, No. 4, pp. 275-286, 2008, http://dx.doi.org/10.1007/s11628-008-0045-3.

[39] E. Bernroider, and S. Koch, "ERP selection process in midsize and large organizations", Business Process Management Journal, Vol. 7,

No. 3, pp. 251-257, 2001, http://dx.doi.org/10.1108/14637150110392746.
[40] Y. Yusuf, A. Gunasekaranb, and M.S. Abthorpe, "Enterprise information systems project implementation: a case study of ERP in Rolls-Royce", International Journal of Production Economics, Vol. 87, pp. 251-266, 2004, , http://dx.doi.org/10.1016/j.ijpe.2003.10.004.
[41] P. Poba-Nzaou, L. Raymond, and B. Fabi, "Adoption and risk of ERP systems in manufacturing SMEs - a positivist case study", Business Process Management Journal, Vol. 14, No. 4, pp. 530-550, 2008, http://dx.doi.org/10.1108/14637150810888064.
[42] P. Liamputtong, Focus group methodology: Principle and practice, Sage Publications, London, UK, 2011.
[43] B.L. Berg, H. Lune, and H. Lune, Qualitative research methods for the social sciences, Pearson, MA: Boston, USA, 2004.
[44] R. Sharma, and P. Yetton, "The contingent effects of management support and task interdependence on successful information systems implementation", MIS Quarterly, Vol. 27, No. 4, pp. 533-555, 2003.
[45] M. Al-Mashari, "A process change-oriented model for ERP application", International Journal of Human-computer Interaction, Vol. 16, No. 1, pp. 39-55, 2003, http://dx.doi.org/10.1207/S15327590IJHC1601_4.
[46] J.E. Umble, R.R. Haft, and M.M. Umble, "Enterprise resource planning: implementation procedures and critical success factors", European Journal of Operational Research, Vol. 146, No. 2, pp. 241-257, 2003, http://dx.doi.org/10.1016/S0377-2217(02)00547-7.
[47] M. Al-Mashari, A. Al-Mudimigh, and M. Zairi, "Enterprise resource planning: a taxonomy of critical factors", European Journal of Operational Research, Vol. 146, No. 2, pp. 352-364, 2003, http://dx.doi.org/10.1016/S0377-2217(02)00554-4.
[48] M.T. Somers, and K.G. Nelson, "A taxonomy of players and activities across the ERP project life cycle", Information & Management, Vol. 41, No. 3, pp. 257-278, 2004, http://dx.doi.org/ 10.1016/S0378-7206(03)00023-5.
[49] S. Dezdar, and S. Ainin, "Examining ERP implementation success from a project environment perspective", Business Process Management Journal, Vol. 17, No. 6, pp. 919-939, 2011, http://dx.doi.org/10.1108/14637151111182693.
[50] L.C.S. Koh, M. Simpson, J. Padmore, N. Dimitriadis, and F. Misopoulos, "An exploratory study of enterprise resource planning adoption in Greek companies", Industrial Management & Data Systems, Vol. 106, No. 7, pp. 1033-1059, 2006, http://dx.doi.org/10.1108/02635570610688913.
[51] C.M. Jones, M. Cline, and S. Ryan, "Exploring knowledge sharing in ERP implementation: an organizational culture framework", Decision Support System, Vol. 41, No. 2, pp. 411-434, 2006, http://dx.doi.org/10.1016/j.dss.2004.06.017.
[52] L.S. Jarvenpaa, and S.D. Staples, "Exploring perceptions of organizational ownership of information and expertise", Journal of Management Information Systems, Vol. 18, No. 1, pp. 151-183, 2001, http://dx.doi.org/10.1080/07421222.2001.11045673.
[53] P.C. Ruppel, and J.S. Harrington, "Sharing knowledge through intranets: a study of organizational culture and intranet implementation", IEEE Transactions on Professional Communication, Vol. 44, No. 1, pp. 37-52, 2001, http://dx.doi.org/ 10.1109/47.911131.
[54] H. Liang, N. Saraf, Q. Hu, and Y. Xue, "Assimilation of enterprise systems: the effect of institutional pressures and the mediating role of top management", MIS Quarterly, Vol. 31, No. 1, pp. 59-87, 2007.
[55] E. Waarts, Y.M. van Everdingen, and J. van Hillegersberg, "The dynamics of factors affecting the adoption of innovations", The Journal of Product Innovation Management, Vol. 19, No. 6, pp. 412-423, 2002, , http://dx.doi.org/10.1111/1540-5885.1960412.
[56] M. Bradford, and S. Richtermeyer, "Realizing value in ERP", Cost Management, Vol. 16, No. 2, pp. 13-19, 2002.
[57] G.T.E. Wang, C.-L.C. Lin, J.J. Jiang, and G. Klein, "Improving enterprise resource planning (ERP) fit to organizational process through knowledge transfer", International Journal of Information Management, Vol. 27, No. 3, pp. 200-212, 2007, http://dx.doi.org/10.1016/j.ijinfomgt.2007.02.002.
[58] S. El Sawah, A. Abd El Fattah Tharwat, and M. Hassan Rasmy, "A quantitative model to predict the Egyptian ERP implementation success index", Business Process Management Journal, Vol. 14, No. 3, pp. 288-306, 2008 http://dx.doi.org/10.1108/14637150810876643.
[59] G.T.E. Wang, and F.H.J. Chen, "Effects of internal support and consultant quality on the consulting process and ERP system quality", Decision Support Systems, Vol. 42, No. 2, pp. 1029-1041, 2006, http://dx.doi.org/10.1016/j.dss.2005.08.005.
[60] S.C.L. Koh, M. Simpson, J. Padmore, N. Dimitriadis, and F. Misopoulos, "An exploratory study of enterprise resource planning adoption in Greek companies", Industrial Management & Data Systems, Vol. 10, No. 7, pp. 1033-1059, 2006, http://dx.doi.org/10.1108/02635570610688913.
[61] S. Dezdar, and S. Ainin, "The influence of organizational factors on successful ERP implementation", Management Decision, Vol. 49, No. 6 pp. 911-926, 2011, http://dx.doi.org/10.1108/00251741111143603.
[62] J. Bradley, "Management based critical success factors in the implementation of Enterprise Resource Planning systems", International Journal of Accounting Information systems, Vol. 9, No. 3, pp. 175-200, 2008, http://dx.doi.org/10.1016/j.accinf.2008.04.001.
[63] F.H. Nah, K.M. Zuckweiler, and L.S. Lau, "ERP Implementation: chief information officers' perceptions of critical success factors", International Journal of Human-Computer Interaction, Vol. 16, No. 1, pp. 5-22, 2003, http://dx.doi.org/ 10.1207/S15327590IJHC1601_2.
[64] E.J. Ettlie, J.V. Perotti, A.D. Joseph, and J.M. Cotteleer, "Strategic predictors of successful enterprise systems deployment", International Journal of Operations & Production Management, Vol. 25, No. 10, pp. 953-972, 2005, http://dx.doi.org/10.1108/01443570510619473.
[65] A. Madapusi, and D. D'Souza, "The influence of ERP system implementation on the operational performance of an organizational", International Journal of Information Management, Vol. 32, No. 1, pp. 24-34, 2012, http://dx.doi.org/10.1016/j.ijinfomgt.2011.06.004.
[66] T.M. Yeh, C.C. Yang, and W.T. Lin, "Service quality and ERP implementation: a conceptual and empirical study of semiconductor-related industries in Taiwan", Computers in Industry, Vol. 58, No. 8-9, pp. 844-854, 2007, http://dx.doi.org/ 10.1016/j.compind.2007.03.002.
[67] A.V. Mabert, A. Soni, and A.M. Venkataramanan, "Enterprise resource planning: measuring value", Production and Inventory Management Journal, Vol. 42, No. 3-4, pp. 46-51, 2001.
[68] H.C.C. Law, and T.W.E. Ngai, "ERP systems adoption: an exploratory study of the organizational factors and impacts of ERP success", Information & Management, Vol. 44, , pp. 418-432, 2007, http://dx.doi.org/ 10.1016/j.im.2007.03.004.
[69] M. Gupta, and A. Kohli, "Enterprise resource planning systems and its implications for operations function", Technovation, Vol. 26, No. 5-6, pp. 687-696, 2006, http://dx.doi.org/10.1016/j.technovation.2004.10.005.
[70] K. Laframboise, and F. Reyes), "Gaining competitive advantage from integrating enterprise resource planning and total quality management", Journal of Supply Chain Management, Vol. 41, No. 3, pp. 49-64, 2005, http://dx.doi.org/ 10.1111/j.1055-6001.2005.04103005.x.
[71] F. Hair, R. Anderson, R. Tatham, and W. Black, Multivariate Data Analysis with Readings, London: Prentice-Hall International, 1995.
[72] R.E. Schumacker, and R.G. Lomax, A Beginner's Guide to Structural Equation Modeling, New York: Routledge Academic, 2010.
[73] E.K. Kelloway, Using LISREL for Structural Equation Modeling: A Researcher's Guide, Thousand Oaks, CA: Sage, 1998.
[74] B.M. Byrne, Structural equation modeling with AMOS: Basic concepts, applications, and programming. UK: Routledge, 2013.
[75] E. Ziemba and I. Kolasa, "Risk factors framework for information systems projects in public organizations-insight from Poland", 2015 Federated Conference on Computer Science and Information Systems (FedCSIS), IEEE, pp. 1575-1583, 2015, http://dx.doi.org/ 10.15439/2015F110.
[76] P. Soja, "Understanding determinants of enterprise system adoption success: Lessons learned from full-scope projects in manufacturing companies", Production Planning & Control, Vol. 21, No. 8, pp. 736-750, 2010, http://dx.doi.org/ 10.1080/09537281003601597.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1253–1262
DOI: 10.15439/2016F1
ACSIS, Vol. 8. ISSN 2300-5963

Antecedents and outcomes of ERP implementation success

Prodromos Chatzoglou
Democritus University of Thrace,
Department of Production and Management Engineering,
Vasillisis Sofias 12, 67100,
Xanthi, Greece
Email: pchatzog@pme.duth.gr

Dimitrios Chatzoudes
Democritus University of Thrace,
Department of Production and Management Engineering,
Vasillisis Sofias 12, 67100,
Xanthi, Greece
Email: dchatzoudes@yahoo.gr

Georgia Apostolopoulou
Hellenic Open University,
School Of Social Sciences,
Parodos Aristotelous 18, 26335,
Patra, Greece
Email: geapostolo84@gmail.com

***Abstract*—Enterprise Resource Planning (ERP) systems have established a reputation in the world of business as indispensables tools that integrate all departments and functions across a company into a single computer system. However, implementing an ERP system does not always result in enhanced organizational performance. In order to ensure successful implementation, companies should study the critical factors having an impact on the whole procedure. In this context, the present study proceeds in developing and testing an original conceptual framework (research model), which explores the factors having an impact on ERP implementation success (internal environment, technology-related issues, implementation team, end-users), as well as the impact of the implementation itself on organisational performance. The proposed conceptual framework was tested, using a newly-developed structured questionnaire, in a sample of 204 Greek companies that have already implemented an ERP system. The Structural Equation Modelling (SEM) technique was used in order to test the research hypotheses.**

I. Introduction

THE explosive growth of Information and Communication Technologies (ICTs) (Information Systems, Enterprise Application Systems, Internet Technologies) has influenced, to a great extent, the way modern organizations operate in the business environment [1] [2]. Technological advances make information easily accessible, providing greater awareness of international economic opportunities [3].

Enterprise Resource Planning (ERP) systems are software systems which can be customized to integrate the business processes of a company, in such a way, that they are visible and accessible by the management in real-time [4]. An ERP system, designed to serve the modern sophisticated management, enables its users to coordinate key business practices across functions more efficiently, collect corporate data more holistically and offer optimal control over the operations of the organisation [5].

When an ERP system is successfully implemented, it promises to manage and integrate all business processes and functions within an organization [6] [7]. The integration, brought by ERP implementation, helps organizations to increase and improve their overall market position, in order to gain competitiveness in a rapidly changing business environment [8]. It also helps different divisions share data and knowledge, reduce costs and improve management of business processes [9] [10].

Due to the potential benefits of ERP systems, most of the organizations invested both time and money in their implementation [11]. However, ERPs have a reputation of costing a lot of money and providing limited results [10]. Some of the causes, cited in the relevant literature, for failed ERP projects include: poor project management planning, lack of business management support, unexpected return on investments, insufficient education and training, and, finally, weakness to redesign business processes [12] [13] [14].

Successful ERP implementation is quite valuable to many organizations, as it provides with numerous benefits. This explains why an ERP system is generally considered to be a vital component for enhancing organizational performance [15] [16] [17] [18] [19] [20] [21].

The ERP literature includes various studies investigating the factors having an impact on the effective implementation of ERP systems [8] [22]. On the other hand, there are fewer studies examining the impact of ERPs on different measures of business success [5] [15] [21] [23]. Despite that, the literature review analysis that has been conducted, failed to identify any empirical studies adopting a multidimensional approach, incorporating both antecedents and outcomes.

The present study aspires to bridge that gap in the relevant literature, developing and testing a three-dimensional conceptual framework (research model). More specifically, the first dimension includes the antecedents of ERP successful implementation (internal environment, technology-related issues, implementation team, end-users), the second dimension the implementation itself (information quality, system quality, service quality), while the third dimension includes three measures of organisational performance (internal efficiency, competitiveness, profitability).

Therefore, the main objective of the study is to identify the factors that drive ERP implementation success, and, consequently, measure the effect of implementation on organization performance. The measurement of each of the

This work was not supported by any organization.

three dimensions is being conducted with the use of multiple measures (sub-factors).

The examination of the proposed conceptual framework was made with the use of a newly-developed structured questionnaire that was distributed to a group of Greek companies. The Structural Equation Modelling (SEM) technique was used in order to test the research hypotheses. The present study is empirical (it is based on primary data), explanatory (examines cause and effect relationships), deductive (tests research hypotheses) and quantitative (analyses quantitative data collected with the use of a structured questionnaire). Its results may be useful for managers, business analysts and IT analysts in dealing with the implementation of ERP systems.

The following section includes a review of the literature; section three presents the conceptual framework of the study, while section four includes the research methodology. Results and conclusions are discussed in sections 5 and 6 respectively.

II. Literature review

An ERP system entails long-term application processes (strategic level) and short-term applications (operational level). As a consequence, it is a given fact that ERP implementation cannot be imposed overnight [24]. On the contrary, it involves constant modifications and gradual implementation, so that in the end it will meet specific needs and bring out desirable end results [25].

Several studies have attempted to describe the critical success factors associated with the implementation and final use of ERP systems [24]. For example, Hong and Kim [26] identified the fit between the ERP system and the organizational climate as a crucial predictor of implementation success. They, also, argued that ERP implementation affects most of the business processes, and influences users directly [26].

Since ERP implementation will, most likely, affect the whole organization, reengineering of business processes is also required for business success [14]. Al-Mashari, Al-Mudimigh and Zairi [14] developed a theoretically and practically grounded taxonomy of ERP critical success factors. They argued that critical factors should gap the bridge between ERP implementation and (a) business processes, (b) Information Technology (IT), (c) structure, (d) culture and management systems, and (e) strategy.

Motwani, Subramanian and Gopalakrishna [27] conducted a comparative case study of four firms that implemented an ERP system, arguing that an evolutionary implementation process, which is supported by careful change management, network relationships and cultural readiness, can lead to successful implementation. Somers and Nelson [28] listed 22 critical success factors and categorized them according to the stage of implementation. They argued that package selection is among the most significant factors.

Ram, Wu and Tagg [29] built a conceptual model exploring the impact of two critical success factors (training and education, and system integration activities) on ERP implementation effectiveness.

Ifinedo and Nahar [30] found out that system quality and information quality are considered as the most important dimensions in the assessment of ERP success. Similarly, Chien and Tsaur [31] concluded that system quality, service quality and information quality seem to be the most important factors when implementing an ERP system. Similar lists of factors have been proposed by several other authors [10] [32].

Esteves and Pastor [33] made a distinction between the strategic and tactical factors, and the technological and organizational factors of implementation. Holland, Light and Gibson [34] grouped critical success factors into two categories: the strategic factors that span during the whole implementation project, and tactical factors that can be applied to particular parts of the project.

Loh and Koh [35] presented a framework of critical success factors in small and medium sized enterprises (SMEs), concluding that the most important factors are: project champion, project management, business plan and vision, and top management support [35]. Hustad and Olsen [36] concluded that SMEs have different challenges than larger organisations, mostly because of their limited resources and competencies [36].

Despite numerous implementation challenges and high implementation costs, ERP systems have become popular, and both small and large companies implement ERP systems in order to remain competitive. However, in some cases, ERP implementation can be very risky and, if organizations do not pay much attention to their limitations and requirements, results may be very unsatisfactory [12] [37].

According to Somers and Nelson [28], everyone who uses ERP systems needs to be trained on how they work and how they relate to business process, early on the implementation phase. Inadequate training and education could be considered as a significant reason for many ERP projects failures [10] [14]. Additionally, proper ‘package’ selection plays a vital role in successful implementation of ERPs, as it is one of the most important steps [10] [28].

During ERP implementation, business process reengineering should take place, in order to take full advantage of the new system [14] [34]. In addition, getting people educated, or trained, and keeping them informed throughout the whole implementation process should be amongst the first organisational priorities [28]. Moreover, through intensive guided learning, superior training and special knowledge activities, ERP consultants can help clients acquire the necessary knowledge for successful implementation [38]. Since users are an integral part of the attempted changes, the ones that will participate in the design and the implementation of new administration procedures must be sufficiently trained [34].

As demonstrated above, a significant number of studies on ERP implementation have been conducted. Many of these studies investigate the critical success factors for ERP implementation, and assist practitioners towards selecting the most suitable ERP software. Moreover, the studies of Al-Mashari, Al-Mudimigh and Zairi [14] and Umble, Haft and Umble [10] appear to be among the most cited papers in the ERP literature (423 and 636 citation respectively: data acquired from the 'Scopus' database, November 2015).

The present study conducted an analytical review of the ERP literature, and developed a conceptual framework that is a synthesis of previous contributions. More specifically, a list including the factors that have been used in order to predict ERP implementation success has been developed, while the most significant factors were, finally, incorporated in the proposed conceptual framework. In comparison with previous studies, the present empirical research does not only examine the antecedents, but, also, includes the outcomes of ERP implementation in its analysis.

III. Conceptual framework

Successful ERP implementation requires the coordination of many activities and a close cooperation between managers, employees, IT specialists, business analysts and consultants [39]. Therefore, any theoretical framework concerning ERP implementation effectiveness should include various dimensions in its analysis.

Based on the literature review analysis that was conducted prior to the development of the conceptual framework, the present study classified the antecedents of ERP implementation into four distinct categories (dimensions): (1) internal environment, (2) technology-related issues, (3) implementation team, (4) end-users. Additionally, each category (dimension) was determined with the use of several factors.

The selection of all research factors was a result of a specific procedure: (1) the Scopus database was used in order to identify previous studies concerning the antecedents of ERP implementation success (65 relevant studies were identified), (2) an extensive list, including the factors used in these studies, was constructed, (c) factors were given a significance index, based on the findings of each study, (d) each factor was categorized into one of the four pre-determined categories (dimensions), (e) the factors with the highest significance index in each category were, finally, selected.

The present study adopts a unique approach on the ERP literature. Instead of, only, examining the antecedents of ERP implementation success, the outcomes of the implementation procedure were, also, taken into consideration. Thus, an original three-dimensional conceptual framework (research model) was developed (see Fig. 1).

A. Internal environment

Successful implementation cannot be achieved, only by managers. Effective leadership is needed in order to achieve the desirable goals [35]. Organizations should, also, review their organizational culture and attitude towards change, before implementing an ERP system [40]. In this study, the dimension of "internal environment" includes four factors (top management support, business process reengineering, organizational culture, change management).

Top management support has been emphasized as a crucial factor in successful ERP implementation by previous studies [10] [14] [41]. Al-Mashari, Al-Mudimigh and Zairi [14] suggested that top management support should not only be offered during the initiation and facilitation stage, but throughout the entire ERP implementation process. Umble, Haft and Umble [10] claimed that successful ERP implementation requires the commitment and constant participation of top management. In a different case, the project is most likely to fail, or fail to deliver the full range of forecasted benefits [11] [12] [42].

Business process reengineering (BPR) has been, often, proposed as a critical success factor for ERP implementation [29]. Typically, BPR is carried out in order to restructure non-value-adding operations, reduce the complexity of business processes and eliminate inefficient processes [43]. Reengineering aims at making the necessary adjustments in order to take full advantage of the new processes offered by the ERP system [44]. Therefore, organizations should be willing to adjust their processes, so as to fit with the new software and minimize the degree of customization needed [34]. Most experts agree that software customization results in higher implementation costs and longer implementation period. Therefore, companies should keep the ERP package "as it is", as much as possible, and reengineer their business processes to conform to the package [45].

An organisational culture of shared values and common objectives is crucial for business success [46]. Organizations should built an organizational culture that is open to change [32], since openness to change plays a pivotal role in today's business environment. When organisational members have different cultures, beliefs and values, they, also, have different perceptions on various organizational changes [47]. In other words, organizational culture is a critical success factor for a project that requires significant changes [48]. Consequently, when the culture of an organisation is prone to change, ERP implementation is made quite easier.

Finally, Motwani, Subramanian and Gopalakrishna [27] argued that ERP projects that are supported by top management, but are not accompanied by adequate change management strategies are likely to fail. On the contrary, an implementation process backed with change management strategies and network relationships has been found to have a positive effect on implementation success [49]. Change management is a primary concern for many organizations involved in ERP implementation [50]. Research has shown that effective change management is critical to successful implementation [28]. Therefore, it is hypothesised:

H1: Internal environment has a positive impact on ERP implementation success.

B. Technology-related issues

During ERP implementation, various technological-related issues need to be addressed. More specifically, (a) the appropriate ERP package should be carefully selected [28]; (b) the overall support provided by the vendor should be taken under serious consideration [51]; and (c) the fit between the implemented system and the technological infrastructure of the organization should be examined [52]. The present study posits that all these dimensions (package selection, consultant (vendor) support, IT infrastructure) have a cumulative effect on ERP implementation success.

The selection of an ERP system, being among the first steps of implementation, appears as a critical factor [6]. After all, the package that will be selected will determine, to a great extend, the success of the project [10] [14]. Despite the fact that almost all ERP packages can be customized according to specific needs, customisation is very expensive and, usually, problematic [53]. Choosing the ERP package that best suits organisational needs and processes is critical to ensure successful implementation [6] [28].

However, no matter how important is to select a compatible ERP package, organizations should, also, select the appropriate vendor that will be able to offer full support. Troubleshooting is necessary for ERP implementation, so as to prepare for unexpected circumstances or, even, crises. This is an ongoing process, since the vendor (consultant) is obligated to assist in all stages on the implementation process [14] [32]. ERP adopting companies have to work closely with ERP vendors in order to determine possible software problems.

Finally, the appropriate IT infrastructure is necessary for the implementation of an ERP system [54]. Since most of ERP transactions are conducted in real-time, a reliable intranet, or local area network needs to be in place [32]. A company with a satisfactory level of IT infrastructure can be expected to implement new technologies, like ERP systems, more successfully than other companies, with low degree of IT readiness. From all the above, it is hypothesised:

H2: Technology has a positive impact on ERP implementation success.

C. Implementation Team

An effective implementation team is a key factor for every successful ERP project. The ERP team should consist of the best people in the organization [8] [22] [32]. At the same time, the implementation strategy is, also, critical, since it will assist the entire organization to adjust its business processes. In addition, a successful ERP implementation process requires excellent project management. Implementation team, the third antecedent of ERP implementation success, includes these three sub-factors (project team, project management, implementation strategy).

First of all, it must be made clear that an ERP implementation project involves all the departments of an organization [55]. According to Bhatti [54], the ERP project team includes: employees, managers, IT personnel, top management, the ERP vendor, and management consultants. Selecting the right employees to take part in the implementation process is critical for its success. The success of ERP projects is related to the skills, knowledge, abilities and experiences of project team members [55] [56].

As stated above, ERP implementation is a multi-level task, involving all business activities, and, often, requiring between one and two years of continuous effort [19] [20]. Therefore, an effective project management strategy should be in place in order to control the whole implementation process. ERP project management includes a clear definition of implementation objectives, the development of both work and resource plans, and a detailed tracking of project progress [34] [57].

Also, an ERP implementation strategy determines how the transfer from the legacy system to the new ERP system will be organised. Adopting an efficient strategy is of vital importance, since strategy sets the whole framework of implementation [58]. Without proper strategy, the whole implementation project is very likely to fail [58] [59].

It is hypothesised that all the above will have a cumulative positive effect on ERP implementation:

H3: Implementation team has a positive impact on ERP implementation success.

D. End-Users

Previous research has shown that no IT-based innovation can be successfully implemented without employee participation [60]. The attitude of end-users toward the ERP system has an impact on implementation success [61]. Characteristics of end-users have been identified as predictors of ERP implementation success, since the full benefits of the ERP system cannot be utilised until end-users are using it properly [10]. In this study, the dimension of "end-users" includes three factors (user involvement, training and education, employee skills).

User involvement is one of the most cited critical success factors in ERP implementation projects [14] [33]. Participation in the ERP implementation process raises the understanding of the new system and helps achieving better use [62]. Despite the level of training employees get during the implementation process [63], their involvement during the whole process is a very critical factor [10].

ERP requires a critical mass of employee knowledge in order to solve real problems within the company [18] [22] [64]. Everyone who uses the ERP system should be trained and educated on how the system works and how it can be used in everyday operations [28]. Organizations should provide training opportunities, on a regular basis, in order to improve the skills and knowledge of their employees. Sufficient training and education can increase the probability of ERP implementation success [32] [45].

Successful ERP implementation demands the constant cooperation of business experts, internal staff and external consultants, as well as the involvement of end-users in

different project phases [32] [54]. Employee skills are very important, since they ensure that the technical and organizational aspects of the project run efficiently [34]. Without the appropriate skills of real system users, the ERP implementation is difficult to be successful.

It is hypothesised that all those factors will have an impact on ERP implementation success:

H4: End-users have a positive impact on ERP implementation success.

E. Impact on organizational performance

ERP implementation success can be measured using two different approaches. Some researchers have proposed quantitative financial measures (e.g. ROI, market share), while some others have proposed qualitative ones (e.g. decision making improvement) [31]. Dezdar and Ainin [11] argued that ERP implementation success depends on the evaluation of its actual users.

Shang and Seddon [43] stated that successful completion of ERP projects has a positive impact on organizational performance. That was supported by Ifinedo and Nahar [30], who claimed that the IS success model of Delone and McLean [65] leads to improvement in organizational performance, through three key antecedents: system quality, information quality and service quality. Consequently, the present study examined whether these three dimensions have an impact on organizational performance.

Information quality refers to the accuracy, timeliness, completeness and consistency of the information provided by the ERP system [31]. If the product (information provided by the ERP) is not delivered on time (timeliness) and does not conform to the needs of its customers (ERP users), then the latter will be dissatisfied and the company will lose business [66]. On the other hand, increased information quality will have a positive organizational impact, in terms of customer satisfaction and, thus, overall organizational performance will increase.

H5a: Information quality has a positive impact on organizational performance.

A well-designed ERP system is necessary for gaining organizational benefits. According to Chien and Tsaur [31], system quality is measured in terms of ease-of-use, functionality, reliability, flexibility and data quality. The expected benefits of system quality include cost reduction, enhanced performance and improved efficiency [67]. On the other hand, a system that is neither well designed nor user-friendly will, probably, create the risk of system failure [68].

H5b: System quality has a positive impact on organizational performance.

Service quality is measured via the reliability, assurance and responsiveness of ERP service providers [31]. This dimension includes the overall quality of services that a particular Information System (IS) provides to an organisation [69]. According to Gorla and Wong [70], service quality is positively associated with organizational impact.

H5c: Service quality has a positive impact on organizational performance.

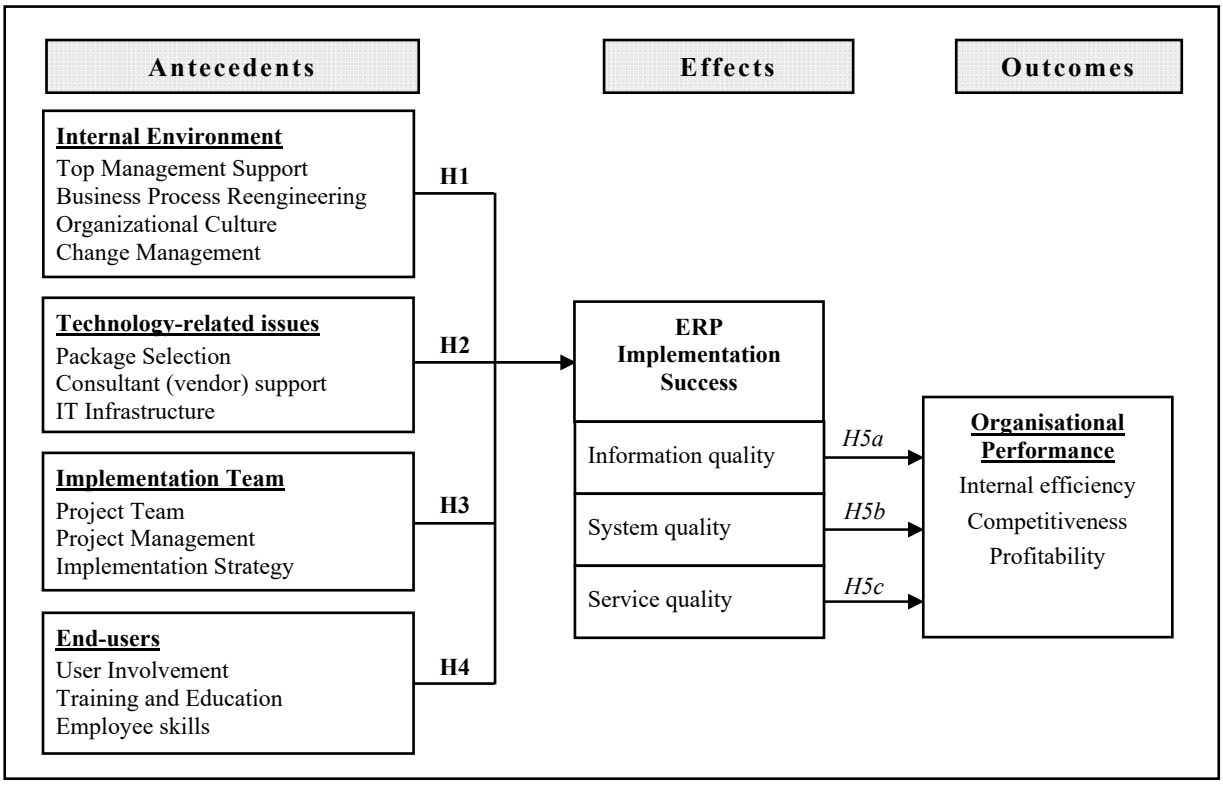


Fig 1. The Proposed Conceptual Framework of the study.

IV. Research methodology

A. Population of the study

The proposed conceptual framework was tested on a sample of Greek companies that have implemented an ERP system. Data concerning the target population were obtained via the web sites of various ERP system providers operating in Greece. Totally, 617 companies that have implemented an ERP system were identified.

B. Measurement

A structured questionnaire, designed specifically for this empirical study, was used in order to collect the appropriate empirical data. All items used to measure the various research factors have been adopted by previous studies [6] [7] [10] [26] [27] [30] [31] [32] [41] [42] [45] [50] [54]. The five-point Likert scale was used for the measurement of all research factors. In total, ninety (90) items were used for the measurement of all research factors.

C. Data collection

IT managers were selected as key respondents, due to their experience and expertise. Questionnaires were sent after telephonic contact with the IT manager. After making all necessary arrangements, 467 questionnaires were distributed to 467 companies that agreed to participate in the survey. The research period lasted three months (October to December 2015). 213 questionnaires were returned, but after conducting all necessary controls 204 were used for data analysis. The 204 returned questionnaires represent a very satisfactory response rate of 43,6%.

D. Validity and reliability

During the construct validity test, each factor (see Table I) was evaluated (a) for its unidimensionality and reliability, (b) for its goodness of fit to the proposed model. The examination of the unidimensionality of each factor was conducted using Explanatory Factor Analysis (EFA). Moreover, Cronbach Alpha was used for estimating the reliability of the same factor. The evaluation of the goodness of fit of all factors was conducted using Confirmatory Factor Analysis (CFA). All tests concluded that, after the extraction of relatively few items, the scales used for the measurement of the research factors are valid and reliable. Table I presents the main results.

V. Empirical results

The conceptual framework was tested using the Structural Equation Modeling (SEM) technique. The estimation of the structural model was conducted with the Maximum Likelihood Estimation method [71].

After experimenting with various different models, it was decided that “organisational performance” should not be

TABLE I. VALIDITY AND RELIABILITY

Factors	**Kaiser-Mayer-Olkin**	**% of Variance**	**Cronbach Alpha**	**Normed X^2**	**C.R.**	**V.E.**	**CFI**
	Explanatory Factor Analysis			**Confirmatory Factor Analysis**			
Top Management Support	0,834	61,769	0,875	2,24	0,69	57,7%	0,97
Business Process Reengineering	0,850	61,335	0,834	2,69	0,71	73,6%	0,93
Organizational Culture	0,741	56,448	0,737	3,29	0,86	78,2%	0,89
Change Management	0,818	75,910	0,891	3,45	0,87	71,6%	0,91
Second-Oder EFA-CFA: Internal Environment	*0,692*	*55,674*	*0,734*	*2,78*	*0,76*	*58,6%*	*0,94*
Package Selection	0,658	76,237	0,687	2,11	0,77	66,8%	0,97
Consultant (vendor) support	0,764	60,086	0,822	3,53	0,81	59,4%	0,95
IT Infrastructure	0,758	68,850	0,756	3,47	0,76	74,3%	0,97
Second-Oder EFA-CFA: Technology-related issues	0,702	73,111	0,839	2,26	0,74	76,8%	0,91
Project Team	0,724	54,850	0,775	2,71	0,81	76,9%	0,91
Project Management	0,775	62,502	0,850	2,33	0,76	66,3%	0,97
Implementation Strategy	0,620	60,615	0,773	3,29	0,74	74,3%	0,97
Second-Oder EFA-CFA: Implementation Team	*0,707*	*71,544*	*0,801*	*3,22*	*0,81*	*83,7%*	*0,97*
User Involvement	0,773	67,888	0,841	2,55	0,82	81,8%	0,99
Training and Education	0,705	60,649	0,779	3,55	0,84	67,3%	0,91
Employee skills	0,711	73,114	0,731	2,36	0,76	78,4%	0,95
Second-Oder EFA-CFA: End-users	*0,637*	*60,780*	*0,691*	*3,45*	*0,71*	*69,8%*	*0,94*
Information Quality	0,762	62,175	0,796	2,14	0,83	79,3%	0,97
System Quality	0,643	67,678	0,757	2,93	0,89	78,3%	0,91
Service Quality	0,547	72,998	0,785	3,09	0,82	78,6%	0,93
Internal efficiency	0,773	79,219	0,865	2,53	0,73	79,1%	0,97
Competitiveness	0,694	59,733	0,769	2,83	0,71	66,6%	0,97
Profitability	0,732	66,159	0,736	3,66	0,84	78,3%	0,99

measured as coherent factor (structure), since the use of its various dimensions offered more in depth information about the investigated phenomenon. More specifically, as it can be seen in Fig. 2, the model that was finally examined includes three dimensions, with a total of ten factors.

In order to evaluate the fit of the overall model the chi-square value (X^2 = 194,61) and the p-value (p = 0,0647) were estimated. These values indicate a good fit of the data to the overall model. However, the sensitivity of the X^2 statistic to the sample size suggests the adoption of other measures for evaluating the overall model, such as the "Normed-X^2" index (2,95), the RSMEA index (0,057) the CFI (0,973) and the GFI (0,967), that all indicate a good fit.

Fig. 2 demonstrates the overall model, along with the path coefficients. In general, the results reveal the mechanism through which the antecedents of ERP implementation are affecting the various dimensions of organizational performance. Both direct and indirect effects are being examined, thus, enhancing the understanding of the investigated phenomenon. The main findings of the study are summarised below:

- Overall, the empirical results confirmed that the research model has satisfactory predictive power, since it can significantly explain the variance of the main dependent factors of the present study ("internal efficiency" by 31%, "internal competitiveness" by 51% and "profitability" by 35%). Moreover, the variance of "system quality" is explained by 78%, "service quality" is being explained by 32%, and "information quality" by 21%.
- While the relationship between several factors was not supported by the empirical data, partial support has been found for most of the hypotheses of the proposed model.
- More specifically, only one of the four antecedents included in the proposed conceptual framework (namely, "end-users") has an impact on all three dimensions capturing ERP implementation success (full support for Hypothesis 1). On the contrary, all other three antecedents have a direct impact on one dimension of ERP implementation success (partial support of Hypotheses 2, 3 and 4). More specifically, "internal environment" and "technology-related issues" have an impact on "system quality", while "implementation team" has an impact on "information quality".
- These findings underline the significance of the human factor in ERP implementation success. It seems that the end-users of the ERP system are the cornerstones for successful implementation. Implementing organizations need to take under serious consideration the involvement of end-user in the whole implementation process, provide training and education, while focusing on increasing their overall IT skills. Bradford and Florin [45], Dezdar and Ainin [22] found similar results.
- Moreover, according to the empirical results, the antecedents of ERP implementation success should be considered as a coherent bundle of activities. Implementing companies should focus on all of these dimensions, since their simultaneous enhancement has a commutative impact on the effectiveness of the implementation process. Nevertheless, focusing on end-users should become the first priority.

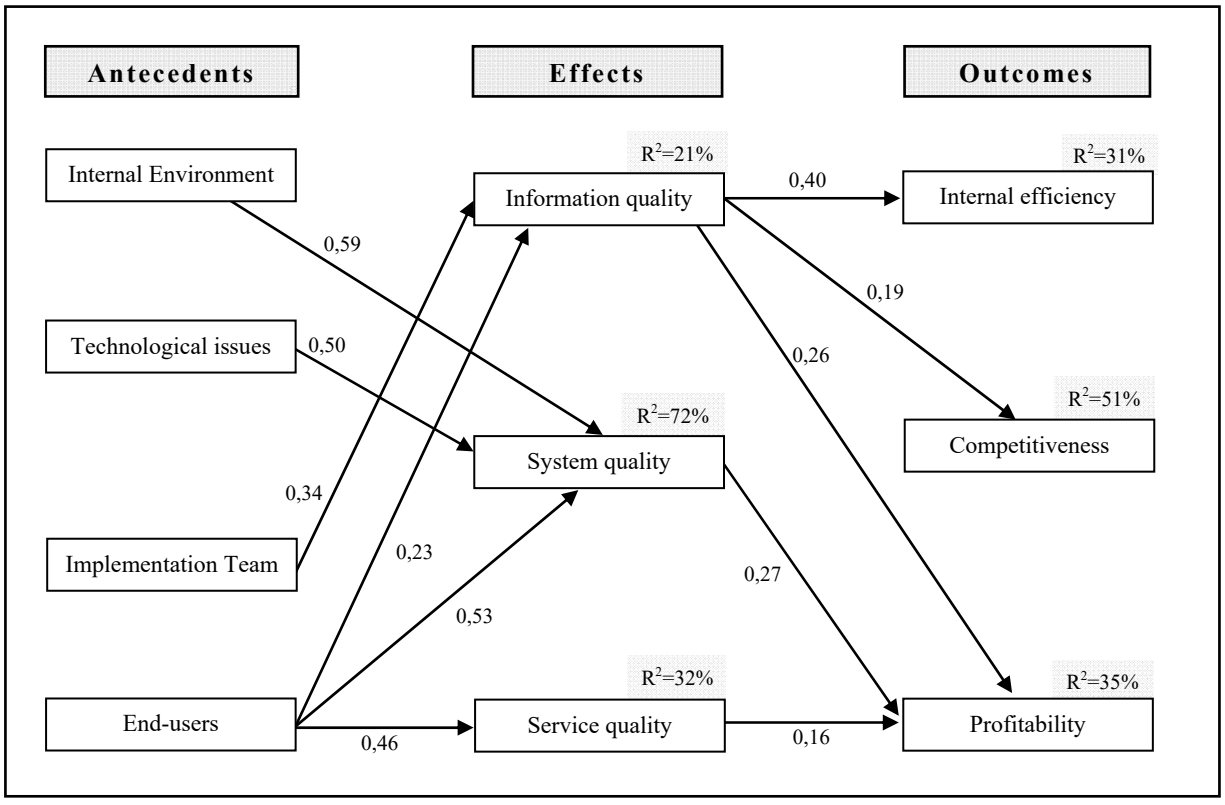


Fig. 2. Main empirical results (all paths are statistically significant).

- The empirical results offer full support for hypothesis 5a, arguing that "information quality" has an impact on all dimensions measuring "organizational performance". Partial support is being offered for hypotheses 5b and 5c, since "system quality" and "service quality" have an impact on "profitability". Despite the little empirical research that has been conducted on the relationship between ERP and organizational performance, the very few existing studies support these results [16] [17] [18] [20].
- The significance of the three-dimensional approach (antecedents, effects and outcomes) introduced in the present study lies in the interrelationships between various research factors. For example, the proposed model explains 51% of the variance in "competitiveness" (outcome). As it can be seen on Fig. 2, there is only one factor having a direct positive effect on "competitiveness" ("information quality", r = 0,19). Despite that, two of the antecedents included in the model ("implementation team", and "end users") have an indirect effect on "competitiveness", through "information quality". Only through that mechanism, the 51% of explained variance can be justified.
- Therefore, based on the above, it can be concluded that enhanced competitiveness is a direct result of information quality, but, despite that, end-users of the ERP system and the team which is responsible for the implementation both have a significant strengthening effect on that relationship. Mapping down these complex relationships has been one of the most important contributions of the present study. Managers are urged to bear in mind the existence of these causal effects, since their main objective should be the enhancement of organizational performance.

VI. Conclusions

The present study developed an conceptual framework that has never been used in the international ERP literature. Future studies can adopt the same approach, further highlighting the relationship between critical factors for ERP implementation success, ERP implementation, and organisational performance. Its main contribution lies on its implemented methodology and conceptualisation.

The empirical results enhance the understanding of the DeLone and McLean Information System (IS) success model [65] [69]. More specifically, a direct effect between "information quality", "system quality" and "service quality" and various measures of organisation performance has been established. On the contrary, the DeLone and McLean model [65] [69] argue that these dimensions have an indirect effect on organisational performance (through "user satisfaction" and "intention to use"). The support of a direct relationship, found in the present study, can be attributed to its conceptualization: the simultaneous examination of antecedents, effects and outcomes increases the explanatory power of the proposed model, offering a more complete picture of ERP implementation.

First and foremost, the empirical results highlight the significance of "end-users" in implementation effectiveness. Among all the antecedents of ERP implementation success, the dimension capturing the contribution of "end-users" on the whole process has the most significant effect. Previous studies [22] [41] have provided similar results, but very few have adopted such a methodological approach, using multiple factors for the measurement of each antecedent of ERP implementation success. For example, Zhang, Lee, Huang, Zhang and Huang [41] have, also, developed an ERP implementation success framework, but their empirical analysis was qualitative (case study research).

Secondly, since all four antecedents influence, each in a different degree, the three dimensions of implementation success, it is concluded that companies should focus on their collective enhancement. After all, all of these four factors (antecedents) have an indirect effect on organisational performance. For instance, the factor measuring the characteristics of the "implementation team" has an indirect impact on "internal efficiency", "competitiveness" and "profitability", through "information quality". Such a conceptual framework, taking under consideration both direct and indirect effects between antecedents and final outcomes, has very seldom been introduced in the ERP literature.

Thirdly, "information quality" appears as the most significant aspect of the implementation process, since it has a direct effect on all performance measures. According to previous studies [14], enhanced information quality during ERP implementation leads to increased performance. Information quality can help organizations choose different supply resources, hence, produce with lower costs and, therefore, develop competitive advantages, while increasing their competitiveness and internal efficiency.

The proposed model has adequate explanatory power, since it explains a significant percentage of the variance of all three main dependent factors. More specifically, it can predict 51% of the variance in business competitiveness, underlining the effect of ERP implementation on measures that previous studies have neglected [22] [43]. The enhanced power of the model lies in its three-dimensional approach, investigating both direct and indirect effects.

On a practical level, the present study offers a comprehensive list of factors having an impact on the successful implementation of an ERP system. Managers should focus on the enhancement of the most significant of these factors, identifying specific objectives for achieving successful implementation and increased performance. Finally, the present study underlines the mediating role of ERP implementation success in the relationship between various antecedents and final outcomes. It seems that such an approach can better describe the hypothesised relationships, since the effects of the four antecedents need to be translated into tangible benefits (information, system, and service quality) in order to have a positive impact on performance.

The study is somehow limited by the poor definition of its population. This limitation is inherent to all studies of the field, since a complete list of ERP implementing companies is not easy to acquire. Further research is suggested with larger samples that would, probably, offer more information and strengthen the results of the present study. Moreover, it would be interesting to examine more factors and gather primary data from all company personnel, so as to achieve a more complete view of the subject under investigation.

REFERENCES

[1] I. Pawełoszek, "Approach to analysis and assessment of ERP system. A software vendor's perspective", IEEE Federated Conference on Computer Science and Information Systems (FedCSIS), pp. 1415-1426, 2015.

[2] M. Relich, "Knowledge acquisition for new product development with the use of an ERP database", IEEE Federated Conference on Computer Science and Information Systems (FedCSIS), pp. 1285-1290, 2013.

[3] J. Keller, and A. Heiko, "The influence of information and communication technology (ICT) on future foresight processes: Results from a Delphi survey", *Technological Forecasting and Social Change*, Vol. 85, pp. 81-92, 2014.

[4] A. Tenhiälä, and P. Helkiö, "Performance effects of using an ERP system for manufacturing planning and control under dynamic market requirements", *Journal of Operations Management*, Vol. 36, pp. 147-164, 2015, http://dx.doi.org/10.1016/j.jom.2014.05.001.

[5] E. Galy, and M. J. Sauceda, "Post-implementation practices of ERP systems and their relationship to financial performance", *Information & Management*, Vol. 51, No. 3, pp. 310-319, 2014, http://dx.doi.org/10.1016/j.im.2014.02.002.

[6] E. M. Shehab, M. W. Sharp, L. Supramaniam, and T. A. Spedding, "Enterprise Resource Planning: an integrative review", *Business Process Management Journal*, Vol. 10, No. 4, pp. 359-386, 2004, http://dx.doi.org/10.1108/14637150410548056.

[7] I. C. Ehie, and M. Madsen, "Identifying critical issues in enterprise resource planning (ERP) implementation", *Computers in Industry*, Vol. 56, No. 6, pp. 545-557, 2005, http://dx.doi.org/10.1016/j.compind.2005.02.006.

[8] M. M. Ahmad, and R. P. Cuenca, "Critical success factors for ERP implementation in SMEs", *Robotics and Computer-Integrated Manufacturing*, Vol. 29, No. 3, pp. 104-111, 2013, http://dx.doi.org/10.1016/j.rcim.2012.04.019.

[9] R. Rajnoha, J. Kádárová, A. Sujová, and G. Kádár, "Business Information Systems: Research Study and Methodological Proposals for ERP Implementation Process Improvement", *Procedia-Social and Behavioral Sciences*, Vol. 109, pp. 165-170, 2014, http://dx.doi.org/10.1016/j.sbspro.2013.12.438.

[10] E. J. Umble, R. R. Haft, and M. M. Umble, "Enterprise resource planning: Implementation procedures and critical success factors", *European Journal of Operational Research*, Vol. 146, No. 2, pp. 241-257, 2003, http://dx.doi.org/ 10.1016/S0377-2217(02)00547-7.

[11] S. Dezdar, and S. Ainin, "Examining ERP implementation success from a project environment perspective", *Business Process Management Journal*, Vol. 17, No. 6, pp. 919-939, 2011, http://dx.doi.org/10.1108/14637151111182693.

[12] A. Amid, M. Moalagh, and A. Z. Ravasan, "Identification and classification of ERP critical failure factors in Iranian industries", *Information Systems*, Vol. 37, No. 3, pp. 227-237, 2012, http://dx.doi.org/10.1016/j.is.2011.10.010.

[13] J. Malaurent, and D. Avison, "From an apparent failure to a success story: ERP in China-Post implementation", *International Journal of Information Management*, Vol. 35, No. 5, pp. 643-646, 2015, http://dx.doi.org/10.1016/j.ijinfomgt.2015.06.004.

[14] M. Al-Mashari, A. Al-Mudimigh, and M. Zairi, "Enterprise resource planning: taxonomy of critical factors", *European Journal of Operational Research*, Vol. 146, No. 2, pp. 352-364, 2003, http://dx.doi.org/10.1016/S0377-2217(02)00554-4.

[15] X. Chan, Y. Y. Lau, and J. M. J. Ng, "Critical evaluation of ERP implementation on firm performance: a case study of AT&T". *International Journal of Logistics Systems and Management*, Vol. 12, No. 1, pp. 52-69, 2012, http://dx.doi.org/ 10.1504/IJLSM.2012.047058.

[16] V. F. Dumitru, N. Albu, C. N. Albu, and M. Dumitru, "ERP implementation and organizational performance. A Romanian case study of best practices", *Amfiteatru Economic*, Vol. 15, No. 34, pp. 518-531, 2013.

[17] H. Ince, S. Z. Imamoglu, H. Keskin, A. Akgun, and M. N. Efe, "The Impact of ERP Systems and Supply Chain Management Practices on Firm Performance: Case of Turkish Companies", *Procedia-Social and Behavioral Sciences*, Vol. 99, pp. 1124-1133, 2013, http://dx.doi.org/10.1016/j.sbspro.2013.10.586.

[18] P. L. Liu, "Empirical study on influence of critical success factors on ERP knowledge management on management performance in high-tech industries in Taiwan", *Expert Systems with Applications*, Vol. 38, No. 8, pp. 696-704, 2011, http://dx.doi.org/10.1016/j.eswa.2011.02.045.

[19] A. Madapusi, and D. D'Souza, "The influence of ERP system implementation on the operational performance of an organization", *International Journal of Information Management*, Vol. 32, No. 1, pp. 24-34, 2012, http://dx.doi.org/10.1016/j.ijinfomgt.2011.06.004.

[20] A. I. Nicolaou, and L. H. Bajor, "ERP systems implementation and firm performance", *Review of Business Information Systems*, Vol. 8, No. 1, pp. 53-60, 2011, http://dx.doi.org/10.19030/rbis.v8i1.4504.

[21] W. H. Tsai, K. C. Lee, J. Y. Liu, S. J. Lin, and Y. W. Chou, "The influence of ERP systems' performance on earnings management", *Enterprise Information Systems*, Vol. 6, No. 4, pp. 491-517, 2012, http://dx.doi.org/10.1080/17517575.2011.622414.

[22] S. Dezdar, and S. Ainin, "The influence of organizational factors on successful ERP implementation", *Management Decision*, Vol. 49, No. 6, pp. 911-926, 2011, http://dx.doi.org/10.1108/00251741111143603.

[23] W. Hwang, and H. Min, "Assessing the impact of ERP on supplier performance", *Industrial Management & Data Systems*, Vol. 113, No. 7, pp. 1025-1047, 2013, http://dx.doi.org/10.1108/IMDS-01-2013-0035.

[24] H. W. Chou, Y. H. Lin, H. S. Lu, H. H. Chang, and S. B. Chou, "Knowledge sharing and ERP system usage in post-implementation stage", *Computers in Human Behavior*, Vol. 33, pp. 16-22, 2014.

[25] O. Zach, B. E. Munkvold, and D. H. Olsen, "ERP system implementation in SMEs: exploring the influences of the SME context", *Enterprise Information Systems*, Vol. 8, No. 2, pp. 309-335, 2014, http://dx.doi.org/10.1080/17517575.2012.702358.

[26] K. K. Hong, and Y. G. Kim, "The critical success factors for ERP implementation: an organizational fit perspective", *Information & Management*, Vol. 40, No. 1, pp. 25-40, 2002, http://dx.doi.org/10.1016/S0378-7206(01)00134-3.

[27] J. Motwani, R. Subramanian, and P. Gopalakrishna, "Critical factors for successful ERP implementation: exploratory findings from four case studies", *Computers in Industry*, Vol. 56, No. 6, pp. 529-544, 2005, http://dx.doi.org/10.1016/j.compind.2005.02.005.

[28] T. M. Somers, and K. Nelson, "The impact of critical success factors across the stages of enterprise resource planning implementations". In: Proceedings of the 34th Annual Hawaii International Conference on System Sciences, 2001, http://dx.doi.org/10.1109/HICSS.2001.927129.

[29] J. Ram, M. L. Wu, and R. Tagg, "Competitive advantage from ERP projects: examining the role of key implementation drivers", *IEEE Engineering Management Review*, Vol. 42, No. 3, pp. 36-53, 2014, http://dx.doi.org/10.1016/j.ijproman.2013.08.004.

[30] P. Ifinedo, and N. Nahar, "Quality, Impact and Success of ERP systems: a Study Involving Some Firms in the Nordic-Baltic Region", *Journal of Information Technology Impact*, Vol. 6, No. 1, pp. 19-46, 2006.

[31] S. W. Chien, and S. M. Tsaur, "Investigating the success of ERP systems: case studies in three Taiwanese high-tech industries", *Computers in Industry*, Vol. 58, No. 8-9, pp. 783-793, 2007, http://dx.doi.org/10.1016/j.compind.2007.02.001.

[32] F. F. Nah, J. L. Lau, and J. Kuang, "Critical factors for successful implementation of enterprise systems", *Business Process Management Journal*, Vol. 7, No. 3, pp. 285-296, 2001, http://dx.doi.org/10.1108/14637150110392782.

[33] J. Esteves, and J. A. Pastor, "Organizational and technological critical success factors behavior along the ERP implementation phases". In: Enterprise information systems VI. Ed. by I. Seruca, J. Cordeiro, S. Hammoudi, and J. Filipe, Springer Netherlands, pp. 63-71, 2006, http://dx.doi.org/10.1007/1-4020-3675-2_8.

[34] P. Holland, B. Light, and N. Gibson, "A critical success factors model for enterprise resource planning implementation". In: Proceedings of the 7th European Conference on Information Systems, pp. 273-297, 1999.

[35] T. Loh, and S. C. L. Koh, "Critical elements for a successful enterprise resource planning implementation in small-and medium-sized enterprises", *International Journal of Production Research*, Vol. 42, No. 17, pp. 3433-3455, 2004, http://dx.doi.org/10.1080/00207540410001671679.

[36] E. Hustad, and D. Olsen, "Critical Issues Across the ERP Life Cycle in Small-and-Medium- Sized Enterprises: Experiences from a Multiple Case Study", *Procedia Technology*, Vol. 9, pp. 179-188, 2013, http://dx.doi.org/10.1016/j.protcy.2013.12.020.

[37] A. A. Hawari, and R. Heeks, "Explaining ERP failure in a developing country: a Jordanian case study", *Journal of Enterprise Information Management*, Vol. 23, No. 2, pp. 135-160, 2010.

[38] E. T. G. Wang, C. C. L. Lin, J. J. Jiang, and G. Klein, "Improving enterprise resource planning (ERP) fit to organizational process through knowledge transfer", *International Journal of Information Management*, Vol. 27, No. 3, pp. 200-212, 2007, http://dx.doi.org/10.1016/j.ijinfomgt.2007.02.002.

[39] E. Alsene, "ERP systems and the coordination of the enterprise", *Business Process Management Journal*, Vol. 13, No. 3, pp. 417-432, 2007, http://dx.doi.org/10.1108/14637150710752326.

[40] E. W. T. Ngai, C. C. H. Law, and F. K. T. Wat, "Examining the critical success factors in the adoption of enterprise resource planning", *Computers in Industry*, Vol. 59, No. 6, pp. 548-564, 2008, http://dx.doi.org/10.1016/j.compind.2007.12.001.

[41] Z. Zhang, M. K. O. Lee, P. Huang, L. Zhang, and X. Huang, "A framework of ERP systems implementation success in China: An empirical study", *International Journal of Production Economics*, Vol. 98, No. 1, pp. 56-80, 2005, http://dx.doi.org/10.1016/j.ijpe.2004.09.004.

[42] V. Gargeya, and C. Brady, "Success and failure factors of adopting SAP in ERP system implementation", *Business Process Management Journal*, Vol. 11, No. 5, pp. 501-516, 2005.

[43] S. Shang, and P. B. Seddon, "Managing process deficiencies with enterprise systems", *Business Process Management Journal*, Vol. 13, No. 3, pp. 405-416, 2007, http://dx.doi.org/10.1108/14637150710752317.

[44] S. C. Gardiner, J. B. Hanna, and M. S. LaTour, "ERP and the reengineering of industrial marketing processes: A prescriptive overview for the new-age marketing manager", *Industrial Marketing Management*, Vol. 31, No. 4, pp. 357-365, 2002, http://dx.doi.org/10.1016/S0019-8501(01)00167-5.

[45] M. Bradford, and J. Florin, "Examining the role of innovation diffusion factors on the implementation success of enterprise resource planning systems", *International Journal of Accounting Information Systems*, Vol. 4, No. 3, pp. 205-225, 2003, http://dx.doi.org/10.1016/S1467-0895(03)00026-5.

[46] T. Cadden, D. Marshall, and G. Cao, "Opposites attract: organisational culture and supply chain performance", *Supply Chain Management: an International Journal*, Vol. 18, No. 1, pp. 86-103, 2013, http://dx.doi.org/10.1108/13598541311293203.

[47] W. Ke, and K. K. Wei, "Organizational culture and leadership in ERP implementation", *Decision Support Systems*, Vol. 45, No. 2, pp. 208-218, 2008, http://dx.doi.org/10.1016/j.dss.2007.02.002.

[48] R. A. Jones, N. L. Jimmieson, and A. Griffiths, "The impact of organizational culture and reshaping capabilities on change implementation success: the mediating role of readiness for change", *Journal of Management Studies*, Vol. 42, No. 2, pp. 361-386, 2005, http://dx.doi.org/10.1111/j.1467-6486.2005.00500.x.

[49] A. Al-Ghamdi, "Change management Strategies and Processes for the successful ERP System Implementation: a Proposed Model", *International Journal of Computer Science and Information Security*, Vol. 11, No. 2, pp. 36-41, 2013.

[50] T. M. Somers, and K. G. Nelson, "A taxonomy of players and activities across the ERP project life cycle", *Information & Management*, Vol. 41, No. 3, pp. 257-278, 2004, http://dx.doi.org/10.1016/S0378-7206(03)00023-5.

[51] F. Cua, and S. Reames, "Big Vendor vs. Little Vendor: Managing the Enterprise Resource Planning (ERP) Project to Overcome the Laggard Sales Barrier", *International Journal of Information Technology Project Management*, Vol. 4, No. 2, pp. 50-74, 2013, http://dx.doi.org/10.4018/jitpm.2013040104.

[52] P. Katerattanakul, J. Lee, and S. Hong, "Effect of business characteristics and ERP implementation on business outcomes: An exploratory study of Korean manufacturing firms", *Management Research Review*, Vol. 37, No. 2, pp. 186-206, 2014.

[53] A. Sarfaraz, K. Jenab, and A. C. D'Souza, "Evaluating ERP implementation choices on the basis of customisation using fuzzy AHP", *International Journal of Production Research*, Vol. 50, No. 23, pp. 7057-7067, 2012, http://dx.doi.org/10.1080/00207543.2012.654409.

[54] R. Bhatti, "Critical Success Factors for the Implementation of Enterprise Resource Planning (ERP): Empirical Validation". In: *The second International Conference on Innovation in Information Technology*, 2005, Accessed at 17-11-2015 from: https://goo.gl/8Nual4.

[55] S. Newell, C. Tansley, and J. Huang, "Social capital and knowledge integration in an ERP project team: the importance of bridging and bonding", *British Journal of Management*, Vol. 15, No. 1, pp. 43-57, 2004, http://dx.doi.org/10.1111/j.1467-8551.2004.00405.x.

[56] C. C. Wei, and M. J. J. Wang, "A comprehensive framework for selecting an ERP system", *International Journal of Project Management*, Vol. 22, No. 2, pp. 161-169, 2004, http://dx.doi.org/10.1016/S0263-7863(02)00064-9.

[57] H. Xu, P. J. Rondeau, and S. Mahenthiran, "Teaching case the challenge of implementing an ERP system in a small and medium enterprise: a teaching case of ERP project management", *Journal of Information Systems Education*, Vol. 22, No. 4, pp. 291-296, 2011.

[58] T. K. Chien, and M. S. Cheng, "The implementation strategy of key task for ERP activities". In: *Proceedings of the 2014 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, Malaysia, pp. 1126-1130, 2014.

[59] E. F. Berger, "A generic model for selecting an ERP implementation strategy". In: *Novel Methods and Technologies for Enterprise Information Systems*, Ed. by F. Piazolo and M. Felderer. Springer International Publishing. Vienna: Austria, pp. 239-247, 2014.

[60] M. Chang, W. Cheung, C. Cheng, and J. Yeung, "Understanding ERP system adoption from the user's perspective", *International Journal of Production Economics*, Vol. 113, No. 2, pp. 928-942, 2008, http://dx.doi.org/10.1016/j.ijpe.2007.08.011.

[61] M. A. Krumbholz, J. Galliers, N. Coulianos, and N. Maiden, "Implementing enterprise resource planning packages in different corporate and national cultures", *Journal of Information Technology*, Vol. 15, No. 4, pp. 267-279, 2000, http://dx.doi.org/10.1080/02683960010008962.

[62] S. Dezdar, and S. Ainin, "Analysis of success measures in enterprise resource planning implementation projects", *International Journal of Business Performance Management*, Vol. 12, No. 4, pp. 334-353, 2011.

[63] L. Zhang, M. K. Lee, Z. Zhang, and P. Banerjee, "Critical success factors of enterprise resource planning systems implementation success in China", In: *Proceedings of the 36th Annual Hawaii International Conference on System Sciences*, 2003, http://dx.doi.org/10.1109/HICSS.2003.1174613.

[64] S. Dezdar, "Strategic and tactical factors for successful ERP projects: insights from an Asian country", *Management Research Review*, Vol. 35, No. 11, pp. 1070-1087, 2012.

[65] W. H. DeLone, and E. R. McLean, "Information Systems Success: the Quest for the Dependent Variable", *Information Systems Research*, Vol. 3, No. 1, pp. 60-95, 1992, http://dx.doi.org/10.1287/isre.3.1.60.

[66] P. M. Clikeman, "Improving Information Quality", *Internal Auditor*, Vol. 56, No. 3, pp. 32-33, 1999.

[67] M. Nofal, and Z. Yusof, "Integration of Business Intelligence and Enterprise Resource Planning within Organizations", *Procedia Technology*, Vol. 11, pp. 658-665, 2013.

[68] E. T. Wang, and J. H. Chen, "Effects of internal support and consultant quality on the consulting process and ERP system quality", *Decision Support Systems*, Vol. 42, No. 2, pp. 1029-1041, 2006.

[69] W. H. Delone, and E. R. McLean, "The Delone and McLean model of information systems success: a ten-year update", *Journal of Management Information Systems*, Vol. 19, No. 4, pp. 9-30, 2003, http://dx.doi.org/10.1080/07421222.2003.11045748.

[70] N. Gorla, T. M. Somers, and B. Wong, "Organizational impact of system quality, information quality, and service quality", *The Journal of Strategic Information Systems*, Vol. 19, No. 3, pp. 207-228, 2010, http://dx.doi.org/10.1016/j.jsis.2010.05.001.

[71] N. K. Bowen, and S. Guo, *Structural equation modeling*, UK: Oxford University Press, 2011.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1263–1271
DOI: 10.15439/2016F181
ACSIS, Vol. 8. ISSN 2300-5963

Attempt to Extend Knowledge of Decision Support Systems for Small and Medium-Sized Enterprises

Jerzy Korczak, Helena Dudycz, Bartłomiej Nita, Piotr Oleksyk, Adrian Kaźmierczak
Wrocław University of Economics Komandorska Str. 118/120,
PL 53-345 Wroclaw, Poland
Email: {jerzy.korczak, helena.dudycz, bartlomiej.nita, piotr.oleksyk, adrian.kazmierczak}@ue.wroc.pl

Abstract—**The article presents a proposal to extend the functionality and knowledge of Business Intelligence systems to answer the requirements of managers of small and medium-sized enterprises (SMEs). It concerns two major aspects of the system, i.e. the interface that takes into account the level of knowledge of the manager, and supports the interpretation of economic and financial information using the built-in domain ontologies. The project is related to the design of smart decision support systems based on financial ontology and on the model of manager knowledge created by eye tracking analysis. An experiment was carried out on real financial data extracted from the database of Business Intelligent BINOCLE, developed by Bilander Co. To create a model of manager knowledge, a number of financial analysts, experts and economists were invited to analyze the pre-defined financial reports. Their tasks were observed and analyzed by the eye-tracking system StudioTM, Tobii. The logs of the system as well as the financial ontology have been used to develop the intelligent interface of a Decision Support System.**

I. Introduction

Decision-making in small and medium-sized enterprises is an extremely difficult process. The most important problems of the functioning of such enterprises are operational continuity, ensuring the ongoing customer service and support as well as technological aspects. Managing a business requires access to the appropriate information system that must always go hand in hand with the methods of financial analysis, allowing managers to monitor the changes in the environment, identify the different types of risk, choose appropriate forms of insurance against these risks, and follow appropriate scenarios of development. Each of the scenarios predicts future financial situations using the methods and tools of financial analysis. Managers of small and medium-sized enterprises (SMEs), on the one hand, are usually confronted with the barriers and limits on information, and on the other hand are able more easily to control costs and expenses using an individual, simplified information processing system, regardless of the requirements of accounting standards. It may be difficult for them to access prospective information allowing the evaluation of the anticipated environment changes.

Managers of SMEs need solutions that support their decisions based on transactional data which, after being processed by the tools of financial analysis, allow them to prepare the draft of decision. Today's information technology assures managers access to the multi-dimensional data stored in various data-bases and enables them to perform multi-criteria analysis. The problem appears, among others, to lie in the excessive numbers of reports which are generated by transactional and executive information systems. In the process of management, the information overload significantly reduces the ability to make the right decision.

In SMEs, financial forecasts needed to make good business decisions oriented towards improving financial efficiency and dynamic development are very often prepared in a cursory manner or simply ignored. This is mainly due to lack of time, and not to sufficient and limited managerial knowledge.

The aim of the article is to point out the possibilities of building the model of financial knowledge for managers in Business Intelligence system using eye tracking tools[1], and applying the model to interpret reported economic data. Data from eye-tracking illustrates not only the perception of the economic reports by a manager, but also enables the manager to identify schemes analysis and determine the level of his or her knowledge. It facilitates also the establishment of a manager knowledge profile and, consequently, it creates a possibility of adapting the interface to an individual's skills. In the project, the whole process of analysing financial data will be supported by built-in ontologies of economic and financial knowledge which is essential for SME managers. This will be possible thanks to research focusing on the development of an intelligent interface supporting interpretation of economic information, which also is associated with modeling of managerial knowledge. The described approach is a continuation of the construction of the intelligent cockpit for managers (InKoM project), whose main objective was to facilitate financial analysis and the

[1] Eye tracking is a collection of techniques for the measurement, acquisition and analysis of data about the position and movements of the eyeballs. It provides a quantitative measurement without referring to subjective, verbal relationships of the respondent [http://eyetracking.pl].

evaluation of the economic status of the company on a competitive market [1].

The structure of the article is as follows. The next section focuses on the issues related to management-based analysis, particularly in relation to SMEs. The next section briefly refers to the Business Intelligence (BI) system in the context of the Key Performance Indicators (KPI) used to build an ontology of economic and financial knowledge. Section 4 is devoted to the modeling of managerial knowledge, together with a synthetic description of the eye-tracking results. The article concludes with a summary of work to date.

II. Supporting of managerial analysis

Management of the SME is a very complex and difficult task. In such organizations, resources for wider financial and accounting services are usually limited. Often, these tasks are assigned to accountants who are excessively burdened with operational activities and supervision of the correct tax settlements. Managers of SMEs often have to focus on the technological problems related to the core business, which limits the possibility of financial projections associated with the preparation of the right decisions to ensure financial security and dynamic development based on improving financial efficiency.

Financial analysis allows the interpretation of the information necessary for the ongoing management of an enterprise[2]. Managers of SMEs are expected to find support in the following areas:

- assessment of the project's effectiveness - the answer to the question whether the achieved profits are adequate to the funds involved, and how to compare the results with other companies with similar operations;
- assessment of cash flows - whether it is possible to repeat a generated surplus in the following periods or whether it will no longer occur in the future.

The main activities of the SMEs are focused on business operations. Information processing and related decisions concern the core activities of the company, and are associated with its specificity and membership in a particular industry. An important element of operational management in SMEs is to control liquidity, while it should be emphasized that classical liquidity ratios used in large companies cannot be always directly applied to SMEs. In this view it is worth designing solutions that make use of analytical accounting records to project liquidity on a monthly basis. There is no doubt that this kind of analysis of liquidity, taking into account the size of the entity, the specificity of the industry, and the type of offered products or services is not possible without expert knowledge. On the basis of empirical studies we found that information requirements for short term management concern [2] information about liquidity, information on revenues generated by the company, and information on costs incurred. But for the long-term management area, information is needed: about the company's indebtedness, about the profitability of planned investments, and about the financial situation of the branch.

The process of decisions support made needs - firstly - the decomposition and the proper selection of management instruments. Managing medium-sized enterprises requires information about the possibility of gridlock preventing stable business in the near future. Secondly, there is a need to acquire information to generate the appropriate level of profit and a minimum level of margins in order to make the right choices. They are focused on obtaining information on opportunities to increase revenue or to carry out activities aimed at cost minimization.

Regarding long-term decisions, managers of small companies are not able to relate to the use of advanced solutions for e.g. capital budgeting. They need analytical models that enable them to identify signals indicative of the need for investment decisions and the level of resources involved. With regard to medium-sized enterprises, this type of analysis should be augmented by simulations on the effectiveness of individual organizational units (responsibility centers) and the profitability of equity involved.

Acquisition of information to support such decisions is possible mainly through the use of patterns developed by experts in financial management. Unfortunately, it is impossible in this case to use the universal solutions - in this case it is necessary to dispose over an intelligent system that using expert knowledge provides ready-made decisions, taking into account the specific nature of the company[3] .

Also important are analysis of the index of debt level and the use of static and dynamic methods of estimating the profitability of investment. For organizations with greater economic importance, the study should be completed by:

- the use of the Balanced Score Card and the early warning system,
- analysis of indicators of debt service, study of the cost of capital: domestic and foreign capital, the weighted average cost of capital,
- pro-forma financial statements - as information about the financial effects of planned long-term actions,
- evaluation of the company's profitability: return on sale ratios, return on assets and equity - to various cuts in the financial results,

[2] It should be emphasized that the selection of appropriate methods of analysis of the financial requirements of SME managers is necessary to determine the company's ability to continue its operations, and to define financial needs, budgets and capital resources of financing assets, signs of danger and risk of changes in the competitive position and trends in various business areas.

[3] The management of long-term financial analysis conducted by managers of small businesses should include: a preliminary analysis of financial statements, ratio analysis, information on liquidity ratios, evaluation of static measures of liquidity, turnover ratios of inventories, receivables and payables analysis, analysis of the operating cycle in terms of the efficiency of indicators of capital engagement, productivity of assets and equity, and information on current and future income and expenses.

- methods of budgeting process.

Generally, the support of decision-making is based on the generation of ready-made paths, together with projections of the effects of planned decisions. For example, in a system of the Binocle Bilander company which is classified as a category of Business Intelligence systems, there are many useful and powerful functionalities that make possible multivariate analysis of financial data. However, due to the limited resources of SMEs, it is necessary to develop ready-made reporting and decision paths for managers and owners of SMEs. These reports and patterns of decision-making should take into account inter alia the support for operational and financial planning, and risk analysis (in particular, the risk of bankruptcy). In addition, they may include support in investment decisions and measuring the effectiveness of the company as a whole and its individual organizational units.

In addition, to improve the decision-making processes in SMEs, three important issues should be taken into account:

- number of KPIs (key measures of achievement),
- methods of forecasting and simulation to facilitate taking corrective actions,
- standard indicators for SMEs needed for benchmarking.

Also crucial is a selection of available reports (option in the Binocle system) to present the most important information for the manager. If needed, these reports may be further detailed.

III. KPI IN BUSINESS INTELLIGENCE SYSTEMS

Both in business practice and in literature one can find different definitions of Business Intelligence (BI). The main goal of each form of BI is to gain access in a timely manner to relevant data to allow making the best decision at a given time [3; 4]. An important element of BI systems is also the visualization of the calculated KPIs. These indicators are the basis of decision-making both at strategic as well as tactical and operational level. The usefulness of indicators depends on the manager's understanding of the concepts, measures and structural links involved. Leading financial analysis of the indicators the manager examines their semantic relationships. The essence of the evaluation is the appropriate computing and use of indicators derived from different financial statements (including balance sheet, profit and loss statement,…). The usability of economic analysis depends on, among other things, the manager's exact understanding of the existing structural semantic links, and relations between indicators and economic terms. According to the degree of data aggregation, the measures might be global or partial, where the indicators can be summed up, differentiated, multiplied or divided.

Ratio analysis used to assess the activities of the company has both advantages and disadvantages. Among the first are: the ease of measurement, the availability of source data, the ability to identify critical areas of economic activities, the universality of the indicators, allowing, among other things, the conduct of comparative analyses with other companies. While the disadvantages are: no indication of the reasons for unfavorable events, the risk of misinterpretation of measures, the lack of universal standard ratios. Finding the causes of adverse events, and taking note of the positive factors, can facilitate the evaluation of the data by analyzing the semantic relationships between economic indicators.

Most BI systems (referred to as traditional BI systems or BI 1.0) are primarily intended for managers who are familiar with data models and are able to build all kinds of scenarios analysis [5]. The literature points to the creation of the next generation BI systems, known as Business Intelligence 2.0 [6; 7]. These systems are characterized by properties such as event management and analysis in real-time, direct access to information at different levels of enterprise management, predictive analytics, enhanced interactive visualization, intuitive interface, support for semantic information retrieval, and widespread and mobile access data (described in [6]). A characteristic feature of this new generation of BI systems is the reliance on the ontology and semantic retrieval of information. In architecture, there are new elements, such as ontology, ontology services and application domain ontology. The use of ontologies and visual information retrieval within analytical tools can be helpful to solve the following problems [8, p. 215]:

- definition of business rules in order to get proactive information and advice in the decision-making process,
- specification of semantic layer describing the relationships between different economic concepts,
- presentation of business information,
- the rapid modification of existing databases and data warehouses.

Development of BI systems is progressing towards the use of visual information retrieval based on a semantic Web. One of the main artifacts in the emergence of the semantic Web is an ontology. In the framework of the project we have elaborated the ontology, covering the scope and the data discussed in the paper.

IV. ONTOLOGICAL FOUNDATION

One way to represent knowledge in information systems is an ontology. In the literature, you can find many definitions of ontology. However, most often the term refers to the definition given by T. Gruber, who describes it as "an explicit specification of a conceptualization" [9, p. 907]. So ontology is a model that defines formally the concepts of a specific area and the semantic relations between them. In literature research of an ontology using in BI systems are described [10; 11;12;13; 14; 15].

Many research projects show that creating an ontology of economic and financial indicators is advantageous in decision making [8]. This is important, because there is no single universal system of economic indicators that would be used in all organizations. Besides, a lot of companies use a number of assessment models of business based on the analysis of various indicators. An ontological approach to modeling

domain knowledge was proposed in the project "Intelligent dashboard for managers (InKoM)" (described in [1; 16]). The main objective of the project was to create an intelligent cockpit for managers of small and medium-sized enterprises, which facilitates analysis and interpretation of the economic situations of the company, and supports analysis of economic and financial data Three new components of the cockpit are important: the financial ontology, the data mining algorithms, and the mechanism of deep retrieval on the Internet. This solution has enabled adequate, expandable and adaptable mapping of economic and financial knowledge, without having to modify the existing system of TETA BI. The new system significantly extended the usefulness of the existing Decision Support System [17].

In developing an intelligent interface, the ontological approach was completed by eye-tracking methods. Available in the BI system, a visual presentation of data permits one quickly to assess the economic situation and take appropriate action. To discover a way of analysing financial reports and statements managers, financial analysts, and students of economics participated in the experiments. The eye-tracking logs during the reading of these documents were used to discover the patterns of operations and to model the financial knowledge of each of the participants. The concepts and analytical operations performed by each manager were matched with the financial ontology available in the system.

V. Case Study – Evaluation of a Company's Profitability

Profitability analysis is necessary to assess the effectiveness of an economic entity. Sales revenues are essential to generate cash inflows that feed any company. These cash flows are used not only to purchase materials, services, and other goods, but also to make investments in fixed assets and working capital necessary to sustain ongoing operations. Profit is, however, a universal measure used in assessment of the financial situation. Unfortunately, the fact that an increase in revenues and earnings does not always generate cash inflows triggers off a serious risk of misinterpretation. Thus, the lack of cash inflows may generate serious financial problems. Profitability analysis without taking into account changes in cash position may yield wrong business decisions [18]. This is why the use of an ontology that describes in detail the issue of profitability allows one to avoid the risk of misinterpretation.

Managers of SMEs require a strong semantic reach and easy to use methods and techniques to support decision making. The aim of the experiment was to examine the usefulness of a financial ontology in the analysis of selected financial indicators and metrics in supporting operational decision making. Participants in the experiment had access to the system containing financial reports and ontology diagrams supporting the analysis of corporate profitability. For the purposes of our research, the balance sheet and profit and loss account of a real company were submitted. The financial statement included information that indicated the seemingly positive performance of the case company. In the periods analyzed, however, serious problems occurred due to bad debts. Those problems should have been perceived by the analyst as threatening the loss of the rationale behind the going concern's basic assumption, i.e. the ability to function without the threat of liquidation for the foreseeable future.

The experiment was initiated by the analysis of an internal managerial report containing widely used measures of financial situation, namely: liquidity profitability, debt, and turnover ratios. The most important factor directly observed by the participants of the experiment was a significant increase in profitability in the analyzed period.

In order to explain the profitability ratio correctly there was a need to take advantage of the ontology describing the issue of profitability, which is presented in Figure 1.

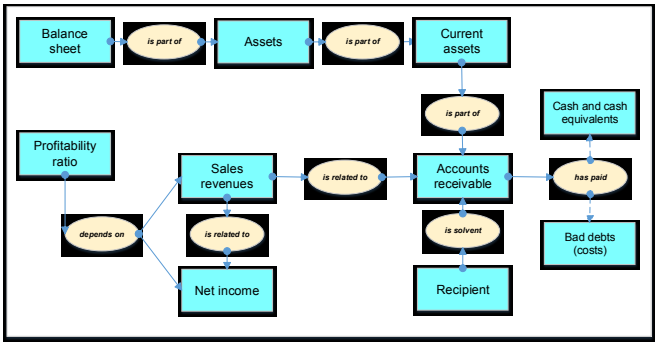


Fig. 1. Ontology: view of profitability

Profitability is the most important measure of effectiveness while running a business. According to the scheme presented in Figure 1, profitability is dependent on sales revenue and total net profit earned by the company. The aim of profitability analysis is to obtain information about the amount of the surplus earned from the sale which remains at the disposal of managers. Ontology describing corporate profitability allows the manager to properly analyze the growth of profitability ratio. It is necessary to examine the increase in revenues, net income as well as changes in cash amount and accounts receivable. According to the ontology, an increase in both profitability and accounts receivable is an unfavorable symptom. A positive sign suggested by the ontology was the increase in profitability along with the increase in cash.

Rate of return on sales is one of the key measures of operational efficiency, which informs us about the margin that has been realized on sales. Increasing ratio of return on sales is a positive symptom for the company because it may confirm that the proper decisions were made with regard to cost reduction or making good business deals allowing the company to increase sales revenues in excess of costs [19, pp. 135–164].

Profit margin, on the one hand, indicates the actual part of company revenues which remains at the disposal of managers and contributes to cover operating costs. On the other hand, profit earned may be used to increase equity or be paid out to the owners as dividends. The greatest disadvantage of return on sales ratio is that it is computed on the accrual basis [20].

The most fundamental concept of accounting is accrual principle that is a requirement of both International Financial Reporting Standards and Generally Accepted Accounting Principles. This concept requires keeping a record of business transactions (such as the rise in revenues or accounts receivable) in the period in which they actually occur, not in the period in which the cash flows related to them occur.

This means that sales transactions recorded as revenues in the accounting system increase net income, which, however, due to various external factors, may never be realized in the form of actual cash inflows (e.g. due to the bankruptcy of the recipient or the initiation of legal proceedings on claims, etc.).

According to the ontology presented in Figure 1, as the profitability increases it is necessary to check whether the increase in revenues is reflected in actual cash flows which are needed in running a business on a daily basis. Increase in profitability may be also triggered off by the increase in accounts receivable, which might be perceived as an unfavorable symptom.

Cash inflows from sales are the most important components of cash flows from operating activities as, in the opinion of many scholars, the most important element of cash flow statement, because in this part information about the most common business operations has been aggregated. This is a segment which usually generates excess cash that is used to finance investment processes, and to purchase fixed assets and intangible assets, as well as to repay financial obligations [2]. The shortage of cash flows from operating activities occurring in subsequent years should be perceived as an unfavorable symptom, because that usually indicates impending bankruptcy if the appropriate remedial steps are not taken.

The problem that causes significant misinterpretations in the proper assessment of the rate of return on sales is the issue of the so-called bad debts. The most common reason for the financial problems of many small and medium-sized companies is a problem with collection of accounts receivables. In such cases credit risk is understood as a potential loss due to the lack of collection of receivables. According to legal requirements in Poland, if there are any doubts with regard to the possibility of collecting receivables, these should be regarded as uncollectable bad debts. Thus, bad debts are those whose probability of recovery is low or close to zero. These accounts are a serious problem for the company, because they require to deal with impairment as well as legal enforcement proceedings that may be problematic.

In the experiment, it was noted that the increase in profitability is due to the accrual principal, and thus virtual, because as revenues significantly increased, accounts receivable increased even more, as presented in Figure 2. The decision maker using financial reports and ontology is guided through analytical steps, which are labeled on Figure 2 with numbers from 1 to 8.

In the analyzed period, sales revenues increased from 732 thousand PLN to 813 thousand PLN. According to the ontology, it is not recommended to base analysis solely on revenue growth, because the comprehensive analysis must include the detailed examination of changes in receivables. In the analyzed period, accounts receivable in the balance sheet increased dramatically from 145 thousand PLN to 321 thousand PLN. Thus, the ontology allows to avoid a threat of misinterpretation with regard to an apparent increase in profitability.

Financial analysis of a company's standing should be comprehensive. The example of such analysis with regard to profitability is presented in Figure 2. Based on the ontology, it is not sufficient to analyze managerial reports containing financial indicators, but it is necessary also to focus on the selected items from the financial statement. In order to properly explain the increase in profitability, the analyst is required to deal with percentage change analysis with regard to:

- revenues from sales,
- net income,
- short-term receivables,
- cash and cash equivalents.

Comprehensive analysis of the most important factors affecting the level of profitability ratio should be completed by sending a warning signal from the decision support system. As presented in Figure 2, the warning signal indicates that there is a need to take a decision to amend the sales process management. In the analysed case, in order to improve average collection of receivables it is recommended to:

- apply debt collection and enforcement techniques (in a short-term transaction),
- look for customers in different market segments, who do not have problems with paying on time (in a long-term one).

Suggestions proposed by the decision support system should be considered against the market conditions and customs of the local industry. Excessive use of debt collection activities can generate a negative signal to potential new customers who will search for suppliers among competing companies patiently waiting for payment. Such actions to be taken by managers were suggested by means of the warning signal generated by the system as it was presented in Figure 2.

Changing the segment of the target market may be feasible in the long term, thus such a decision should be preceded by insightful analysis. Moreover, it is necessary to examine the potential impact of restructuring on the future performance of the company. In addition, it is necessary to analyse the company's ability to carry out the investment project oriented towards the company's adjustment to new customers.

Any decision should be revised using the same techniques that helped to identify the original problem. In our experiment, we studied the financial performance demonstrated by the company that was achieved in the subsequent period. In Figure 3, we present an analysis of the research findings and their interpretation.

1

Managerial report		2012	2013
Profitability ratio	%	15,6%	19,2%

3

Profit and loss account		2012	2013
Sales revenues	PLN	732 167,0	812 967,0
Net income	**PLN**	**114 435,0**	**156 097,0**

5

Balance sheet / Assets		2012	2013
Accounts receivable	PLN	145 567,0	321 170,0

7

Balance sheet / Assets		2012	2013
Cash and cash equivalents	PLN	245 138,0	249 876,0

6 What was the change in cash?

8 **INTERPRETATION:**
Virtual increase in profitability

SUGGESTION:
Extensive debt collection activities

Fig. 2. Supporting the proper interpretation of profitability ratios (task 1)

1

Managerial report		2013	2014
Profitability ratio	%	19,2%	21,6%

3

Profit and loss account		2013	2014
Sales revenues	PLN	812 967,0	978 045,0
Net income	**PLN**	**156 097,0**	**211 078,0**

5

Balance sheet / Assets		2013	2014
Accounts receivable	PLN	321 170,0	339 067,0

7

Balance sheet / Assets		2013	2014
Cash and cash equivalents	PLN	249 876,0	311 479,0

6 What was the change in cash?

8 **INTERPRETATION:**
Satisfactory increase in profitability

SUGGESTION:
Continuation of existing practices

Fig. 3. Supporting the proper interpretation of profitability ratios (task 2)

Analysis of the solutions applied in the previous period allows them to be considered as accurate. Figure 3 presents the performance of the company, among which the most important signal is the increase in sales revenue and the

increase in cash flow. Simultaneously, the level of accounts receivable has stabilized. Figure 3 presents the financial performance of the case company. Similarly to Figure 2, in Figure 3 we labeled analytical steps to be performed by the decision maker with numbers from 1 to 8. Among other measures, the most important signs are the increase in revenues from 813 thousand PLN to 978 thousand PLN and an increase in cash amount from 250 thousand PLN to 311 thousand PLN while the short-term receivables amounts remained stable (not exceeding 10%).

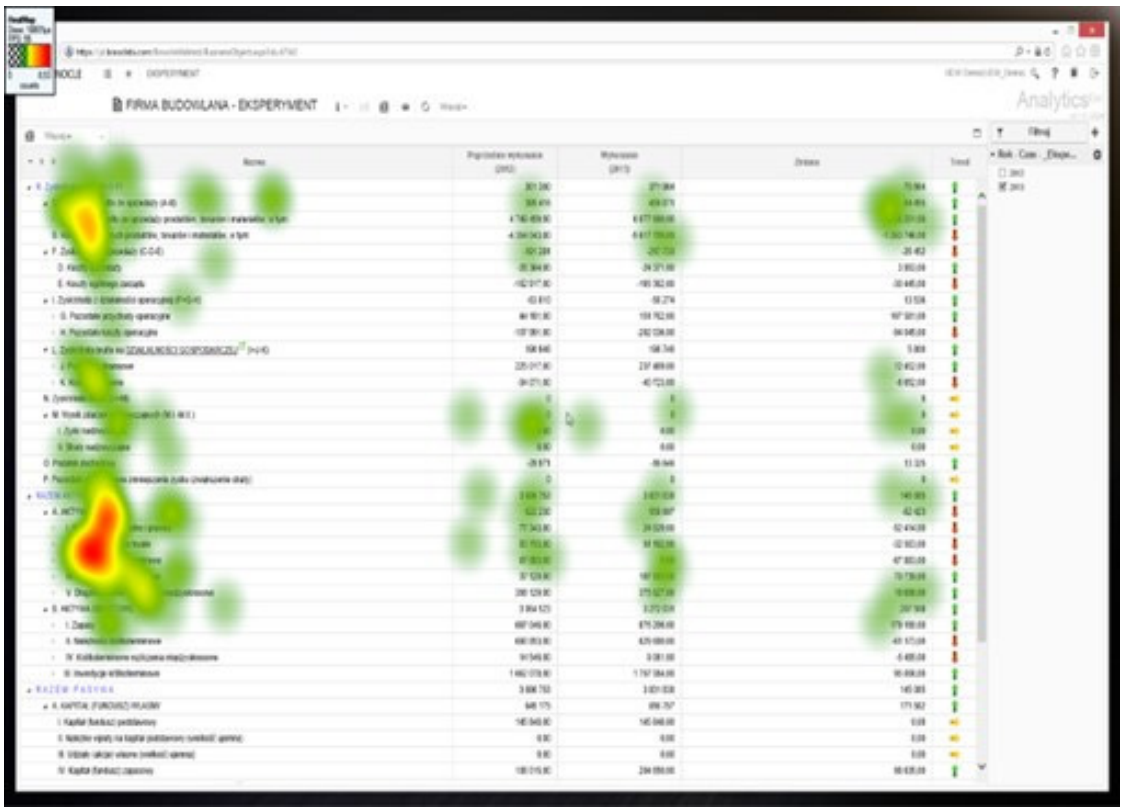


Fig. 4. Example heat map

This highlights the relevance of decisions taken by means of a decision support system based on a developed ontology. The proper interpretation of the information contained in the financial statements varies depending on the qualifications of novices and experienced managers. Novices only paid attention to the increase in revenues and profit growth, but neglected the change in accounts receivables. Most of the experienced managers focused their attention on the growth of accounts receivables. The behavioral patterns of experienced managers confirmed the usefulness and comprehensiveness of the ontology. Bad debts are a serious problem which should be analyzed first, before the revenue growth. Decision making without the use of ontology may lead to wrong conclusions, e.g. introducing sales promotion to increase sales growth. Using the ontology enabled the appropriate decisions to be made with respect to receivables management.

Comprehensive insight into both the selected managerial indicators and metrics as well as the changes in the value of chosen items from the financial statement is required for the proper interpretation of a company's performance. This would not be possible without the use of systematic knowledge in the form of ontology. Ontology is therefore an essential element of the decision support system that allows us to make the right business decisions oriented toward company development and avoiding the threat of bankruptcy.

VI. Approach to Modelling of Managerial Knowledge

In the construction of a new decision support system interface, an important role is played by the manager profile, in particular identifying his or her financial and economic knowledge. There are a number of methods for knowledge and reactions modelling of managers: crowdsourcing [21], testing physiological brain [22], explorations of the observation of the manager's eye movements, otherwise called eye tracking. In the project this last test method was applied. Eye tracking provides the opportunity to obtain various results, among which may be mentioned: the raw data, the scanning path, and a heat map. Figure 4 illustrates the sequence of analysis of the Balance Sheet by a manager. The diagram shows the parts of the document which have particularly attracted the attention of the manager: the area

close to the color red. Green areas on the thermal map show other areas for attention.

The results of the experiments raised three questions: (1) how and why does the experienced manager proceed; (2) how to identify the short-comings and errors committed by inexperienced managers, and (3) how to recognize a manager's fatigue and stress. The answers to these and other questions provided patterns of schematic perception of economic information. Using process mining methods, the preliminary rules and sequences of activities were discovered for novices and managers with experience. Raw quantitative data, collected by tracking software, can be analysed and used in the process of searching for perception patterns. As a result of the transformation of the data, it is possible to obtain a model of the analytical operations of managers. Such data, along with analytical models and ontology of financial analysis, are the basis of an intelligent decision support system. Patterns, models of financial analysis, and ontology are considered as knowledge in an intelligent financial decision support system.

VII. Conclusions and future research

This article presents a new attempt to create an intelligent interface for a Decision Support System. The idea was founded on the use of domain ontology and methods of patterns discovery of managerial knowledge from eye-tracking logs. The studies show that managers of small and medium-sized enterprises often have little benefit from achievements in the field of broadly understood financial analysis and IT technology, to strengthen their competitive position on the market and maintain financial credibility. The problem is often caused by the lack of the knowledge required to correctly interpret the financial reports as well as economic indicators. Also, they have difficulty with the knowledgeable use of the information systems that contain too many functions and tools that exceed their knowledge and skills. Usually, numerous ways of reporting, and the complex visualization tools, deepen the difficulties in the perception of the financial situation of an enterprise and its environment.

The use of improved visualization and access to semantic searches are primary characteristics of the BI 2.0 system. One of the main artefacts in semantic networks is the ontology of business information. A created ontology is required in which data and relationships of objects from various fields of knowledge are defined.

The experiments showed the usefulness of ontology where each concept and relationship can be considered in a multifaceted way. In the case studies, we have demonstrated that decision making without the use of ontology may lead to wrong conclusions, e.g. introducing sales promotion to increase sales growth. Analysis of indicators alone without relation to other components of financial reporting can lead to the choice of improper interpretations. Ontology is therefore an essential element in the decision support system that allows a manager to make correct business decisions oriented toward company development and avoiding the threat of bankruptcy.

The use of eye-tracking allowed us to develop preliminary patterns of analysis of financial reports and interpret the most important indicators.

The sequences of actions, discovered on the basis of the eye-tracking of experts, allows for defining of best practices to analyze financial data. However, these patterns are not sufficient to generate the interface. They must be associated with the financial ontology that help to interpret the data and provide adequate recommendations.

Preliminary results of the experiments demonstrated large differences between the experienced managers and novices in the manner of data perception and a way of analysis of financial reports. To develop a user-oriented interface, further in-depth studies on identifying the level of manager knowledge are needed. The project will be continued, especially on modeling of manager knowledge using eye-tracking. Models of patterns exploration in analytical sequences of actions will also be developed.

Acknowledgements. The authors thank the staff and the companies Bilander and Tobii for their support in developing the prototype. Thanks go also to the financial experts: Wojciech Hasik, Mariola Kotłowska and Wojciech Ostojski, and students of the Faculty of Management, Informatics and Finance of Wrocław University of Economics.

References

[1] J. Korczak, H. Dudycz and M. Dyczkowski, "Intelligent dashboard for SME managers. Architecture and functions", in: *Proc. of the Federated Conference on Computer Science and Information Systems*, M. Ganzha, L. Maciaszek, M. Paprzycki, Eds., Polskie Towarzystwo Informatyczne, IEEE Computer Society Press, Warsaw, Los Alamitos, CA, 2012, pp. 1003–1007

[2] M. Samonas, *Financial Forecasting, Analysis and Modelling*, John Wiley and Sons, Chichester 2015

[3] H. Dudycz, Visualization methods in Business Intelligence systems – An overview, *Business Informatics (16). Data Mining and Business Intelligence*, J. Korczak, Ed., Research Papers of Wrocław University of Economics, 2010, no. 104, pp. 9-24.

[4] D. Sell, L. Cabral, E. Motta, J. Domingue and R. Pacheco, Adding Semantics to Business Intelligence, 2008, http://dip.sema.nticweb.org/documents/WebSpaperOUV2.pdf.

[5] N. Raden, Business Intelligence 2.0: Simpler, More Accessible, Inevitable, 2007 http://www.informationweek.com/news/software/bi/197002610

[6] S. Nelson, *Business Intelligence 2.0: Are we there yet?* SAS Global Forum, 2010 http://support.sas.com/resources/papers/proceedings10/040-2010.pdf

[7] J. Trujillo, A. Mate, Business Intelligence 2.0: a general overview, in: *Business Intelligence: First European Summer School, eBISS 2011,* M.-A. Aufaure, E. Zimanyi, Eds., Lecture Notes in Business Information Processing (LNBIP) 96, Springer, 2012, pp. 98-116

[8] H. Dudycz., *The Topic Map as a Visual Representation of Economic Knowledge* (in Polish), Wrocław University of Economics, Wrocław, 2013

[9] T. R. Gruber, *Toward Principles for the Design of Ontologies Used for Knowledge Sharing*, Technical Report KSL, Knowledge Systems Laboratory, Stanford University, 1993, http://tomgruber.org/writing/onto-design.pdf

[10] M. Aruldoss, D. Maladhy and V. Prasanna Venkatesan, A framework for business intelligence application using ontological classification, International Journal of Engineering Science and Technology, 2011, vol. 3, no. 2, pp. 1213-1221

[11] A. Cheng, Y.-C. Lu, C. Sheu, An ontology-based business intelligence application in financial knowledge management system, Expert Systems with Applications, 2009, vol. 36, issue 2, part 2, pp. 3614-3622

[12] B. Neumayr, M. Schrefl, K. Linner, Semantic cockpit: an ontology-driven, interactive Business Intelligence tool for comparative data analysis, in: Advances in Conceptual Modeling. Recent Developments and New Directions, Lecture Notes in Computer Science (LNCS) 6999, Springer-Verlag Berlin Heidelberg 2011, pp. 55-64

[13] F. Pinto, M. F. Santos and A. Marques, Ontology based data mining – A contribution to business intelligence, 10th WSEAS International Conference on Mathematics and Computers in Business and Economics (MCBE '09), Czech Republic, 2009, March 23-25, pp. 210-216.

[14] H. Saggion, A. Funk, D. Maynard and K. Bontcheva, Ontology-based information extraction for business intelligence, in: Proceedings of the 6th International Semantic Web Conference and 2nd Asian Semantic Web Conference, Busan, Springer, Berlin/Heidelberg, 2007, pp. 843-856.

[15] D. Sell, D. C. da Silva, F. D. Beppler, M. Napoli, F. B. Ghisi, R. Pacheco and J. L. Todesco, SBI: a semantic framework to support business Intelligence, in Proceeding of the first international workshop on Ontology-supported business intelligence, Article no. 11, ACM New York, 2008.

[16] J. Korczak, H. Dudycz and M. Dyczkowski, "Design of financial knowledge in dashboard for SME managers", in: Proc. of the 2013 Federated Conference on Computer Science and Information Systems. Annals of Computer Science and Information Systems, vol. 1, M. Ganzha, L. Maciaszek, M. Paprzycki, Eds. Polskie Towarzystwo Informatyczne, IEEE Computer Society Press, Warsaw, Los Alamitos, CA, 2013, pp. 1111–1118

[17] M. Dyczkowski, J. Korczak and H. Dudycz, Multi-criteria evaluation of the intelligent dashboard for SME managers based on scorecard framework, in: Proc. of the 2014 Federated Conference on Computer Science and Information Systems. Annals of Computer Science and Information Systems, M. Ganzha, L. Maciaszek, M. Paprzycki, Eds., New York City, 2014, vol. 2, pp. 1147–115

[18] K. Berman, J. Knight and J. Case, Financial Intelligence, Harvard Business Review Press, Boston 2013

[19] B. Nita, Managerial Reporting (in Polish), Warszawa: PWN, 2014

[20] L. Revsine, D.W. Collins, W.B. Johnson and H.F Mittelstaedt, Financial Reporting and Analysis, McGraw Hill, New York 2012

[21] E. Schenk, C. Guittard, Towards a characterization of crowdsourcing practices, Journal of Innovation Economics & Management, 2011, no. 7, pp. 93-107

[22] H. J. Hwang, K. Kwon and C. H. Im, Neurofeedback-based motor imagery training for brain-computer interface (BCI), Journal of Neuroscience Methods, Republic of Korea, 2009

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1273–1282
DOI: 10.15439/2016F285
ACSIS, Vol. 8. ISSN 2300-5963

Information and Communication Technologies for Supporting Prosumers Knowledge Sharing – Evidence from Poland and United Kingdom

Ewa Ziemba
Faculty of Finance and Insurance
University of Economics in Katowice
ul. 1 Maja 50, 40-287 Katowice
Poland
ewa.ziemba@ue.katowice.pl

Monika Eisenbardt
Faculty of Finance and Insurance
University of Economics in Katowice
ul. 1 Maja 50, 40-287 Katowice
Poland
monika.eisenbardt@ue.katowice.pl

Roisin Mullins
Faculty of Business and Management
University of Wales Trinity Saint David
Lampeter, Ceredigion
United Kingdom
r.mullins@uwtsd.ac.uk

***Abstract*—Information and communication technologies (ICTs) can enhance the knowledge sharing by lowering temporal and spatial barriers between prosumers and enterprises, and improving access to prosumers' knowledge for enterprises. A major challenge for enterprises involves investing in the appropriate ICTs that help facilitate prosumers' knowledge engagement and knowledge transfer. The purpose of the paper is to indicate which ICTs are currently used and expected to be used by prosumers for knowledge sharing. The reported outcomes are the result of a questionnaire survey that yielded responses from 783 Polish and 171 UK based prosumers. The results indicate the primary ICT choices for use and expected use by Poland and UK based prosumers revealing important differences between these countries. The mobile applications being favored amongst the UK respondents whereas the dedicated enterprise website is the favored ICT amongst Polish respondents. Further, the variety of ICTs provided by enterprises may be too limiting to promote the type of knowledge sharing and communications expected to reassure the prosumers.**

I. Introduction

Knowledge is currently viewed as a fundamental driver for the commercial success of enterprises and is crucial to their competitive advantage [1]-[4]. Moreover, customer knowledge becomes an essential intangible asset for every line of business [5], leads to better response and respect toward customers [6], [7], makes a contribution toward new and innovative products [8], [9], contributes to the improvement of business value [10], and enhances the competitiveness of businesses [11].

Consumers who share their knowledge with enterprises with the aim of creating values and benefits for enterprises and their own consumption are known as "prosumers," whereas the process in which they share knowledge with enterprises is consistent with the notion of "prosumption" [12]-[17]. In general, prosumption refers to situations in which prosumers share knowledge not only with enterprises, but also with other prosumers to produce things of value for enterprises, and also for themselves.

Given that advances in ICTs have made it easier to share knowledge and these ICT developments have made the world increasing interconnected, many enterprises recognise there are challenges to employ the appropriate ICTs to facilitate knowledge sharing with prosumers. In considering the complexity of prosumption, prosumers' knowledge sharing initiatives and the variety of ICTs, enterprise must often confront these challenging tasks in deciding what type of ICTs to deploy in support of their prosumption initiatives.

The existing studies mostly examine ICTs for knowledge management in enterprises [18]-[23]. Researchers argued that ICTs play an important role in acquiring, codifying, storing, creating, sharing and applying knowledge that can be crucial for effective decision making and control at all levels.

The authors of this paper, following an extensive review of the literature, did not uncover any deep studies to interpret how ICTs support prosumers' knowledge sharing with enterprises. This reveals a need to study the ICTs that should be adopted and used by enterprises to better enable prosumers' knowledge sharing. Therefore, conducting research among prosumers and enterprises should contribute to greater understanding of the use of ICTs for prosumers' knowledge sharing and should help fill the gap in the existing body of knowledge.

In light of the above limitations, this paper focuses on investigating the choice of ICTs supporting prosumers' knowledge sharing in Poland and the UK. Its aim is to indicate the ICTs that are currently used by prosumers in comparison with the preferred ICTs expected to be used by prosumers.

The paper is structured as follows. Section I is an introduction to the subject. Section II states the theoretical background of ICTs for supporting prosumers' knowledge sharing and poses a research question. Section III describes the research methodology. Section IV presents the research findings on ICTs used and expected to be used by prosumers to facilitate knowledge sharing. Section V provides the study's contributions and limitations, and implications for the findings and considerations for future investigative work.

II. Theoretical Background And Research Question

A. Concept of prosumption and prosumer

The concept of prosumption has emerged from consumption theory. It focuses on the role that can be played by pro-active consumers willing to cooperate with enterprises.

The term '*prosumer'* was coined by Toffler [15]. According to him, selected enterprises' tasks (mainly manual

tasks), previously performed by enterprises' employees, are increasingly performed by consumers in accordance with the *do-it-yourself* principle, and by implication consumers become co-creators of products and services. Indeed, the terms of prosumption and prosumer have evolved over the years [24], [25]. As a result, modern approaches to prosumption differ greatly from Toffler's proposal. Table 1 presents the characteristics of two approaches to prosumption: Toffler's approach versus the modern approach. The modern approach to prosumption emphasizes the creativity of prosumers, as well as being connected to the value of prosumers' knowledge for enterprises. Enterprises can use their knowledge to attain business goals and, as a consequence, engage prosumers in business tasks.

B. Prosumers' knowledge sharing with enterprises

Prosumers' knowledge is the most important asset for most sectors engaged in contemporary business and is one of the most important contributors in improving business value and enhancing business performance [6], [27]-[30]. It is categorized into three types [31], [32]:

- *Knowledge about prosumers* represents both prosumers' needs and requirements; it may encompass characteristics in prosumers' behavior, their demographics and previous purchasing patterns; it may allow an understanding of prosumers' motivation in order to adjust and personalize products' or services';
- *Knowledge for prosumers* is created to satisfy prosumers' needs; it may include knowledge about enterprises, products and services; it may support prosumers in their buying cycle, impact on prosumers' perception of enterprises and offers, and become the base of knowledge from prosumers; and
- *Knowledge from prosumers* is created through the prosumers' experience with enterprises; it may embrace ideas, thoughts, reviews, opinions, discussions, advice and rankings that enterprises receive from their prosumers and use them to enhance their products and services.

The sharing of these various kinds of prosumers' knowledge between the prosumers and enterprises is critical in order to produce things that are of value not only for enterprises but also for prosumers. It could be characterized as a process in which prosumers' knowledge is exchanged among prosumers and enterprises. In this process prosumers share what they have learned and transfer what they know to enterprises that have a business interest in it and that have found this new knowledge to be useful for business improvement [33]. In this process the value of knowledge appreciates when it is shared [34].

In this study, the term "prosumers' knowledge sharing" means providing *knowledge from prosumers* (prosumers' ideas of products developments and creation, thoughts, reviews, opinions, discussions, advice and rankings) to enterprises and other prosumers. This approach is in line with the proposal of Wang and Noe, who distinguished knowledge sharing from knowledge exchange [35].

TABLE I.
NATURE OF PROSUMPTION

Toffler's approach	**Modern approach**
Prosumer's role	
Less complex tasks, previously carried out by enterprises' employees are performed by prosumers	• Sharing knowledge and experience with enterprise • Participating in enterprise business processes
Prosumer's knowledge	
Tasks performed by prosumers were tightly connected with their manual skills	Prosumers' knowledge is a source of innovative, creative business solutions, and processes improvement
Prosumer's relationship with enterprises	
Static, based on taking over of less important tasks from employees and performing them themselves	Active, based on collaboration, co-participation, co-design, and co-creation
Prosumers' communication with enterprises	
One-way, impeded, most often indirect	Two-way, multi-channel, easy and direct
Main advantages for enterprises	
Delegating simple tasks and activities to prosumers	• Using prosumers' knowledge for achieving business goals • Following prosumers' needs • Establishing relationships with prosumers and prosumer-friendly images of enterprises • Supporting enterprises' business processes by prosumers' knowledge
Main advantages for prosumers	
Self-service accordance with prosumers expectations	• Expressing own opinions about enterprises and their products • Adjusting products or services to own needs • Getting various types of financial and non-financial rewards

Source: own elaboration on the basis of [26].

According to them, "knowledge exchange includes both knowledge sharing (or employees providing knowledge to others) and knowledge seeking (or employees searching for knowledge from others)." It should however be noted that "knowledge sharing" can be also used interchangeably with "knowledge exchange" [36].

C. ICTs supporting prosumers' knowledge sharing

Some studies show that ICTs, especially CRM systems [37], Business Intelligence systems [38], and social media [39]-[44] can be used for knowledge management.

Additionally, researchers have examined ICT-tools for knowledge sharing [19], [45]. Jiebing, Bin, and Yongjiang [46] provided a conceptual framework to explore the linking mechanisms between customer knowledge management and ICT-based business model innovation. Studies concerned with the role of ICTs in knowledge sharing enlist such primary technologies as blogs, e-mail systems, e-collaborative systems, e-forums, knowledge repository, instant messaging, audio conferencing, podcasts, video conferencing, and wiki in the context of challenges faced by the practitioners in distributed projects [47] or in the context of Nonaka and Takeuchi's SECI model [48]. The focus of the SECI model on knowledge creation explores the cycle of generating tacit knowledge through to explicit knowledge and recreating tacit knowledge. The knowledge change in the SECI model is summarised as tacit to tacit (Socialization), tacit to explicit (Externalization), explicit to explicit (Combination), explicit to tacit (Internalization) [48], [49].

Only a few of the studies explore the application of social media for sharing customer knowledge. For example Chua and Banerjee [40] presented how Starbucks redefined the roles of its customers through the use of social media by transforming them from passive recipients of beverages to active contributors of innovation. Jalonen [50] explored the interplay between knowledge and emotion in the organisational knowledge creation process in the context of social media. Okazaki et al. [51] found a clear connection among customer engagement, prosumption, and Web 2.0 in a context of service-dominant logic. Moreover, they identified social networks created by prosumers. Based on the literature review, Zembik [52] explored various types of social media and their role as source of knowledge about, for, and from customers. Ziemba and Mullins [32] proposed the conceptual customer stratification framework which explains the stages required by a business to observe customers social media discussions.

D. Research question

After extensively searching through the literature, it was observed that there is a research gap in the existing body of knowledge related to ICTs used currently by prosumers and expected to be used by them to support prosumers' knowledge sharing. Also there is no research focusing on comparative analysis between less developed countries (like Poland) and better developed (like the UK) in the above mentioned area. In order to bridge the gap this study examines ICTs facilitating Polish and UK based prosumers' knowledge sharing and focuses on addressing the following research question:

RQ: Which ICTs facilitating prosumers' knowledge sharing are currently used and expected to be used by prosumers?

III. Research Methodology

Research methods included a critical review of the literature, logical deduction, case studies, a survey questionnaire, and statistical analysis. The research process followed the following steps.

The first step. The critical review of existing studies related to "prosumption," "prosumer," "customer," "consumer," "knowledge," "knowledge sharing," "ICT," "information technology" enabled to examine some ICTs supporting prosumers' knowledge sharing. The review embraces five bibliographic databases: Ebsco, ProQuest, Emerald Management, Scopus and ISI Web of Knowledge.

The second step. Interpretation of the case studies reporting prosumers' knowledge sharing informed the identification of the ICTs that are used by prosumers to share knowledge with enterprises.

The third step. An initial pilot survey questionnaire was designed. The questionnaire was divided into two parts. After a few demographics questions all participants were obliged to answer the question: *Have you ever assessed or commented on products or companies, proposed products improvements to the companies or designed new products?* This question enabled the division of respondents into consumers (not active in this area) and prosumers (active ones). The questionnaire contained questions concerning specified ICTs employed by enterprises to support prosumers' knowledge sharing. The questions were: (1) Which ICTs offered by enterprises have you used to share your knowledge, ideas and proposals about products or enterprises? (2) If you could in a free and unlimited way share your knowledge about products or enterprises, propose ideas of products developments or design new products – please indicate which ICTs would you like to use? The former question was directed only to prosumers. The latter was directed to both – prosumers and consumers. Various kinds of ICTs were listed for those questions. For each listed ICTs the respondents could choose one of five responses, according to a 5-point Likert scale: (1) definitely not (never), (2) probably not, (3) I don't know (no answer), (4) probably yes, (5) definitely yes (many times).

The fourth step. In November 2014 the more in-depth pilot survey was conducted in Poland. The purpose was substantive and methodological scrutiny of the questionnaire. To conduct reliability analysis, Cronbach's coefficient alpha was used. Cronbach's alpha for 16 analyzed items was 0.881. Hinton et al. [53] suggested four different ranges of reliability, i.e. the excellent range (0.90 and above), the high

(0.70-0.90), the high moderate (0.50-0.70) and the low (0.50 and below). Thus, it can be concluded that the scale had high reliability and it could be used in the research process. Moreover, substantive scrutiny of the questionnaire enabled the researchers to perform minor changes in order to improve the quality of the questionnaire.

The fifth step. Applying the CAWI (Computer-Assisted Web Interview) method and employing the Polish platform Ankietka.pl, and the English platform Bristol Online Survey (BOS), hosted at the University of Bristol, the survey questionnaires was uploaded to the websites. Data collection took place between the end of December 2014 and March 2015 in Poland, and between February and April 2016 in the United Kingdom. In Poland, the designed sample size was 2.500 people, comprising people of different age, gender, and ICT skills. In the UK the online survey letter and URL was initially posted to 1000 individuals comprising people of different age, gender, and ICT skills, and presented to a random sample of the target population. Using online tools permits contact with an accessible audience as the survey appears on search engine lists due to metatags and appropriate placing of keywords.

After screening the responses and excluding outliers, there was a final research sample of 783 usable, correct and complete questionnaires from Poland and 171 from the United Kingdom. The data was stored in Microsoft Excel format. The demographic analysis of the research sample is presented in Table 2.

The sixth step. As the process of collecting data was completed the reliability was calculated. The Cronbach's alpha coefficient with all 16 items confirmed a high internal consistency (0.882). Additionally, the values of Cronbach's alpha for each item, with the assumption that a given item was deleted, were calculated. The Cronbach's alpha values for the items were between 0.883 and 0.845. The results showed that the removal of some items would not lead to the improvement of internal consistency among items on the scale. Overall, the original alpha scores with all 16 items show a strong internal consistency and reliability.

The seventh step. In order to answer the research questions the statistical analysis was employed. The descriptive analysis of ICTs was prepared; the mean, median, mode, and distribution of ICTs used and expected to be used by prosumers were calculated.

IV. Research Findings

A. ICTs used and expected to be used by prosumers

In order to answer the research question, detailed analysis concerning ICTs used and expected to be used by prosumers, to share knowledge about products or enterprises, propose ideas of products developments or design new products, was made. The results are presented in Table 3.

"Used ICTs" reflect which ICTs are currently offered by enterprises to prosumers and used by them to share knowledge. It is noticeable that ICTs used by Polish and UK

TABLE II.
DEMOGRAPHIC ANALYSIS OF THE RESEARCH SAMPLE

	Poland		**United Kingdom**	
Demographic profile	**Number of respondents**	**Percentage of respondents**	**Number of respondents**	**Percentage of respondents**
Gender				
female	599	76.5%	98	57.30%
male	184	23.5%	73	42.70%
Age				
Builders generation – over 65 years old	14	1.8%	8	4.68%
Baby-Boomers generation – 51–65 years old	35	4.5%	25	14.62%
X generation – 36–50 years old	108	13.8%	67	39.18%
Y generation – 21–35 years old	369	47.1%	68	39.77%
Z generation – less than 21 years old	257	32.8%	3	1.75%
Level of education				
higher education	217	27.7%	89	52.05%
secondary education	559	71.4%	75	43.86%
less than secondary education	7	0.9%	7	4.09%
Place of residence				
city with a population of more than 100.000	419	53.5%	96	56.14%
city with a population of less than 100.000	244	31.2%	53	30.99%
rural area	120	15.3%	22	12.87%

Source: own elaboration.

based prosumers varies a lot. Polish prosumers mainly use enterprises' websites (the mean value is 3.72), e-mails (the mean value is 3.52), and Internet forums (the mean value is 3.40). UK based prosumers mainly use mobile applications (the mean value is 4.01), Facebook fanpages (the mean value is 3.78), and enterprises' specialized applications (the mean value is 3.59). Interestingly, the Facebook fanpages result for the median and mode values are 4.00 for 'used ICT' for both the UK and Poland prosumers. It means that the majority of prosumers have ticked the answer 'probably yes', so they probably were using these ICTs to share knowledge with enterprises or other prosumers.

It is useful to underline, that differences between the mean values of a number of used ICTs are significant in both countries. The most substantial difference relates to mobile applications – the mean value is 2.56 for Poland, whereas it is 4.01 for UK. Similarly, the mean value of enterprises' specialized applications is 2.40 for Poland, whereas it is 3.59 for the UK. It indicates that UK based prosumers use those ICTs more frequently than Polish ones. The outcomes show that Polish prosumers use only popular information websites more frequently than UK based prosumers. The mean value is 2.82 and the mode value is 4.00 for Poland, whereas the mean value is 2.33 and the mode value is 2.00 for UK. Admittedly, the differences between the mean values are not significant for mobile applications and enterprises' specialized applications. Nonetheless, the mode values analysis shows that the majority of Polish prosumers have chosen the answer 'probably yes', so popular information websites are probably offered to them by enterprises; whereas the majority of UK based prosumers have chosen the answer 'probably not', so these websites are probably offered to them but these are not the preferred prosumer exchange choice.

The overall analysis of used ICTs shows that UK based prosumers use ICTs for knowledge sharing more frequently than Polish ones. In addition, Polish prosumers use mainly standard and well known ICTs, whereas UK based prosumers use the latest kinds of ICTs.

"Expected ICTs" reflect which ICTs are needed by prosumers to share knowledge. The research findings show that UK based prosumers mainly expect to engage using mobile applications. The mean value is 3.74. The median and mode values are 4.00. Furthermore, they expect to engage directly with enterprises' websites and Facebook fanpages (the mean values are 3.53 in both cases). Polish

TABLE III.
ICTS USED AND EXPECTED TO BE USED BY POLISH AND UK PROSUMERS ENGAGED IN KNOWLEDGE SHARING

ICTs	'Used ICTs'						'Expected ICTs'					
	POLAND			UK			POLAND			UK		
	Mean	Median	Mode	Mean	Median	Mode	Mean	Median	Mode	Mean	Median	Mode
E-mails	3.52	4	4	3.34	4	4	3.77	4	4	3.29	4	4
Internet forums	3.40	4	4	3.33	4	4	3.54	4	4	3.18	4	4
Enterprises' websites	3.72	4	4	3.53	4	4	4.00	4	4	3.53	4	4
Popular information websites	2.82	3	4	2.33	2	2	3.44	4	4	2.53	2	2
Industry specialized portals	2.87	3	4	3.26	4	4	3.57	4	4	3.25	4	4
Mobile applications	2.56	2	1	4.01	4	4	3.28	4	4	3.74	4	4
Enterprises' specialized applications	2.40	2	1	3.59	4	4	3.43	4	4	3.25	4	4
File sharing portals	2.54	2	1	2.96	2	2	3.16	3	4	2.87	2	2
Facebook fanpages	3.11	4	4	3.78	4	4	3.38	4	4	3.53	4	4
Crowdsourcing portals	1.61	1	1	2.26	2	2	2.34	2	2	2.33	2	2
Business blogs	1.98	2	1	2.51	2	2	2.85	3	4	2.64	2	2
Private blogs	2.18	2	1	2.24	2	2	2.73	3	2	2.25	2	2
Online auctions	2,99	3	4	2,47	2	2	3,07	3	4	2,40	2	2
Price comparison websites	2,99	3	4	2,92	3	4	3,38	4	4	2,77	2	2
Enterprises' helplines/ helpdesks	2.15	2	1	3.13	4	4	2.52	2	2	3.01	3	4
Online surveys	3.16	4	4	2.99	3	2	3.10	3	4	2.89	2	2

Source: own elaboration.

prosumers mainly expect to engage via enterprises' websites. The mean, median and mode values are 4.00. They also choose e-mails (the mean value is 3.77) and industry specialized portals (the mean values is 3.57).

The overall analysis of ICTs presented in Table 3 shows that in the case of Poland all the mean values of *"used ICTs"* (except for online surveys) are lower than the mean values of "*expected ICTs*". It may show that ICTs which are currently offered to Polish prosumers by enterprises may not meet their expectations. Thus, Polish prosumers would like enterprises to offer them a greater range of ICTs. It could influence their willingness to share their knowledge with enterprises. In the case of UK based prosumers the majority of the reported mean values for "*expected ICTs*" are slightly lower than the mean values for "*used ICTs*" (11 from 16). It may illustrate that ICTs which are currently offered to UK based prosumers by enterprises meet or even slightly exceed their expectations. Four ICTs are expected to be used to a higher degree than are currently used and are referred to as popular information websites, crowdsourcing portals, business blogs and private blogs. Perhaps an indication of a willingness to switch one ICT channel for another one where these ICTs may be seen to be more specific to envelop a critical mass of 'close' engagement and discussion which enhances the prosumers effort. Only one channel that of enterprises' websites reported the same median for expected ICTs and used ICTs. The differences between the mean values are not significant in any case.

In order to compare ICTs used and expected to be used by prosumers of both countries two analyses are presented below. The analyses embrace only these prosumers who ticked (4) or (5) answering the questionnaire questions. It is indicating that they probably or definitely use or expect to use ICTs to share knowledge.

B. ICTs used by prosumers – distribution analysis

The research findings identify the ICTs used by Polish and UK based prosumers to enable knowledge sharing with enterprises as shown in Figure 1.

Figure 1 shows that there are no significant differences between Polish and UK based prosumers related to standard ICTs used by them, such as e-mails, Internet forums, enterprises' websites, and price comparison websites. Nonetheless, there are significant differences concerning other ICTs used by Polish and UK based prosumers.

The biggest differences pertain to mobile applications (indicated by 88.2% of UK based prosumers in relation to 33% of Polish prosumers), enterprises' specialized applications (indicated by 71.1% of UK based prosumers in comparison with 25.5% of Polish prosumers), and enterprises' helplines/ helpdesks (indicated by 56.6% of UK based prosumers in relation to 18.8% of Polish prosumers). The outcomes show also that only in two cases – which are online auctions and popular information websites, Polish prosumers use them in a considerably greater range then UK. For example, online auctions were indicated by 46.6% of Polish prosumers in relation to 25% of UK based prosumers. Similarly, popular information websites were indicated by 39.7% of Polish prosumers in relation to 17.1% of UK based prosumers. Overall the analysis shows that UK based prosumers use and probably engage with UK based enterprises where the choice of ICTs for knowledge sharing is a more extensive range than the Polish enterprise ICT offer.

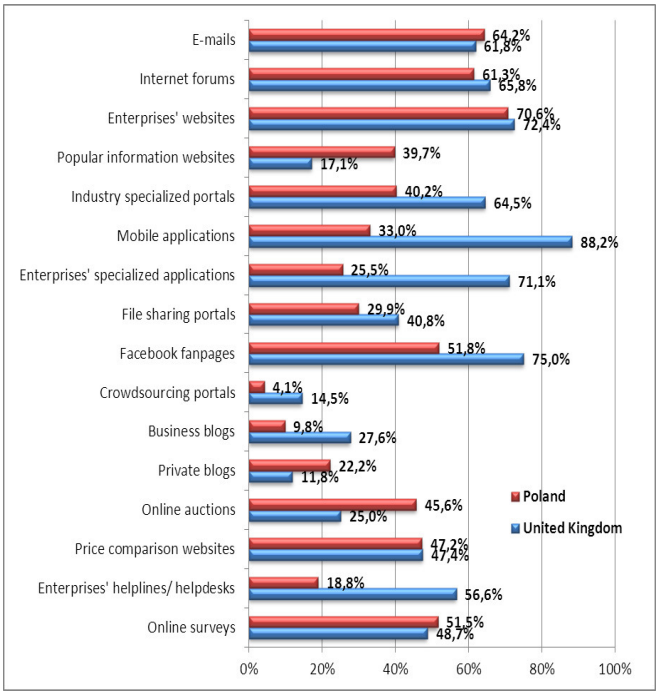


Fig. 1. ICTs used by prosumers for knowledge sharing
Source: own elaboration

C. ICTs expected to be used by prosumers – distribution analysis

The research findings of ICTs expected to be used by Polish and UK based prosumers for their knowledge sharing with enterprises is shown in Figure 2.

Figure 2 shows that in eleven cases (from a total of 16) Polish prosumers more frequently expect ICTs for knowledge sharing than UK based prosumers. The biggest difference relates to popular information websites indicated by 58.2% of Polish prosumers and 24.6% of UK based prosumers. Whereas, UK based prosumers more frequently expect enterprises' helplines/ helpdesks than Polish prosumers. This is indicated by 49.1% of UK based prosumers and 22.2% of Polish prosumers. Similarly, mobile applications are expected by 75.4% of UK based prosumers and 50.1% of Polish prosumers. The considerable difference relates also to online auctions indicated by 41.3% of Polish prosumers and 21.6% of UK based prosumers, as well as to price comparison websites indicated by 55% of Polish prosumers and 38.6% of UK based prosumers.

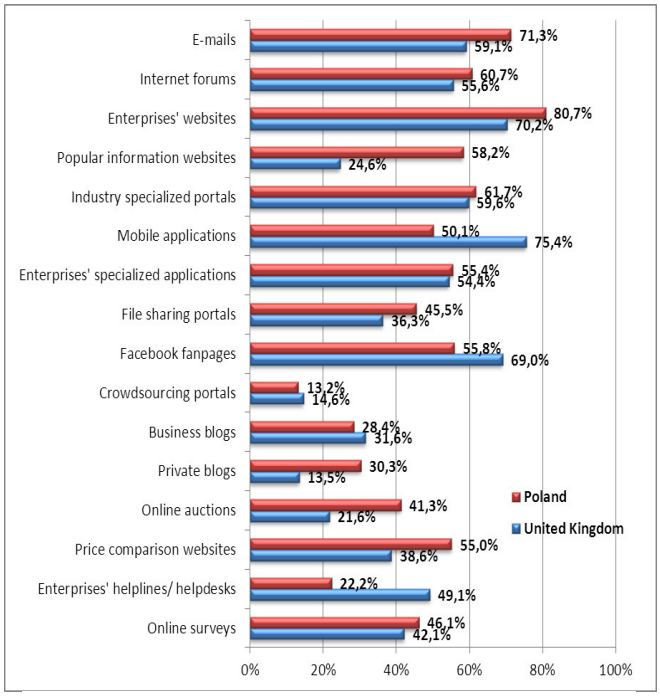


Fig. 2. ICTs expected to be used by prosumers for knowledge sharing
Source: own elaboration

Generally, the prosumers usage and expectations of ICTs for knowledge sharing are only slightly different in the UK. It may show that enterprises offer those ICTs for knowledge sharing that are expected by enterprises. Whereas, the prosumers usage and expectations of ICTs for knowledge sharing are significantly different in Poland. It may show that currently the enterprises do not meet their expectation.

V. Discussion of Findings

The trends identified in the demographics breakdown for the Polish and UK based respondents follow a similar pattern to those outlined in Smith [54]. Consistent with other studies females are more inclined to respond to surveys questioning the use and intended use of ICTs for knowledge sharing with enterprises as our findings show for Poland as well as the UK respondents.

There were similarities in the categorization of participants in the age categories, in the case of the builders generation the responses were 1.8% from Poland and 4.68% from the UK, and would not be unexpected given the training and technical competences of this generation and their culture of communicating more face-to-face rather than through online questionnaires. Also generation Y responses were 41.7% from Poland and 39.77% from the UK and this age category expect to use devices to communicate online and are comfortable with this mode of communication. These outcomes are consistent with research from authors [55], [56] whose goal was to elaborate critical success factors for ICTs adoption by people in Poland. The overall analysis of outcomes also shows differences between these generations. For generation Y it is self-satisfaction with e-products and e-services delivered by enterprises and public administration that is crucial, whilst for builders generation their awareness of ICTs is critical.

The differences reflected in the categorization of participants presents an interesting breakdown, where generation X reported Poland responses as 4.5% whereas 14.62% for the UK responses since this age range in the UK use technology in the workplace or home, are often self-taught in using technology and commit time to engaging in knowledge sharing. Generation X in responses from Poland was 13.8% while UK was 39.18% and this is a marked difference in responses indicating a possible culture of more accepted online communication in the UK for this age range. Finally generation Z indicated Poland at 32.8% whereas UK was only 1.75%, an interesting marked differences in responses and this needs further research to determine if the survey was more visible to this age range in Poland where their use of technology is embedded in their everyday social interactions. However, the research reported [55] indicated that for Polish generation X the most crucial success factors for ICT adoption is the need to make one's own live easier, whilst for Polish generation Z it is the financial situation of the household. It may partially explain the difference relating to generation Z prosumers. Polish prosumers use ICTs to enhance the opportunities of financial benefits or merits.

The educational differences reported between Poland and UK respondents fit well with the categorization of age responses especially as few responses from both countries were from those who are less educated, showing educational attainment may be an indicator for participating in knowledge sharing.

The respondent's place of residence was equally captured for both countries with half of the respondents from both countries living in a city with a population of more than 100.000, and this is interesting as the greater the chances to communicate offline in larger population centers the more likely the respondents are to use time for online communication, and this is a cultural communication shift noted in other recent studies.

The results indicated no significant differences between Polish and UK based prosumers in their use of standard ICTs, such as e-mails, Internet forums, and websites of enterprises. The most substantial difference in the ICTs used and expected by prosumers relates to mobile applications – the mean value is 2.56 for Poland (33%), whereas it is 4.01 for UK (88.2%) and this is consistent with research to support this use of mobile as an enabler of knowledge sharing.

The overall analysis of ICTs used shows that UK based prosumers use ICTs for knowledge sharing more frequently than Polish ones and this may be associated with levels of education achieved as there was almost twice as many UK (52.05%) responses than Polish respondents (27.7%) with higher education attainment.

The findings also show that mobile applications are the *'expected ICTs'* needed by UK based prosumers to share

knowledge, followed by use of Facebook fanpages and enterprises' websites. Whereas, Polish prosumers mainly expect to engage with use of enterprises' websites, followed by e-mails and industry specialized portals. Interestingly, crowdsourcing portals and blogs are those ICTs which, UK based and Polish prosumers, do not expect or cite a preference to use.

The findings also show that Polish prosumers more frequently expect to use ICTs for knowledge sharing than UK based prosumers who have an expectation to use the enterprises' helplines/ helpdesks instead or to complement their online initiated discussions.

A recommendation is that the enterprises need to take consideration of the culture of contemporary communication choices associated with the wide age ranges of prosumers. Finally the enterprises need to embed a comprehensive choice of ICT's particular to their prosumers needs to actively encourage knowledge sharing.

VI. Conclusions

A. Research contribution

This work contributes to existing research on prosumption, especially prosumers' knowledge sharing with the use of ICTs by:

- Indicating the ICTs currently used by prosumers to promote knowledge sharing with enterprises; and
- Indicating the ICTs expected to be used by prosumers to stimulate knowledge sharing with enterprises.

Firstly, this study indicates that mobile application use is expected to a greater degree by UK based prosumers. However, they also expect to use the enterprises' helplines/ helpdesks indicating the diverse expectations and somewhat divergent needs of UK based prosumers and the opportunities this presents to enterprises.

Secondly, the outcomes show that ICTs which are currently offered to UK based prosumers by enterprises meet or even slightly exceed their expectations. However, the prosumers usage and expectations of ICTs for knowledge sharing are significantly different in Poland suggesting that the enterprises do not meet their expectation, and this may result in less engagement in knowledge sharing.

B. Implication for research and practice

This study can be useful for researchers. They may use this methodology and do similar analyses with different sample groups in Poland, United Kingdom, and other countries; additionally many comparisons between different groups and countries can be made. Moreover, the methodology constitutes a very comprehensive basis for identifying ICTs to support knowledge sharing, both, about prosumers, as well as for and from prosumers, but researchers may develop, verify and improve this methodology and its implementation. In addition, researchers may use these research findings and employ them in studies of enterprises. Their goal could be the analysis of ICTs and the possibilities of adjusting them to the expectations of prosumers.

Moreover, for practitioners, the results of this study can be used to improve activities aimed at prosumption adoption, especially helping them understand which ICTs should be used to support prosumers' knowledge sharing.

C. Limitations and future research

As with many other studies, this study has its limitations. The first one was the selection of the survey respondents. Most of them were young people below 35 years old. It is advisable to extend the future research to elderly persons, inter alia prosumers 50+.

The second limitation was the relatively low number of respondents from United Kingdom in comparison with the number of respondents from Poland. Resulting from the low UK responses and timing of the survey the research will continue in the UK to ensure a higher response rate for deeper analysis. Since the initial results reveal interesting findings the research will continue to generate a higher response rate, and for this reason this paper is rather preliminary, and recognizes the analysis are not to be generalized. Therefore the research will be extended with detailed analysis in further works.

As the third limitation, it is possible to specify a methodological limitation. The research sample embraced only prosumers, not enterprises. It is advisable to extend the research to enterprises. All these above issues should be carefully considered and assimilated in the future works.

Acknowledgment

This research has been supported by a grant entitled "Transformation of business and public administration by information technology and information systems" from the University of Economics in Katowice, Poland, 2014-2016.

References

[1] A. Jaki and B. Mikuła, Eds., *Knowledge – Economy – Society. Managing Organizations: Concepts and Their Applications* (in Polish). Cracow: University of Economics, 2014.

[2] J. Kisielnicki, *Zarządzanie i informatyka* (in Polish). Warsaw: Placet, 2014.

[3] A. Kowalczyk and B. Nogalski, *Zarządzanie wiedzą*: *koncepcje i narzędzia* (in Polish). Warsaw: Difin, 2007.

[4] F.P. Omotayo, "Knowledge Management as an important tool in organisational management: A review of literature," *Library Philosophy and Practice* (e-journal), Paper 1238, 2015. Available at: http://digitalcommons.unl.edu/libphilprac/1238/.

[5] J. Rowley, "Eight questions for customer knowledge management in e-business," *Journal of Knowledge Management*, vol. 6(5), pp. 500–511, 2002, http://dx.doi.org/10.1108/13673270210450441.

[6] B. Aghamirian, B. Dorri, and B. Aghamirian, "Effects of customer knowledge management's eight factors in e-commerce," *Management Science and Engineering*, vol. 7(4), pp. 1–11, 2013.

[7] S.-M. Tseng, "The effect of knowledge management capability and customer knowledge gaps on corporate performance," *Journal of Enterprise Information Management*, vol. 29(1), pp. 51–71, 2016, http://dx.doi.org/10.1108/JEIM-03-2015-0021.

[8] D.C. Brabham, "Motivations for participation in a crowdsourcing application to improve public engagement in transit planning," *Journal of Applied Communication Research,* vol. 40(3), pp. 307–328, 2012, http://dx.doi.org/10.1080/00909882.2012.693940.

[9] W. Tsai, M. Tsai, S. Li, and C. Lin, "Harmonizing firms' knowledge and strategies with organizational capabilities," *Journal of Computer Information Systems,* 53(1), pp. 23–32, 2012.

[10] A.-M. Croteau and P. Li, "Critical success factors of CRM technological initiatives," *Canadian Journal of Administrative Sciences*, vol. 20(1), pp. 21–34, 2003, http://dx.doi.org/10.1111/j.1936-4490.2003.tb00303.x.

[11] E.-J. Song and M.-S. Kang, "A study on the platform of knowledge integration for customer feedback in B2C service industry," *International Journal of Information and Communication Technology*, vol. 8(1), pp. 26–36, 2016, http://dx.doi.org/10.1504/ijict.2016.073637.

[12] C. Fuchs, "Web 2.0. Prosumption, and Surveillance," *Surveillance & Society*, vol. 8(3), pp. 288–309, 2011.

[13] G. Ritzer and N. Jurgenson, "Production, consumption, prosumption: The nature of capitalism in the age of the digital 'prosumer'," *Journal of Consumer Culture March*, vol. 10(1), pp. 13–36, 2010, http://dx.doi.org/10.1177/1469540509354673.

[14] D. Tapscott and A.S. Williams, *Wikinomics: How mass collaboration changes everything*. New York: Penguin Group, 2006.

[15] A. Toffler, *The Third Wave*. New York: Bantam Books, 1980.

[16] C. Xie, R.P. Bagozzi, and S.V. Troye, "Trying to prosume: Toward a theory of consumers as co-creators of value," *Journal of the Academy of Marketing Science*, vol. 36, pp. 109–122, 2008.

[17] E. Ziemba, "Conceptual model of information technology support for prosumption," in *Proc. of Int. Conf. on Management, Leadership and Governance – ICMLG 2013*, pp. 355–363. Bangkok University, February 07-08, 2013.

[18] Ł. Osuszek and S. Stanek, "Knowledge management and decision support in adaptive case management platforms," in *Proc. of the 2015 Federated Conf. on Computer Science and Information Systems*, 2015, pp. 1539–1549, 2015, http://dx.doi.org/10.15439/2015f60.

[19] Ö.G. Bayram and H. Demirtel, "Effect of ICT on information sharing in enterprises: The case of Ministry of Development," in *Proc. of European Conf. on Knowledge Management ECKM*, Kidmore End: Academic Conferences International Limited, pp. 94–101, Sep. 2014.

[20] Y.-Y. Chen and H.-L. Huang, "Strategic orientation of knowledge management and information technology and their effects on performance," in *Proc. of Pacific Asia Conference on Information Systems PACIS 2014*, Chengdu, China, 24-28 June 2014.

[21] M.T. García-Álvarez, "Analysis of the effects of ICTs in knowledge management and innovation: The case of Zara Group," *Computers in Human Behavior*, vol. 51, pp. 994–1002, 2015, http://dx.doi.org/10.1016/j.chb.2014.10.007.

[22] F.N.D Piraquive, V.H.M García, R.G. Crespo, and D. Liberona, "Knowledge management, innovation and efficiency of service enterprises through ICTs appropriation and usage," *Lecture Notes in Business Information Processing*, vol. 185, pp. 300–310, 2014, http://dx.doi.org/10.1007/978-3-319-08618-7_29.

[23] R. Subashini, S. Rita, and M. Vivek, "The role of ICTs in knowledge management (KM) for organizational effectiveness," *Communications in Computer and Information Science*, vol. 270, pp. 542–549, 2012, http://dx.doi.org/10.1007/978-3-642-29216-3_59.

[24] M. Izvercianu, S. Seran, and C.F. Buciuman, "Changing marketing tools and principles in prosumer innovation management," *European Conf. on Management, Leadership & Governance*, Kidmore End: Int. Limited, pp. 246–255, 2012.

[25] D. Jelonek, C. Stępniak, and T. Turek, "Prosumpcja w regionalnych społecznościach elektronicznych dla potrzeb przedsięwzięć miejskich (in Polish)," in *Zeszyty Naukowe Uniwersytetu Ekonomicznego w Katowicach,* no. 243, J. Gołuchowski and A. Frączkiewicz-Wronka, Eds., Katowice: University of Economics, pp. 151–164, 2015.

[26] E. Ziemba and M. Eisenbardt, "Examining prosumers participation in business processes," *Online Journal of Applied Knowledge Management,* vol. 2(1), pp. 219–229, 2015

[27] A.S. Cui and F. Wu, "Utilizing customer knowledge in innovation: antecedents and impact of customer involvement on new product performance," *Journal of the Academy of Marketing Science*, pp. 1–23, March 2015, http://dx.doi.org/10.1007/s11747-015-0433-x.

[28] M.F.A.K. Panni, "CKM and its influence on organizational marketing performance: Proposing an integrated conceptual framework," in *Customer-centric marketing strategies: Tools for building organizational performance*, H.R. Kaufman and M.F.A.K. Panni, Eds., Hershey: IGI Global, pp. 103–125, 2015, http://dx.doi.org/10.4018/978-1-4666-2524-2.ch006.

[29] M.R. Shihab and A.A. Lestari, "The impact of customer knowledge acquisition to knowledge management benefits: A case study in Indonesian banking and insurance industries," in *Proc. of 2014 Int. Conf. on Advanced Computer Science and Information Systems*, pp. 301–306, 2014, http://dx.doi.org/10.1109/icacsis.2014.7065867.

[30] N. Taherparvar, R. Esmaeilpour, and M. Dostar, "Customer knowledge management, innovation capability, and business performance: A case study of the banking industry," *Journal of Knowledge Management*, vol. 3(18), pp. 591–610, 2014, http://dx.doi.org/10.1108/jkm-11-2013-0446.

[31] J.O. Chan, "Big data customer knowledge management," *Communications of the IIMA*, vol. 14(3), Article 5, 2014, Available at: http://scholarworks.lib.csusb.edu/ciima/vol14/iss3/5.

[32] E. Ziemba and R. Mullins, "Identifying more about customers: the phenomenon of the switch to the knowledge exchange," *Journal of Applied Knowledge Management*, vol. 4(1), pp. 165–179, 2016.

[33] M.Y Cheng, J.S.Y. Ho, and P.M. Lau, "Knowledge sharing in academic institutions: a study of multimedia university Malaysia," *Electronic Journal of Knowledge Management,* vol. 3(7), pp. 313–324, 2009.

[34] E. Ziemba and M. Eisenbardt, "Prosumers' eagerness for knowledge sharing with enterprises – a Polish study," *Online Journal of Applied Knowledge Management,* vol. 2(1), pp. 40–58, 2014.

[35] S. Wang and R.A. Noe, "Knowledge sharing: A review and directions for future research," *Human Resource Management Review*, vol. 20, pp. 115–131, 2010.

[36] H.-G. Lin, "Knowledge sharing and firm innovation capability: An empirical study," *International Journal of Manpower*, vol. 28(3/4), pp. 315–332, 2007.

[37] S. Bagheri, R.J. Kusters, and J. Trieneken, "Business-IT alignment in PSS value networks – Linking customer knowledge management to social customer relationship management," in *Proc. of 17th Int. Conf. on Enterprise Information Systems ICEIS*, Barcelona-Spain, pp. 249–257, 2015, http://dx.doi.org/10.5220/0005370002490257.

[38] M.C. Lee, "Business Intelligence, knowledge management and customer relationship management – Technological support in enterprise competitive competence," *Business Intelligence: Concepts, Methodologies, Tools, and Applications*, pp. 216–23, 2015, http://dx.doi.org/10.4018/978-1-4666-9562-7.ch011.

[39] P. Bharati, W. Zhang, and A. Chaudhury, "Better knowledge with social media? Exploring the roles of social capital and organizational knowledge management," *Journal of Knowledge Management,* vol. 19(3), pp.456–475, http://dx.doi.org/10.1108/jkm-11-2014-0467.

[40] A.Y.K. Chua and S. Banerjee, "Customer knowledge management via social media: The case of Starbucks," *Journal of Knowledge Management*, vol. 17(2), pp. 237–249, 2013, http://dx.doi.org/10.1108/13673271311315196.

[41] D.P. Ford and R.M. Mason, "A multilevel perspective of tensions between knowledge management and social media," *Journal of Organizational Computing and Electronic Commerce*, vol. 23(1-2), pp. 7–33, 2013, http://dx.doi.org/10.1080/10919392.2013.748604.

[42] C. Heller-Baird and G. Parasnis, "From social media to social customer relationship management," *Strategy & Leadership*, vol. 39(5), pp. 30–37, 2011, http://dx.doi.org/10.1108/10878571111161507.

[43] M. Levy, "WEB 2.0 implications on knowledge management," *Journal of Knowledge Management*, vol. 13(1), pp. 120–134, 2009, http://dx.doi.org/10.1108/13673270910931215.

[44] Z. Zhang, "Customer knowledge management and the strategies of social software," *Business Process Management Journal*, vol. 17(1), pp. 82–106, 2011, http://dx.doi.org/10.1108/14637151111105599.

[45] Å. Fast-Berglund and E. Blom, "Evaluating ICT-tools for knowledge sharing and assembly support," in *Proc. of the 5th Int. Conf. on Applied Human Factors and Ergonomics AHFE 2014*, T. Ahram,

W. Karwowski and T. Marek, Eds., pp. 2734–2742, Krakow, Poland, 19-23 July, 2014.

[46] W. Jiebing, G. Bin, and S. Yongjiang, "Customer knowledge management and IT-enabled business model innovation: A conceptual framework and a case study from China," *European Management Journal,* vol. 31(4), pp. 359–372, 2013, http://dx.doi.org/10.1016/j.emj.2013.02.001.

[47] M.A. Razzak, R. Ahmed, "*Knowledge sharing in distributed agile projects: techniques, strategies and challenges*," in *Proc. of the 2014 Federated Conf. on Computer Science and Information Systems*, pp. 1431–1440, 2014, http://dx.doi.org/10.15439/2014f280.

[48] S.C. Lee and R.S. Kelkar, "ICT and knowledge management: Perspectives from SECI model," *The Electronic Library*, vol. 31(2), pp. 226–243, 2013, http://dx.doi.org/10.1108/02640471311312401.

[49] I. Nonaka, "Dynamic theory of organizational knowledge creation," *Organization Science*, vol 5(1), pp. 14–37, 1994, http://dx.doi.org/10.1287/orsc.5.1.14.

[50] H. Jalonen, "Social media and emotions in organisational knowledge creation," in *Proc. of the 2014 Federated Conf. on Computer Science and Information Systems*, pp. 1371–1379, 2014 http://dx.doi.org/10.15439/2014f39.

[51] S. Okazaki, A.M. Díaz-Martín, M. Rozano, and H.D. Menéndez-Benito, "Using Twitter to engage with customers: a data mining approach," *Internet Research*, vol. 25(3), pp. 416–434, 2015, http://dx.doi.org/10.1108/intr-11-2013-0249.

[52] M. Zembik, "Social media as a source of knowledge for customers and enterprises," *Journal of Applied Knowledge Management*, vol. 2(2), pp. 132–148, 2014.

[53] P.R. Hinton, C. Brownlow, I. McMurvay, and B. Cozens, *SPSS Explained*. East Sussex: Routledge, 2004.

[54] G. Smith, "Does gender influence online survey participation?: A record-linkage analysis of university faculty online survey response behavior," *ERIC Doc. Reproduction Service*, no. ED 501717, 2008. Available at: http://eric.ed.gov/?id=ED501717.

[55] E. Ziemba, Ed., *Czynniki sukcesu i poziom wykorzystania technologii informacyjno-komunikacyjnych w Polsce* (in Polish), Warsaw: CeDeWu, 2015.

[56] E. Ziemba, "Factors affecting the adoption and usage of ICTs within Polish households," *Interdisciplinary Journal of Information, Knowledge, and Management*, vol. 11, pp. 89–113, 2016.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1283–1287
DOI: 10.15439/2016F216
ACSIS, Vol. 8. ISSN 2300-5963

Knowledge integration in multi-agent decision support system for financial e-services

Marcin Hernes
Wrocław University of Economics
ul. Komandorska 118/120, 53-345 Wrocław, Poland
Email: marcin.hernes@ue.wroc.pl

Jadwiga Sobieska-Karpińska
The Witelon State University of Applied Sciences in Legnica,
Sejmowa 5A, 59-220 Legnica, Poland
Email: jadwiga.sobieska.karpinska@gmail.com

Abstract— **Providing financial e-services in the all areas involves taking decisions processes. Existing systems, also multi-agent systems, usually include only one of the earlier mentioned areas, and they are closed systems, available only to a small group of users. In addition, agents' knowledge in these systems is characterized by a certain degree of heterogeneity. Since in the decisive process one final decision is required, knowledge shall be automatically integrated. The aim of this paper is presentation of the author's method for knowledge integration in multi-agent decision support system of financial e-services. The first part of paper presents an architecture of the developed system, functioning of selected agents and a structure of agents' knowledge representation. Next, the developed method for integration of knowledge has been described. The last part of paper presents the results of research experiment to evaluate the effectiveness of the system and the developed method.**

I. Introduction

In the era of e-economy one can observe a sudden increase in the level of offered e-services connected with finances, embracing all financial services available to clients via the Internet [1]. The Polish Agency for Enterprise Development [2] has defined financial e-services as all operations connected with finances conducted via electronic media. These types of services are provided within the following areas [2, 3, 4]:

- investment (securities, currencies),
- banking,
- insurance,
- financial consulting,
- managing one's own finances,
- payments.

Providing financial e-services in the mentioned areas involves taking decisions processes. These processes most often need to be executed in a near-real time, and they are always characterized by a certain degree of risk. They are often supported by multi-agent decision support system [5]. Such system may generate decisions constituting hints or tips to investors, or alternatively decisions may be taken automatically (taking into account criteria specified by an investor – e.g. return rate, risk). However, existing systems usually include only one of the earlier mentioned areas, and they are closed systems, available only to a small group of users. It needs to be noticed that in multi-agent decision support system, agents use various sources of data and different methods of supporting decisions. Consequently, variants of decisions presented by individual agents may differ. Therefore, agents' knowledge is characterized by a certain degree of heterogeneity. Since in the decisive process one final decision is required, knowledge shall be automatically integrated. The integration shall be performed in relation to a given area, and also in relation to all areas financial e-services are provided.

Related works on the subject also mention various methods of knowledge integration, for example negotiations [6], or deduction-calculation methods [7]. Negotiations enable effective integration of knowledge by reaching a compromise, however they require exchanging a large number of communications between agents, which results in decreased efficiency of the multi-agent system. The deduction-calculation methods (e.g. ones based on the theory of games, classical mechanics, or methods of choice) enable one to obtain a great computational or calculation capacity of a system, however they do not guarantee a proper result of knowledge integration [8]. Thus, applying negotiation or deduction and calculation methods cannot guarantee an adequate level of satisfaction from taken decisions. In order to eliminate the presented problems, consensus methods may be applied which enable integration of knowledge in a real time and guarantee reaching a good compromise at a lower level of risk, which may consequently lead to selecting decisions producing profits satisfactory for a decision maker [8, 9]. In a consensus each party/side is taken into account, each party "loses" the least, and each one contributes to the consensus, and all parties accept the consensus, so the consensus constitutes a representation of all agents.

The aim of this paper is presentation of the author's method for knowledge integration in multi-agent decision support system for financial e-services. The first part of paper presents an architecture of the developed system, functioning of selected agents and a structure of agents' knowledge representation. Next, the developed method for integration of knowledge has been described. The last part of paper presents the results of research experiment to evaluate the effectiveness of the system and the developed method.

II. ARCHITECTURE OF SYSTEM AND DESCRIPTION OF SELECTED AGENTS

Multi-agent decision support system for financial e-services consist of following elements:

1. Collectives of agents. The purpose of the members of the collective is the analysis of information from the market, generating a decision and taking an action. Each agent in the collective running on the basis of a different method of decision support. Each collective makes decisions from a different area (for example, Collective 1 makes the banking decisions, Collective 2 makes the investment decisions, Collective 3 makes the insurance decisions). Collective members knowledge state is represented by uniform structure (described in section V).
2. Knowledge integration module, which integrate the knowledge of individual members of the collectives (by using a consensus method - integration is performed independently for each collective). One, final decision is determined (for each collective determined a separate decision). The final decision is then presented to users.
3. Users - financial investors or software agents making decisions on behalf of the investor.

In our prototype of system, agents running in order to determine decisions in the area of investment and banking use methods of technical analysis, fundamental analysis, artificial intelligence (such as genetic algorithms, artificial neural networks, expert systems), two collectives : financial investments and banking. The next part of paper describes functioning of randomly selected agents: *MixedTechnical*, *BollingerPlus* and *Fundamental*.

A. . The MixedTechnical agent

MixedTechnical agent has been developed on the Java Agent Development Framework (JADE), which is a platform to facilitate the creation of agents and multi-agent systems in the Java programming language [10]. Agent generates buy/sell decisions on the basis of commonly used technical analysis indicators (Lento, 2008): Average Directional Index (ADX), Relative Strenght Index (RSI), Rate of Change (ROC), Commodity Channel Index (CCI), Moving Average of Oscillator (OsMA), Moving Average Convergence Divergence (MACD), Stop and Reverse (SAR), Williams %R, Moving Average (MA). Agent generates decisions depending on what the decision is suggested by a larger number of indicators used. Agent's knowledge is represented by using three-values logic (value „1" denotes „buy" decision, value „-1" denotes „sell", decision, value „0" denotes „do nothing").

B. . The BollingerPlus agent

The BollingerPlus agent is created on the basis of the Bollinger Bands indicator [11]. These bands are volatility constraints placed above and below a moving average. Volatility is expressed by the standard deviation, which changes as volatility increases and decreases.

The bands automatically widen when volatility increases and narrow when volatility decreases. The buy decision's probability level is calculated when the price is close to the upper Bollinger Band or breaks above it, and the sell decision is calculated when the price is close to the lower Bollinger Band or falls below it.

C. The Fundamental agent

Agent *Fundamental* makes decisions using fundamental analysis. For this purpose he performs the analysis of text documents [1] containing the experts' opinions on the economic situation or the organization's situation. The main purpose of the analysis is to determine the general sentiment of opinion, i.e. to determine whether the opinion is positive (suggesting a "buy" decision) or negative (suggesting a "sell" decision) or neutral (suggesting the "leave unchanged" decision). The analysis is done by the agent built by using The Learning Intelligent Distribution Agent (LIDA) architecture [12]. The advantage of this architecture is its emergent-symbolic character, making it possible to process both structured (numerical and symbolic) and unstructured (stored in natural language) information. This agent consist of following modules: sensory memory, perceptual memory, workspace, episodic memory, declarative memory, attentional codelets, global workspace, action selection, sensory-motor memory. The functioning of the agent is performed in the frame a cognitive cycle.

Considering the process of analysis of expert opinions, the environment of agent functioning is a set of text documents containing these opinions (opinions are placed e.g. on financial portals). Agent looking for opinions, and then stores them in a repository (system's database).

Text analysis is performed in the following way:

1. A semantic network containing terms and connections between them is created in the perceptual memory on the basis of a learning set. The perceptual memory stores also synonyms and different variations of words (thesaurus). In the perceptual memory of LIDA agents terms are represented by means of nodes, whereas connections are represented by means of links.
2. Individual text documents are added one by one into the sensory memory.
3. Opinions are analyzed by codelets, i.e. programs which search through texts according to certain criteria specified by means of configuration parameters.
4. Results of analysis, in the semantic network form, are transferred to the workspace (a current situational model is created).
5. In the next step, the situational model is passed to the global workspace and from the procedural memory the following patterns of action are automatically selected: „saving results of opinion analysis into a data base and „loading next opinion into the sensory memory".

[1] Text documents' analysis, performed by LIDA agent, has been characterized in work [13] in details.

III. METHOD FOR KNOWLEDGE REPRESENTATION

Each agent running within the given collective presents its decision in the form of a specific structure of knowledge. In the considered system a structure developed by [10] has been used. The structure is definied as follows:

Definition 1

A structure for representation a decision D of finite set of financial assets[2] $E=\{e_1,e_2,\ldots,e_N\}$ is called a sequence:

$D=\left\langle\{EW^+\},\{EW^\pm\},\{EW^-\},Z,SP,DT\right\rangle$ where:

1) $EW^+=\left\langle e_o,pe_o\right\rangle,\left\langle e_q,pe_q\right\rangle,\ldots,\left\langle e_p,pe_p\right\rangle$.

Couple $\left\langle e_x,pe_x\right\rangle$, where: $e_x\in E$ and $pe_x\in[0,1]$ denote a financial asset and this asset's participation in set EW^+.

Financial asset $e_x\in\left\langle e_x,pe_x\right\rangle$ is denoted by e_x^+ when $\left\langle e_x,pe_x\right\rangle\in EW^+$. The set EW^+ is called a positive set; in other words, it is a set of financial assets with respect to which an agent has the knowledge or information that they should be buy.

2) $EW^\pm=\left\langle e_r,pe_r\right\rangle,\left\langle e_s,pe_s\right\rangle,\ldots,\left\langle e_t,pe_t\right\rangle$.

Couple $\left\langle e_x,pe_x\right\rangle$, where: $e_x\in E$ and $pe_x\in[0,1]$ denote a financial asset and this asset's participation in set $EW^\pm$.

Financial asset $e_x\in\left\langle e_x,pe_x\right\rangle$ is denoted by $e_x^\pm$ when $\left\langle e_x,pe_x\right\rangle\in EW^\pm$. The set $EW^\pm$ is called a neutral set, in other words, it is a set of financial assets, with respect to which an agent has no knowledge or information whether to buy or sell them. If these assets are held by an investor, they should not be sold, or if they are not in the possession of the investor, they should not be bought.

3) $EW^-=\left\langle e_u,pe_u\right\rangle,\left\langle e_v,pe_v\right\rangle,\ldots,\left\langle e_w,pe_w\right\rangle$.

Couple $\left\langle e_x,pe_x\right\rangle$, w here: $e_x\in E$ and $pe_x\in[0,1]$, denote a financial asset and this asset's participation in set EW^-.

Financial asset $e_x\in\left\langle e_x,pe_x\right\rangle$ is denoted by e_x^- when $\left\langle e_x,pe_x\right\rangle\in EW^-$. The set EW^- is called a negative set; in other words it is a set of financial assets with respect to which an agent has the knowledge or information that they should be sell.

4) $Z\in[0,1]$ - decision rate of return forecast.

5) $SP\in[0,1]$ - degree of certainty of rate Z. It can be calculated on the basis of the level of risk related to the decision.

6) DT - date of decision.

IV. METHOD FOR KNOWLEDGE INTEGRATION

Fig 1 presents schema of the knowledge integration process.

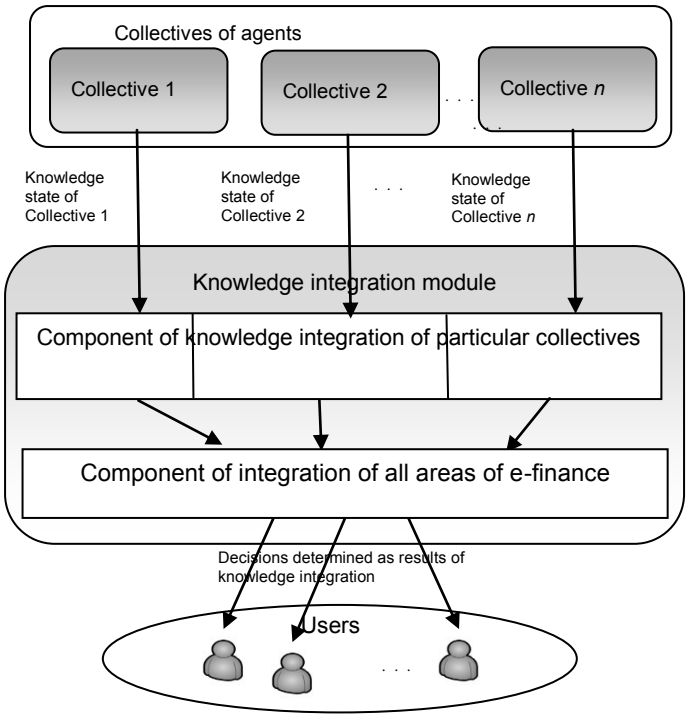


Fig. 1. Schema of the knowledge integration process

In order to knowledge integration, a consensus algorithm is used, formally defined as follows:

Input: Profile (set) $A=\{A^{(1)}, A^{(2)}, \ldots. A^{(M)}\}$consist of M structures of knowledge.
Output: Consensus
$CON=\left\langle CON_+, CON_\pm, CON_-, CON_Z, CON_{SP}, CON_{DT}\right\rangle$ for profile A.
START
1: $CON_+, CON_\pm, CON_- = \varnothing$, $CON_Z, CON_{SP}, CON_{DT} = 0$.
2: $CON_Z = \frac{1}{M}\sum_{i=1}^{M} Z^i$.
3: $CON_{SP} = \frac{1}{M}\sum_{i=1}^{M} SP^i$.
4: $CON_{DT} = \frac{1}{M}\sum_{i=1}^{M} DT^i$
let $d := \sum_{i=1}^{M}\left[\Psi\left(CON, A^{(i)}\right)\right]^2$ **and** $j:=1$.
5: If $e_j \in CON_+$ **then** $CON':=\left\langle CON_+\setminus\{e_j\}, CON_\pm, CON_-, CON_Z, CON_{SP}, CON_{DT}\right\rangle$
Go to:8,
If $e_j \notin CON_+$ **then** go to: 6.
6: If $t_+(j) = 0$ **then** go to: 9.

[2] assets including cash assets as a contractual right to receive cash assets , the right to exchange financial instruments with another entity under favorable conditions, and equity instruments issued by other entities [14].

7: If $e_j \cap CON \neq \emptyset$ **and** $e_j \in CON_{\pm}$ **or** $e_j \in CON_{\pm}$ **then**

$CON' = \langle CON_+ \cup \{e_j\}, CON_{\pm} \setminus \{e_j\}, CON_- \setminus \{e_j\}, CON_Z, CON_{SP}, CON_{DT} \rangle$

If $e_j \cap CON = \emptyset$ **then**

$CON' = \langle CON_+ \cup \{e_j\}, CON_{\pm}, CON_-, CON_Z, CON_{SP}, CON_{DT} \rangle$

8: If $\sum_{i=1}^{M} \left[\Psi(CON', A^{(i)})\right]^2 < d$ **then** $d := \sum_{i=1}^{M} \left[\Psi(CON', A^{(i)})\right]^2$ **and** $CON := CON'$.

9: If $e_j \in CON_{\pm}$ **then** $CON' := \langle CON_+, CON_{\pm} \setminus \{e_j\}, CON_-, CON_Z, CON_{SP}, CON_{DT} \rangle$. Go to: 12.

If $e_j \notin CON_{\pm}$ **then** go to: 10.

10: If $t_{\pm}(j) = 0$ **then** go to: 13.

11: If $e_j \cap CON \neq \emptyset$ **and** $e_j \in CON_+$ **or** $e_j \in CON_-$ **then**

$CON' = \langle CON_+ \setminus \{e_j\}, CON_{\pm} \cup \{e_j\}, CON_- \setminus \{e_j\}, CON_Z, CON_{SP}, CON_{DT} \rangle$

If $e_j \cap CON = \emptyset$ **then**

$\langle CON_+, CON_{\pm} \cup \{e_j\}, CON_-, CON_Z, CON_{SP}, CON_{DT} \rangle$.

12: If $\sum_{i=1}^{M} \left[\Psi(CON', A^{(i)})\right]^2 < d$ **then**

$d := \sum_{i=1}^{M} \left[\Psi(CON', A^{(i)})\right]^2$ **and** $CON := CON'$, go to: 13.

else go to: 16.

13: If $e_j \in CON_-$ **then**

$CON' = \langle CON_+, CON_{\pm}, CON_- \setminus \{e_j\}, CON_Z, CON_{SP}, CON_{DT} \rangle$ **and** go to: 16.

14: If $t_-(j) = 0$ **then** go to: 17.

15: If $e_j \cap CON \neq \emptyset$ **and** $e_j \in CON_+$ **or** $e_j \in CON_{\pm}$ **then**

$CON' = \langle CON_+ \setminus \{e_j\}, CON_{\pm} \setminus \{e_j\}, CON_- \cup \{e_j\}, CON_Z, CON_{SP}, CON_{DT} \rangle$

If $e_j \cap CON = \emptyset$ **then**

$CON' = \langle CON_+, CON_{\pm}, CON_- \cup \{e_j\}, CON_Z, CON_{SP}, CON_{DT} \rangle$.

16: If $\sum_{i=1}^{M} \left[\Psi(CON', A^{(i)})\right]^2 < d$ **then**

$d := \sum_{i=1}^{M} \left[\Psi(CON', A^{(i)})\right]^2$ **and** $CON := CON'$.

17: If $j<N$ **then** $j := j+1$. Go to: 2

else: STOP.

STOP.

V. RESEARCH EXPERIMENT

The aim of the experiment was to examine the effectiveness of the system and developed method for knowledge integration. The collective knowledge of agents on the area of investment and knowledge integration module (named Agent Supervisor), was evaluated. Data from randomly selected quotations has been used (Sygnity company). Test was conducted, which the following assumptions:

1. Quotations from randomly selected periods was used:
   - 01-02-2016 9.30 to 06-02-2016 17.00,
   - 08-02-2016 9.30 to 13-02-2016 17.00,
   - 15-02-2016 9.30 to 20-02-2016 17.00.
2. Following agents was evaluated:
   - MixedTechnical,
   - BollingerPlus,
   - Fundamental,
   - Supervisor.
3. It was assumed that the initial capital, which has an investor is 1000 PLN, and the rate of return on investment assumed difference between that amount and the amount the investor will have the last sell transactions in a given period. The rate of return is expressed in nominal units (PLN).
4. The transaction cost was not considered.
5. It was assumed that in each transaction the investor engages 100% of the capital held.
6. Analysis of the quality of collective knowledge was carried out using the following measurement (ratios):
   - rate of return (ratio x_1),
   - the number of profitable transactions (ratio x_2),
   - the number of unprofitable transactions (ratio x_3),
   - the average rate of return per transaction (r. x_4),
   - Sharpe ratio (ratio x_5),
   - the average coefficient of variation (ratio x_6).

   In order to comparison quality of collective knowledge, the following function was used[15]:

$$y = (a_1 x_1 + a_2 x_2 + a_3 (1 - x_3) + a_4 x_4 + a_5 x_5 + a_6 (1 - x_6)).$$

   where x_i denote the normalized values of ratios mentioned in item 6 from x_1 to x_6. It was adopted in the test that coefficients a_1 to a_6=1/6.
7. The results obtained by the tested agents were compared with the results of the Buy-and-Hold (B&H) benchmark [3]– table 1.

Summing up the results of the evaluation of knowledge of a group of agents and the *Supervisor* agent one may notice that in the considered periods their decisions generated both profits as well as losses. While evaluating the efficiency of the system one needs to take into account not only the return rate but also other indicators, including the level of risk connected with an investment, which the evaluation function employed in the article enables. In the first period, *BollingerPlus* proved to be the best agent, and the remaining agents got a higher note than the *B & H* benchmark note. In the second period, the *Supervisor* agent received a note higher that the remaining agents and the *B & H* benchmark. Taking into account the third period, one may notice that the ranking of notes looks similar to the second period.

Taking into consideration all the periods one may conclude that most often (2 of 3 periods) the highest note was given to the *Supervisor* agent, even though the return rate achieved by the agent had not always been the highest one. The note however results from a low level of risk connected with investing on the basis of decisions generated as a result of knowledge integration. However, in all periods, the *MixedTechnical* agent received low notes as due to a relatively high level of risk it generated little return rates. The *Fundamental* agent received average level notes in all periods which may be connected with the fact that the agent generated a very small number of decisions.

[3] benchmark is also implemented as agent's algorithm.

TABLE I.
RESULTS OF AGENTS PERFORMANCE ANALYSIS

Agent's name	Period	Rate of return	Number of transaction		Rate of return per transaction	Sharpe ratio	Average coeffecient of variation	Evaluation function
			profitable	unprofitable				
MixedTechnical	1	3,25	5	3	0,41	0,46	23,72	0,25
	2	-83,68	4	7	-7,61	-1,03	14,18	0,18
	3	34,17	4	1	6,83	0,57	6,44	0,50
BollingerPlus	1	10,10	5	2	1,44	0,64	18,29	0,51
	2	-15,14	3	5	-1,89	-0,43	7,83	0,34
	3	25,48	5	2	3,64	0,29	11,87	0,42
Fundamental	1	2,43	4	2	0,41	0,78	8,21	0,26
	2	-21,75	1	3	-5,44	0,64	6,45	0,53
	3	25,43	2	1	8,48	0,71	2,34	0,48
Supervisor	1	2,26	3	1	0,57	0,82	2,47	0,48
	2	-3,97	2	1	-1,32	0,73	1,98	0,56
	3	21,30	3	0	7,10	0,94	2,90	0,59
B & H	1	-25,41	0	1	-25,41	0	0	0,11
	2	-73,78	0	1	-73,78	0	0	0,08
	3	13,63	1	0	13,63	0	0	0,22

. On the basis of the research experiment one can draw the conclusion that integration of agents' knowledge enables selecting decisions which produce benefits satisfactory to a user. Integration of knowledge performed by the Supervisor agent using consensus determining algorithm is performed in a near to real time. The agent selects a final decision on the basis of suggestions generated by all remaining agents, which consequently leads to a decreased level of risk.

VI. CONCLUSION

Functioning of a multi-agent decision support system for financial e-services requires continuous, automatic integration of agents' knowledge. The process enables the elimination of decisions generated by group members whose knowledge status has been evaluated as being poor, which means that decisions taken by such agents may most of the time bring unsatisfactory results. Thanks to that, their influence on the final decision established with the use of knowledge integration module and presented to a user is eliminated. Additionally developed algorithm enables taking into account summed-up knowledge of a group as it includes knowledge of all members of a group. The issues discussed in the paper imply further research concerning, for example, implementation of agents performing behavioural analysis and developing a multistep method of integration which would include improving knowledge of agents.

REFERENCES

[1] K. Dandapani, "Success and failure in web-based financial services", *Communications of the ACM*, 47(5), pp.31–33., 2004.

[2] PARP, *Definicja e-finansów* http://www.web.gov.pl/e-finansowanie/76_141.html, (05.12.2015).

[3] J. Hu and N. Zhong, "A Multilevel Integration Approach for Developing E-Finance Portals: Challenges and Perspectives", in: *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence* (WI '07). IEEE Computer Society, Washington, DC, USA, 2007, pp. 825-828. DOI=http://dx.doi.org/10.1109/WI.2007.121.

[4] K. Narayanasamy, D. Rasiah and T. M. Tan, "The adoption and concerns of e-finance in Malaysia", *Electron. Commerce Res.* 11 (4), 2011, pp. 383–400.

[5] A. Abroud, Y., V. Choong, S. Muthaiyah, D. Fie and G. Yong, "Adopting e-finance: decomposing the technology acceptance model for investors", *Service Business* 9 (1), 2015, pp. 161-182.

[6] P. Dyk and M. Lenar, "Applying negotiation methods to resolve conflicts in multi-agent environments", in: *Multimedia and Network Information systems*, MISSI 2006 Zgrzywa A. (ed.), Oficyna Wydawnicza PWr, Wrocław 2006.

[7] J.P. Barthlemy, "Dictatorial consensus function on n-trees", *Mathematical Social Science*, nr 25, 1992.

[8] M. Maleszka, B. Mianowska and N. T. Nguyen "A method for collaborative recommendation using knowledge integration tools and hierarchical structure of user profiles", *Knowledge-Based Systems*, Volume 47, pp. 1-13, 2013.

[9] M. Hernes and J. Sobieska-Karpińska , "Application of the consensus method in a multi-agent financial decision support system", *Information Systems and e-Business Management* 14 (1), Springer Berlin Heidelberg, 2016, DOI: 10.1007/s10257-015-0280-9.

[10] JADE Tutorial, http://jade.tilab.com/doc/tutorials/JADEProgramming-Tutorial-for-beginners.pdf (15.12.2015).

[11] J. Bollinger, *Bollinger on Bollinger Bands*, McGraw Hill, 2001.

[12] S. Franklin, F. G. Patterson, "The LIDA architecture: Adding new modes of learning to an intelligent, autonomous, software agent", in: *Proceedings of the International Conference on Integrated Design and Process Technology*. CA: Society for Design and Process Science, San Diego, 2006.

[13] A. Bytniewski, M. Hernes, "Analiza opinii klientów o produkcie dokonywana w kognitywnym zintegrowanym systemie informatycznym zarządzania", In: Porębska-Miąc, T., Sroka, H. (eds.), *Systemy Wspomagania Organizacji*. Katowice: Wydawnictwo Uniwersytetu Ekonomicznego w Katowicach, 2014.

[14] G. K. Świderska, W. Więcław (eds.), *Sprawozdanie finansowe według polskich i międzynarodowych standardów rachunkowości*. Difin/MAC sp. z o.o, Warszawa, 2012, p. 780

[15] J. Korczak, M. Hernes, M. Bac, "Performance evaluation of decision-making agents' in the multi-agent system", in: *Proceedings of Federated Conference Computer Science and Information Systems* (FedCSIS), Warszawa, 2014.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1289–1296
DOI: 10.15439/2016F236
ACSIS, Vol. 8. ISSN 2300-5963

The Role of Polish Crowdfunding Platforms in Film Productions – an Exploratory Study

Urszula Świerczyńska-Kaczor
The Jan Kochanowski University in Kielce
ul. Żeromskiego 15,
25-369 Kielce, Poland
Email: swierczynska@ujk.edu.pl

Paweł Kossecki
Paweł Kossecki
The Lodz Film School,
ul. Targowa 61/63,
90-323 Lodz, Poland,
E-mail: kossecki@poczta.onet.pl

***Abstract*—This paper aims to contribute to the better understanding, and in consequence, better development and implementation of crowdfunding projects for filmmaking. This study covers two areas of analysis: project-level research and founders-level (creators-level) research. In the first area, the article presents an analysis of documentary film projects based on descriptive statistics and the clustering of the film projects. In the second area, the paper sheds light on the Polish filmmakers' attitude to crowdfunding, and this analysis was based on a survey. This exploratory study led to the following conclusions: 1. The descriptive statistics indicate that the 'average documentary film' on the Polish market reaches a higher level of funding compared to non-documentary projects, and higher numbers of supporters. 2. The film projects on the Polish crowdfunding platforms can be segmented into six clusters. 3. The survey conducted among Polish filmmakers indicates that the experts strongly differed in their views about crowdfunding in general, and specifically, in the crowdfunding for documentary films. Such diversity of opinions and attitudes may be linked to the novelty of crowdfunding and therefore, the experts' difficulty of assessing the present and future role of crowdfunding for filmmaking. 4. This study shows that over 85% of experts agree with the sponsor's involvement with film production. Such a high level of expert agreement is important as nowadays 'being a prosumer' is one of the major trends of consumer behavior.**

I. Introduction

THE magnitude of crowdfunding has been steadily growing: in 2014 the global total volume of funding was about 16.2 billion US$ compared to 0.8 billion US$ in 2010 (Belleflamme, Omrani, & Peitz, 2015), the European Commission recorded the growth of crowdfunding platforms from 445 million euro in 2011 to 1 billion euro in 2013 (Borello, De Crescenzo, & Pichler, 2015), and World Bank predicts that crowdfunding will reach 93 billion US$ by 2025 (Kshetri, 2015). Although, crowdfunding covers very different kinds of projects, the cultural industries, such as film, music and video games have been taking a large share of the crowdfunding market (Boeuf, Darveau, Legoux, 2014). Major crowdfunding platforms gather thousands of artistic projects, e.g. on 1 May 2016, on Kickstarter there were over 294 thousand projects in all artistic categories, nearly 55 thousand projects in the category of video/film, and within this category over 600 'live' projects (kickstarter.com, 01.05.206).

In Poland crowdfunding is still in the very early stages of introduction to possible adopters, both to funders (called also sponsors or backers) and founders (project creators). So far, Polish crowdfunding platforms have attracted a relatively small group of users, for example, on 2 May, 2016 the crowdfunding platform polakpotrafi.pl reported 2263 projects in all categories (polakpotrafi.pl, 02.05.2016). However, we presume that the near future would bring further expansion of crowdfunding on the Polish market due to various factors, for instance, diminishing barriers for Internet payment and growing user knowledge about crowdfunding.

The academic research of crowdfunding is still in the nascent stage, especially if we consider the research conducted in a part of a particular domain, e.g. the crowdfunding for disruptive innovations, or the crowdfunding for art and cultural events. This article aims to fill the gap in the understanding of crowdfunding in the domain of film production. In this study we particularly focus on crowdfunding projects linked to documentary films.

The paper is structured as follows: in the next section we discuss the research problem and the scope of the conducted analyses, and then we present the context of research. The third section focuses on the analysis of the documentary films on selected crowdfunding platforms (Kickstarter, polakpotrafi.pl, and wispieramkulture.pl). This overview of the film projects is followed by a presentation of the results of a survey conducted among Polish filmmakers. Finally, the last part of the article highlights the conclusions, and future research.

I. THE RESEARCH PROBLEM

A. The scope of research

Our research aims to contribute to better understanding the process of film crowdfunding, with particular focus on documentary films and the Polish filmmaking industry. In this article, we present a study conducted in the following two areas:

A. the overview of the film and documentary film projects on selected crowdfunding platforms;
B. the Polish filmmakers' attitudes to crowdfunding;

B. Overview of film and documentary film projects on the selected crowdfunding platform

The idea of crowdfunding began from the crowdfunding of small-scale music and film projects (Hörisch,2015). Nowadays, it seems that crowdfunding covers almost all kinds of human activities (from art, technology, medicine, scientific discovery to film production), and crowdfunding projects significantly vary in time duration (from 'one-time' events to projects enabling the expansion of business ventures). In literature, crowdfunding is classified into four different models (see - Ryu & Kim, 2016; Lam, & Law, 2016), although in practice crowdfunding can be based on their variations (see – Vasileiadou, Huijben, & Raven, 2015). The main crowdfunding models are: donation-based crowdfunding, lending-based crowdfunding, reward-based crowdfunding, and equity-based crowdfunding.

The majority of the film projects are based on a reward-based crowdfunding in which the funders receive different kinds of rewards as compensation for their support. In some situations, funders make a very small donation (for example, 5 zł) without even expecting a reward at all (donation-based crowdfunding). In most projects, the creator of the project (the founder) offers the film (DVD or online viewing) as a reward, therefore film crowdfunding can be perceived as pre-ordering the product by the potential audience. The creator often encourages the crowdfunders to participate and co-create the final product (e.g. playing parts in the film), and therefore the crowdfunder acts as a prosumer.

Apart from financing the project, crowdfunding can play other roles, for example being a way of promotion or a tool of validating the potential market for ideas (Hörisch, 2015). These two roles can be noticed also in film projects, as crowdfunding can bring public attention to a particular movie production (promotion role), and – if the project fails in funding – it may be a signal that the idea was not properly developed (the validation).

The knowledge of the project creator in how to develop the project, (for example, what level of financing is feasible, how to reward the backers), is crucial for enhancing the project's probability to meet the financial goal (the requested financial support). On many reward-based crowdfunding platforms (e.g. Kickstarter.com or polakpotrafi.pl) if the project does not meet the threshold of funding, the money returns to the backers, therefore reaching the financial threshold for each project determines the project's actual execution. Any crowdfounder must consider that many factors may constitute the success or failure of crowdfunded projects, for example: the entrepreneur's social capital (see – Zheng et al., 2014), the entrepreneur's network of close contacts (Mendes-Da-Silva et. al, 2016), the way in which the project demonstrates its legitimacy (Frydrych et al., 2014), or the level of required funding, the time of project duration, and its contribution frequency from sponsors (Cordova, Dolci, & Gianfrate, 2015).

This study focuses on the analysis of crowdfunding for documentary film projects on the Polish market. To what extent does the funding of these film projects vary? Do most projects aim for very high or low funding? What is the 'average' level of funding, or number of sponsors?

To analyze crowdfunding for filmmaking we gathered the data from two selected Polish reward-based crowdfunding platforms – polakpotrafi.pl and wspieramkulutre.pl. The analyses were based on the descriptive statistics and K-Means analysis aiming to identify the clusters of film projects. To better understand the features of Polish crowdfunding compared to global platforms, we also conducted an analysis of selected documentary film projects on Kickstarter.

C. The filmmakers' attitudes to crowdfunding

The future of crowdfunding for films is, and will be, affected by the filmmakers' knowledge and attitude to the crowdfunding phenomenon. How do the professional filmmakers perceive the role of crowdfunding? What kind of barriers do they perceive for film crowdfunding? Do they agree with sponsor participation and film co-creation as a possible reward? In what way do experts assess the probability of success for documentary films? We looked for the answers by conducting a survey among the professionals. In April 2016, 37 respondents - Polish film producers and students of film production – took part in the survey, and filled in a questionnaire. As film production students are more likely to seek new funding sources compared to well-known producers, the opinion of this group is especially interesting and can be an indicator of crowdfunding development in the near future.

II. THE CONTEXT OF RESEARCH

In this section of the article, we would like to point to the following aspects of crowdfunding: the type of crowdfunding platforms, the sponsors, and developing the reward options.

A. Type of crowdfunding platforms

To date, equity-based crowdfunding (in which the founders agree on sharing the profit) is much less popular and takes a much smaller part of the crowdfunding market than reward-based crowdfunding, however, its popularity may grow in future (Son Turan, 2015). With reference to filmmaking, we would like to point to the following two examples:

- British film project - The Age of Stupid - in which the profits were pledged to crowdfunders (Belleflamme, Lambert, & Schwienbacher, 2014).
- CinemaShares.com – the crowdfunding platform which allows movie fans to buy shares of film projects (see - https://cinemashares.com, 02.05.2016).

B. The sponsors

Drivers and deterrents of crowdfunding differ from the perspective of the creator, funder (sponsor) and crowdfunding platforms (see – Kuti, & Madarász, 2014/3). So far, there is little knowledge about the specific sponsors' needs, behavior, motivation or characteristics in crowdfunding for filmmaking.

The study of Ryu, & Kim (2016) is helpful in the understanding of the film crowdfunders' characteristics, although the study covers different kinds of projects (not only films). The authors identified four types of crowdfunding sponsors, which are named descriptively as 'angelic backers', 'reward hunters', 'avid fans' and 'tasteful hermits'. These identified clusters of sponsors differ not only in their characteristics (for example, the value of philanthropic and reward motivation), but also they tend to pledge money at different stages of the project life-cycle: from project launching to project closing. Ryu, & Kim (2016) also pointed out that the project genre determines what kind of sponsor group is attracted, and the authors stated that "[a]ngelic backers tended to support films, plays, and charity projects, while reward hunters were more focused on art and design and game projects" (p. 49). The study of Galuszka, & Bystrov (2014) which refers to the music industry, showed that the sponsor's motivation is not only connected with financial reward, but also to the willingness to support their favorite artists or the fandom.

Another important aspect is the viewer's satisfaction. In this field, the research of Xu et al. (2016) can bring an interesting insight, although again, the study embraces different types of projects (not necessarily films). Xu et al. (2016) examined the sponsor's satisfaction, and the ascendant factors from an asymmetrical perspective and configurational models. The authors pointed to the role of project implementation perspective, project novelty, sponsor participation, entrepreneur activeness and sponsor demographics as the important variables influencing sponsor satisfaction.

C. Developing the rewards options

Thürridl, & Kamleitner (2016) examined different types of rewards in the crowdfunding (in general, not only film), indicating their important features, such as purpose/reward type, tangibility, scarcity, geographical location, monetary value/reward tier, recognition, level of collaboration, and the core future. Their study showed that the most popular strategies of rewarding sponsors in the film category are 'Add-On Highly Appreciated' and 'Top it Up'. In 'Top it Up' the rewards accumulate when a sponsor selects more valuable options, for example: for selecting the first tier the sponsor receives a 'thank you e-mail', but the second tier includes the reward from the first option, and additional benefits e.g. DVD film copy. 'Add-On Highly Appreciated' strategy underlines the value of the recognition of sponsor contribution to the project, for example, the filmmaker puts the name of the sponsor in the film credits.

In many cases of film crowdfunding, the crowdfunders pre-order the final product as a reward, which means that they take the risk that the film (final product) would not meet their expectations or the film would not be produced at all. Although fraud is possible in crowdfunding, the study conducted by Mollick (2014) indicates that very few projects failed to deliver the promised products, although delivery was often late. Some crowdfunding platforms, e.g. the crowdfunding platform Seed&Spark, offer the possibility to see the accepted movies on its website (https://www.seedandspark.com/, 03.05.2016), which may be a factor in enhancing the sponsor's trust to the founder.

III. THE OVERVIEW OF THE DOCUMENTARY FILMS ON CROWDFUNDING PLATFORMS

A. Global crowdfunding platform – kickstarter.com

Kickstarter is one of the major reward-based crowdfunding platforms, at the time of writing, for creators from the US, the UK, Canada, Australia, New Zealand, the Netherlands, Denmark, Ireland, Norway, Sweden, Germany, France, Spain, Italy, Austria, Belgium, Switzerland, and Luxembourg. In order to overview the characteristics of the documentary film projects on kickstarter.com, we looked into data from two samples:

a. Sample 1 - On 13 April 2016 the search engine on kickstarter.com showed 14 920 projects classified as "documentary projects in film/video on Earth'. For further analysis, we selected 100 projects using, as the criterion, 'the end day' - the first listed project will end in 2 hours, the last project will end in 19 days. In the next step, from this list we selected 19 projects which had reached the level of successful funding at the moment of sampling (note – 1. as some of these 19 successful projects still had a few more

days for running, the actual funding for these projects may be higher than we captured in our sample; 2. other projects from our sample list may well reach the level of funding in the next few days, but as they were not successful at the moment of sampling we did not include them for further analysis)

b. Sample 2 - On 1 May, 2016, the search engine of the Kickstarter.com platform showed 15 014 projects categorized as 'documentary projects in film/video on Earth'. In the next step, we entered in search the criterion of 'the most funded', and we sampled the 10 top projects.

The most funded documentary film project (sample 2) gathered over 1 mil US$ with support from over 8 500 backers (see Tab. I). The average funding for the top ten projects was over 476 thousand dollars with support from over 7 thousand backers. The top ten projects raised over 4.7 mil US$ altogether.

The 19 successful documentary projects analyzed from sample 1 raised US$ 240 994 altogether, with the average project receiving over 16 thousand dollars due to attracting over 160 backers. The overfunding for the 19 projects was 52% on average, meaning that the average project received more than one and a half times the funding that the founders applied for. Moreover, the 'average project' from the 'top 10' was overfunded by more than two times.

B. Polish crowdfunding platforms – polakpotrafi.pl and wspieramkulture.pl

For the analysis of Polish crowdfunding, we looked into the projects listed on two platforms: polakpotrafi.pl and wspieramkulture.pl. We gathered the data about the successful projects categorized as a film (on wspieramkulture.pl) or video/film (on polakpotrafi.pl), which were listed on these platforms in April 2016. It led to the analysis of 29 successful projects on wspieramkulture.pl and 108 successful projects on polakpotrafi.pl. In the next step of analysis, we classified 41 (38%) projects of the 108 projects as documentary films on polakpotrafi.pl, and 6 projects (21%) of the 29 projects as documentary films on wspieramkulture.pl. As the documentary films could have different features, we adopted a 'soft' approach for classification - taking as the main criterion to what extent the film is based on facts.

Our analysis was focused on two groups of projects in the film category: the documentary film vs non-documentary projects (for example: music video, fiction film). Tab. II presents the general overview of the documentary projects compared to the non-documentary projects on both Polish platforms. Average documentary film projects received funding of over 14 thousand złotych on the polakpotrafi platform, and nearly 16 thousand złotych on the wspieramkulure platform. The average documentary film project gathered the about 130 supporters. The statistics indicate a high level of diversification of projects, and a higher level of support for the documentary film projects compared to other film projects.

In order to identify the clusters for analyzed projects listed on polakpotrafi.pl, we used the following variables in the K-Means analysis:

a. The type of film project: documentary or non-documentary;

b. The level of overfunding the project received. The projects vary on the level to which the actual funding exceeded the financial threshold. We categorized the film projects into three categories based on the ratio 'the actual

TABLE I.
THE DESCRIPTIVE STATISTICS OF ANALYZED PROJECTS - KICKSTARTER.COM

	19 successful projects from the 100 projects taken from 'end day' search				**Top 10 documentary projects from the search 'most funded'**			
	Average	**Minimum**	**Max**	**Standard deviation**	**Average**	**Minimum**	**Max**	**Standard deviation**
Actual funding	$16 196.21	$555	$16 3517	$37 885	$476 224	$239 020	$1 126 036	$303 684
Financial goal (financial threshold)	$12 683.89	$250	$14 3436	$33 206.85	$303 543	$60 000	$650 000	$210 368
Ratio: actual funding/financial goal	1.52	1.04	2.6	0.52	2.02	1.10	5.4	1.3
Number of backers	164	14	1 685	384	7 074	2 621	16 850	4 169
Funding per a backer	$89.36	$18.28	$296	$58.94	$72.59	$28.33	$129.61	$31.68
Number of new backers	122	2	1 455	327.78	3 669	1 254	6 853	1 804.63
Percentage of new backers	66%	10%	90%	18%	57%	29%	87%	21%

Note: the currency of projects varied – there were British pounds, Australian dollars and American dollars. The data in the table is based on calculation of all the projects' funding in American dollars

funding/aimed financial goal for project'. Therefore, we identified three groups of projects: 'just reach' – ratio 100-119%, 'medium overfunding' – ratio 120-149%; 'high overfunding' – 150-199%, and 'exceptional' – ratio above 200%;

c. The numbers of sponsors – the project with 'low numbers of sponsors': 1 to 49; 'medium number of sponsors' – from 50 to 99 supporters; 'high numbers of sponsors' – from 100-199 supporters; 'very high numbers of sponsors' – above 200 supporters.

The K-Means analysis indicated six clusters for projects (Tab. III): three clusters for documentary films and three clusters for non-documentary projects. The three clusters of documentary projects are as follows:

- The largest cluster is the group of documentary films which 'just reach' the financial goal with few or relatively few supporters;
- The second cluster of documentary films is the group of projects which 'just reach' the financial goal, but these projects managed to engage high numbers of supporters;
- The third cluster is the group of documentary films which exceeded the level of financing on a medium level and attracted high numbers of supporters;

IV. Polish filmmakers' views on crowdfunding

A. The methodology of conducting the survey

In the next phase of research, we asked professionals – filmmakers - about their attitude to crowdfunding. A group of 37 professionals (20 women, 17 men) - producers and the students of film production - took part in the survey which aimed to gather information about the barriers for crowdfunding film projects, the preferable structure of the rewards, and future trends on the market. We included the group of students as our respondents due to various reasons: 1. they are likely to seek funding outside the usual procedures, 2. they are likely to implement the Internet in their work, 3. their attitude will impact the development of the film industry as they soon enter (or they have just entered) the film market. In our research we used a

TABLE II.
THE STATISTICS OF ANALYZED PROJECTS – POLAKPOTRAFI.PL AND WSPIERAMKULTURE.PL

	Successful projects on polakpotrafi in film/video category			**Successful project on wspieramkulture.pl in film category**		
	All successful projects in film category N=108	**The documentary films n (documentary) =41**	**Non-documentary projects n (non-documentary) = 67**	**All projects N=29**	**The documentary films n=6**	**Non-documentary projects n=23**
Average actual funding per a project	zł 10 011	zł 14 392	zł 7 329	zł 8 537	zł 15 905	zł 6 615
Average – requested funding (threshold)	zł 7 737	zł 10 849	zł 5 832	-	-	-
Minimum actual funding	zł 601	zł 1 401	zł 601	zł 510	zł 850	zł 510
Maximum actual funding	zł 58 054	zł 58 054	zł 48 365	zł 36 995	zł 36 995	zł 25 205
Average - number of sponsors per a project	96	133	73	57	127	38
Average funding per sponsor	zł 112	zł 114	zł 110	For projects without 'partner' (n=21)		
				zł 168	zł 269	zł 137

TABLE III.
CLUSTERS OF FILM PROJECTS – POLAKPOTARFI.PL

	Type of film project	**Number of sponsors**	**The level of funding**	**Number of cases**	**Percentage (%)**
1	Documentary	LOW	JUST REACH	26	24.07
2	Documentary	HIGH	JUST REACH	7	6.48
3	Documentary	HIGH	MEDIUM OVERFUNDING	11	10.19
3	Non-documetnary	LOW	JUST REACH	39	36.11
5	Non-documentary	HIGH	HIGH OVERFUNDING	8	7.40
6	Non-documentary	MEDIUM	JUST REACH	17	15.74

questionnaire with closed and open questions. The survey was conducted in April, 2016.

B. The barriers for film crowdfunding

The filmmakers' opinions about the barriers were highly diversified. The experts perceived and assessed factors which can hinder the development of crowdfunding differently (see Tab IV). To sum up the filmmakers' responses, we point to the following conclusions:

a. An important barrier in using crowdfunding for film projects lies in the sponsors' uncertainty of the project execution. Over 80% of experts agreed with the statement that the important barrier for financing film is the sponsors' uncertainty in whether the money would be spent in a reasonable way. Also about 70% of experts shared the view that sponsors can be uncertain of whether the film would be produced at all.

b. There is no major crowdfunding platform developed exclusively for filmmakers on Polish market (almost 60% of experts pointed to this barrier).

c. Lack of knowledge about crowdfunding. Half of the group of experts expressed the view that Polish filmmakers lack detailed knowledge about the terms and rules of crowdfunding. Over 60% of experts claimed that Polish filmmakers do not 'trust' this way of financing film projects.

d. Uncertainty about the legal aspects of crowdfunding. Almost 60% of experts shared the opinion that the Polish legal regulations for crowdfunding can be perceived as 'vague'.

e. The majority of experts do not perceive that the barriers lie in: 1) the lack of a proper schema of rewards, 2) the high cost of seeking financial support using crowdfunding, or 3) sponsors' concerns about online privacy.

TABLE IV.
EXPERTS' VIEWS OF BARRIERS FOR FILM CROWDFUNDING ON POLISH MARKET

The stated barrier	To what extent the experts agree with the statement (1-strongly disagree, 4 -strongly agree)	The percentage of experts which agree with the statement (4 or 5 points)
The sponsor's uncertainty that the money would be used in a reasonable way for film production	3.1	81%
The sponsor's uncertainty that the film would be produced at all	3.1	68%
There is no major crowdfunding platform developed for filmmakers on Polish market	2.8	59%
Lack of filmmakers' trust in crowdfunding	2.7	62%
Vague legal regulations for crowdfunding	2.8	59%
Lack of filmmakers' knowledge about crowdfunding	2.7	51%
The prospective sponsors' lack knowledge about crowdfunding	2.6	49%
High costs of crowdfunding for film projects	2.3	41%
Sponsors' concerns about their privacy and data gathered by crowdfunding platforms	2.2	41%
Too low reward schema for sponsors offered by Polish filmmakers	2.1	30%

TABLE V.
THE FILMMAKERS' ASSESSMENT OF THE REWARD OPTIONS

The reward	To what extent the reward is suitable for Polish film market (1 – not suitable at all; 5 – highly suitable)	The percentage of experts perceiving the reward schema as suitable for the Polish film market
Involving backers in events connected with the film, e.g. film premiere, thank you note on social media website	3.6	92%
The copy of the film on DVD as a reward	3.6	89%
The opportunity for a sponsor to be engaged in film production e.g. playing small parts in a film	3.4	86%
Film merchandise as rewards, e.g. T-shirts, mugs	3.0	65%
Sharing the profit earned by a film project with backers	2.1	32%

C. Developing a reward schema

In the next part of the survey experts were asked how they perceived different forms of rewarding backers (see Tab. V). The results are:

a. Majority of respondents perceived the following rewards as suitable for film production: involving backers in film events, offering a copy of the film DVD and engaging the supporters directly in the film production process, e.g. playing small parts in the film. The offering film merchandise was perceived as suitable by 65% of experts.

b. The experts do not agree with the idea of sharing the film profit with backers. Only one third of the experts consider the reward schema based on sharing the film profit with sponsors as a suitable solution on the Polish market.

D. The experts' views about their involvement in crowdfunding and future trends

We asked experts the question based on the following scenario: "Imagine that today you are planning the process of film production. To what extent do you consider crowdfunding as a way of funding for your film project?".

a. Over 90% of experts – 34 of the 37 respondents – stated that they consider crowdfunding as a supplementary way of funding the film project.

b. Three respondents stated that they would not consider crowdfunding at all.

c. None of respondents aimed to fund the film project mainly with crowdfunding.

The filmmakers also perceived the future role of crowdfunding differently. The group of experts was split almost in half in their assessment.

a. 51% of respondents agreed with the opinion that during the next five years the importance of crowdfunding for film projects would be growing significantly;

b. 49% of respondents expressed the opinion that during the next five years crowdfunding still remains only an 'additional' – not significant – way of funding film projects.

The experts also estimated the probability of success of the documentary film project differently.

a. Almost half of respondents (46%) stated that documentary films would probably be less successful than other film projects.

b. 24% of experts estimated the chances of success for documentary films as being similar to other film projects.

c. Almost 30% of experts perceived the documentary films as being probably more successful than other film projects.

V. Conclusions

This exploratory study, based on the analysis of selected film projects on crowdfunding platforms and the results of a survey conducted among Polish filmmakers, led to the following conclusions:

A. The statistical data indicates that the 'average documentary film' on the Polish market reaches a higher level of funding compared to non-documentary projects, and higher numbers of supporters.

B. The analysis indicates six different clusters of film projects on Polish crowdfunding platforms. Three clusters include documentary film projects:

- first cluster: documentary films 'just reaching' the threshold of funding, with relatively few sponsors;
- second cluster: documentary films 'just reaching' the threshold with a high number of sponsors;
- third cluster: documentary films reaching 'medium level of overfunding' with a high number of sponsors.

Although overfunding may seem be the indicator of film success and, at first glance, has only a positive effect on movie projects, the overfunding of film projects is also linked with possible problems connected with the growing scale of the project (for example, it can change the scale of producing and delivering the film merchandise)

C. The filmmakers taking part in the survey strongly differed in their views about crowdfunding in general, and specifically, the crowdfunding for documentary films. Such diversity of opinions and attitudes may be linked to the novelty of crowdfunding and therefore, the experts' difficulty of assessing the present and future state of crowdfunding.

D. As the majority of film projects are reward-based crowdfunding, it is important for project creators to develop properly working reward options. This study shows that over 85% of experts agree with the sponsor's involvement with film production. Such high level of expert agreement for sponsor-filmmaker cooperation is important, as one of the major trends of consumer behavior is based on the 'being a prosumer' attitude.

This study is as an exploratory study; therefore, the obtained result is difficult to generalize. However, this exploratory study gives the background for future research, particularly in such areas as:

- building trust between sponsors and founders in crowdfunding for filmmaking;
- the promotional role of crowdfunding and its incorporation in the promotion of a film project;
- the role of crowdfunding in marketing research for the film industry.

References

[1] P. Belleflamme, T. Lambert, and A. Schwienbacher (2014), "Crowdfunding: Tapping the right crowd", *Journal of Business Venturing* 29 (2014), 585–609, http://dx.doi.org/10.1016/j.jbusvent.2013.07.003

[2] P. Belleflamme, N. Omrani, and M. Peitz (2015), "The economics of crowdfunding platforms", *Information Economics and Policy* 33 (2015) 11–28, http://dx.doi.org/10.1016/j.infoecopol.2015.08.003

[3] B. Boeuf, J. Darveau, and R. Legoux (2014), "Financing Creativity: Crowdfunding as a New Approach for Theatre Projects", *International*

Journal Of Arts Management, Volume 16, Number 3, Spring 2014, 33-48

[4] G. Borello, V. De Crescenzo, and F. Pichler (2015), "The Funding Gap and The Role of Financial Return Crowdfunding: Some Evidence From European Platforms", Journal of Internet Banking and Commerce, JIBC April 2015, Vol. 20, No. 1, 1-20

[5] A. Cordova, J. Dolci, and G. Gianfrate (2015), "The determinants of crowdfunding success: evidence from technology projects", Procedia —Social and Behavioral Sciences 181 (2015) 115–124, doi: 10.1016/j.sbspro.2015.04.872

[6] D. Frydrych, A. J. Bock, T. Kinder, and B. Koeck (2014), "Exploring entrepreneurial legitimacy in reward-based crowdfunding", Venture Capital, 2014, Vol. 16, No. 3, 247–269, http://dx.doi.org/10.1080/13691066.2014.916512

[7] P. Galuszka, and V. Bystrov (2014), "Crowdfunding: A Case Study of a New Model of Financing Music Production", Journal of Internet Commerce, 13:233–252, 2014, DOI:10.1080/15332861.2014.961349

[8] J. Hörisch (2015), "Crowdfunding for environmental ventures: an empirical analysis of the influence of environmental orientation on the success of crowdfunding initiatives", Journal of Cleaner Production 107 (2015) 636-645, http://dx.doi.org/10.1016/j.jclepro.2015.05.046

[9] N. Kshetri (2015), "Success of Crowd-based Online Technology in Fundraising: An Institutional Perspective", Journal of International Management 21 (2015) 100–116, http://dx.doi.org/10.1016/j.intman.2015.03.004

[10] M. Kuti, and G. Madarász (2014/3), "Crowdfunding", Public Finance Quarterly, 2014/3, 355-366

[11] P. T. I. Lam, and A. O. K. Law (2016), "Crowdfunding for renewable and sustainable energy projects: An exploratory case study approach", Renewable and Sustainable Energy Reviews 60 (2016) 11–20, http://dx.doi.org/10.1016/j.rser.2016.01.046

[12] W. Mendes-Da-Silva, L. Rossoni, B. S. Conte, C. C. Gattaz, and E. R. Francisco (2016), "The impacts of fundraising periods and geographic distance on financing music production via crowdfunding in Brazil", Journal of Cultural Economics (2016) 40:75–99, DOI: 10.1007/s10824-015-9248-3

[13] E. Mollick (2014), "The dynamics of crowdfunding: An exploratory study", Journal of Business Venturing 29 (2014) 1–16, http://dx.doi.org/10.1016/j.jbusvent.2013.06.005

[14] S. Ryu, and Y.-G. Kim (2016), "A typology of crowdfunding sponsors: Birds of a feather flock together?", Electronic Commerce Research and Applications 16 (2016) 43–54, http://dx.doi.org/10.1016/j.elerap.2016.01.006

[15] S. Son Turan (2015), "Financial Innovation - Crowdfunding: Friend or Foe?", Procedia - Social and Behavioral Sciences 195 (2015) 353–362, doi: 10.1016/j.sbspro.2015.06.334

[16] C. Thürridl, and B. Kamleitner (2016), "What Goes Around Comes Around? Rewards As Strategic Assets In Crowdfunding", California Management Review Vol. 58, No. 2 Winter 2016, 88-110

[17] E. Vasileiadou, J.C.C.M Huijben, and R.P.J.M. Raven (2015), "Three is a crowd? Exploring the potential of crowdfunding for renewable energy in the Netherlands", Journal of Cleaner Production xxx (2015) 1-14, http://dx.doi.org/10.1016/j.jclepro.2015.06.028

[18] B. Xu, H. Zheng, Y. Xu, and T. Wang (2016), "Configurational paths to sponsor satisfaction in crowdfunding", Journal of Business Research 69 (2016) 915–927, http://dx.doi.org/10.1016/j.jbusres.2015.06.040

[19] H. Zheng, D Li., J. Wu, and Y. Xu (2014), "The role of multidimensional social capital in crowdfunding: A comparative study in China and US", Information & Management 51 (2014) 488–496, http://dx.doi.org/10.1016/j.im.2014.03.003.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1297–1302
DOI: 10.15439/2016F478
ACSIS, Vol. 8. ISSN 2300-5963

SIMMI 4.0 – A Maturity Model for Classifying the Enterprise-wide IT and Software Landscape Focusing on Industry 4.0

Christian Leyh
Technische Universität Dresden
Chair of Information Systems, esp. IS in Manufacturing and Commerce
Helmholtzstr. 10, 01069 Dresden, Germany
Email: Christian.Leyh@tu-dresden.de

Thomas Schäffer
University of Applied Sciences Heilbronn
Faculty of Business Administration
Max-Planck-Str. 39, 74081 Heilbronn, Germany
Email: Thomas.Schaeffer@hs-heilbronn.de

Katja Bley
Technische Universität Dresden
Chair of Information Systems, esp. IS in Manufacturing and Commerce
Helmholtzstr. 10, 01069 Dresden, Germany
Email: Katja.Bley@tu-dresden.de

Sven Forstenhäusler
University of Applied Sciences Heilbronn
Faculty of Business Administration
Max-Planck-Str. 39, 74081 Heilbronn, Germany
Email: sforsten@stud.hs-heilbronn.de

***Abstract* — The increasing digitalization of business and society leads to drastic changes within companies. Nearly all enterprises have to face enormous challenges when dealing with topics such as Industry 4.0/Industrial Internet. One of these challenges represents the realistic classification of the company's own IT infrastructure. In this paper we present a maturity model (SIMMI 4.0 – System Integration Maturity Model Industry 4.0) that enables a company to classify its IT system landscape with focus on Industry 4.0 requirements. SIMMI 4.0 consists of 5 stages. Each describes several characteristics of digitization, which allows a company to assess itself. Additionally, recommended activities are presented for each stage of digitization, which can enable a company to reach the next stage of maturity. We also present several possible topics for future research to improve and refine the developed maturity model.**

I. Motivation

ONE of the most important challenges that companies currently face is the digitization of business processes and of the enterprise itself. They have to join in global digital networking, improve automation of individual or even all business processes, and reengineer existing business models to gain momentum in digital innovation. It has never been more important for enterprises to be able to rely on IT-enabled capabilities, as well as to count on a deep understanding of information technology in general and in digital innovation in particular. Without a doubt, nearly all enterprises have to undergo an increasing digital transformation to remain competitive in global markets. In these efforts, the specific challenge for companies is to realize the increasing integration of virtual, digital programs with real objects or products in their everyday business in order to subsequently adapt, enhance, or optimize the processes [1]–[4].

For a while, trends such as Industry 4.0/Industrial Internet affected mainly large companies, especially since small- and medium-sized enterprises (SMEs) often judged such topics as too complex and expensive and partially classified them as not relevant. However, digitization is no longer limited to large companies and does not only concern separate functional areas such as the IT department. Rather, it takes place throughout the entire value chain of all companies [5]. Overall, together with this increasing digitalization of companies, the definitions of value-adding and supportive processes become vague, whereby the traditional supply chain of a company with its downstream processes develops into a holistic supply/value network. Thus, also SMEs open themselves for the complex topic of Industry 4.0 and try to reshape their business processes and business models in this direction. To face up to this development the use of adequate information and communication technology (ICT) is essential. However, what is missing at this point is the companies' level of knowledge concerning their own digitization. A number of studies already exist applying to this topic (e.g., [6], [7]). By using various interrogation techniques, the authors figure out which information and enterprise systems are used in business (especially in SMEs) and in what shape the IT-infrastructure of the company appears. There is, however, the question of how an IT landscape must be designed so that a company can "move" in the field of Industry 4.0. Recognizing and evaluating what systems are needed, and in which way and for what purpose, still embodies a challenge for companies.

This is where the present paper comes and this results in the main research question for our research:

What should a maturity model look like to assess a company's IT system landscape in the context of Industry 4.0?

Industry 4.0, as the fourth stage of the industrial revolution is entitled, consists of an increasing digitization of products and systems, together with their interconnectedness. Thereby, the physical world is connected to the virtual world. The characteristics of Industry 4.0 are: e.g., horizontal integration across whole value networks, strong vertical integration within the company, and a digital

transparency of the engineering across the entire value chain [8], [9]. However, a universal definition for the term "Industry 4.0" does not exist. From the aforementioned descriptions and further characteristics of Industry 4.0 we deduce a working definition as the foundation for our research:

Industry 4.0 describes the transition from centralized production towards one that is very flexible and self-controlled. Within this production the products and all affected systems, as well as all process steps of the engineering, are digitized and interconnected to share and pass information and to distribute this along the vertical and the horizontal value chains, and even beyond that in extensive value networks.

To answer the research question we present a tool (a maturity model) that enables companies to classify their own provided IT system landscape in the needs of an Industry 4.0 system landscape. This is also the core of our paper. We describe the components (dimensions and stages) of "SIMMI 4.0" (System Integration Maturity Model Industry 4.0) necessary to fulfill the requirements of an Industry 4.0 environment. Afterwards we finish with a short summary and an outlook for future research in this field.
Furthermore, related work with detailed insight into the field of Industry 4.0, as well as in the field of existing maturity models and into the development of our maturity model "SIMMI 4.0" are given in [10].

II. SIMMI 4.0 – System Integration Maturity Model Industry 4.0

As a starting point for model development, a further literature analysis was conducted. Contrary to the related work literature analysis in [10], the aim of this analysis was to gain an understanding about the existing level of knowledge about Industry 4.0, and, therefore, to deduce the essential requirements for IT systems in the context of Industry 4.0. Several databases (e.g., EBSCO, ScienceDirect, SpringerLink, and Google Scholar) were searched using the following terms and combinations of these terms: Information systems, Industry 4.0, Maturity models, Integration, Digitization, Internet of things and services, Cyber-physical systems, Value networks, IT systems, Enterprise systems, and Business information systems. Some of the resulting requirements from this literature analysis are presented as follows.

A. Requirements for IT systems in the context of Industry 4.0

In their final report about Industry 4.0, [8] highlighted three key requirements fostered by Industry 4.0 and thus should be supported by the enterprise application system landscape:

Vertical integration along the hierarchical levels of a company: While the different enterprise systems support their own tasks very well, the data of the respective systems, such as Enterprise Resource Planning (ERP) systems, Supply Chain Management (SCM) systems, Management Information Systems (MIS), Product Life cycle Management (PLM) systems, etc., is often stored in separate databases (sometimes data interfaces are provided) and partly stored in different formats. This sub-optimal level of integration must be improved for implementing Industry 4.0 business processes and activities.

Horizontal integration across value networks: For the implementation and use of different enterprise systems, failures and leakages throughout the flow of information must be avoided. In fact, the information must be accessible and useable at the right time in the right "place" along the entire supply chain and therefore for all business partners. Furthermore, the exchange of such information flows must be (completely) automatized.

Digital continuity of engineering: This means supporting a product's engineering consistently and continuously along the entire supply chain by using adequate and appropriate enterprise systems and includes the production system development process as well.

Also, stemming from the literature review (especially from analyzed study results), cross-sectional technologies were identifiable as an important part of the enterprise systems. These technologies are defined below and their relevance to Industry 4.0 will be explained:

Service-oriented architecture: For example, the project "Platform Industry 4.0" has published a whitepaper that names the development of a reference architecture based on a Service-oriented architecture (SOA) as an important prerequisite for the implementation of Industry 4.0 [11].

Cloud Computing: Industry 4.0 not only leads to a digitization of separate production facilities, but also that of the enterprise's information technology at the production plant(s) as well as all companies digitally interconnected along the supply chain. Considering cloud computing, these aspects are provided as different services; therefore, this could help enterprises operate in the field of Industry 4.0 effectively and efficiently.

Information aggregation and processing: In this context, aggregation of information implies that data can be easily identified from various integrated enterprise systems through different ways of treatment, such as clustering, filtering, and correlation. In a next step, this data is made available to every user or machine that needs it. This illustrates not only that the data of the production floor/of the production systems (e.g., various interconnected machines, (semi-) products, sensors etc.) is aggregated and transferred towards the company's higher levels and enterprise systems (e.g., ERP systems, SCM systems), but also that the data needs to be transferred in the opposite direction to the production floor [12].

IT security: In Industry 4.0, the company will be connected with/to the internet not just at an operational or higher level.

As part of the Internet of things and services, the production level/production floor, maybe even the control level of several machines themselves, as well as all levels up to the strategic level of companies will be connected through a continuous link to the internet. For this reason, IT security will be a major challenge for establishing different kinds of IT systems. Here, IT security is defined as adequate protection of all information available in form of electronic data. In addition, it must be ensured that the IT systems themselves and their services are available at all times for the users and work properly [13], [14].

B. Components of SIMMI 4.0

Depending on its aims and strategic positions as well as on its arrangements in terms of Industry 4.0, not every company needs to fully implement all the dimensions of SIMMI 4.0. There are several gradations per dimension, which in turn result in different stages within the maturity model. These dimensions can have different characteristics in terms of scope and intensity for each company. Therefore, Table I in the Appendix gives a summary of our proposal for SIMMI 4.0. In the following chapters, the dimensions and stages of SIMMI 4.0 are described in detail.

B.1 Dimensions of SIMMI 4.0

Several dimensions of the development of SIMMI 4.0 are deduced from the requirements from our literature analysis. With these dimensions, SIMMI 4.0 can enable a company to assess its IT system landscape.

Dimension – Vertical integration: This dimension focuses on the components of the lowest level of an enterprise, where different physical things ((semi-) products, machines, etc.) need to exchange information throughout the level itself and with the levels above. The most important criterion here is that this exchange is possible in both directions.

Dimension – Horizontal integration: Industry 4.0 requires horizontal integration across the different value networks. Accordingly, an essential criterion has emerged from the requirements above. An automated and integrated information flow is necessary along the horizontal enterprise level as well as beyond the enterprise borders. Without this information flow, a business-wide value network is not realizable, meaning that the various enterprise systems of the different partners in the value networks require interoperability at the data level. Therefore, a continuous and consistent information flow is needed [15], [16].

Dimension – Digital product development: For the engineering's digital continuity it is especially important that each process step is represented digitally. For this purpose, at least one enterprise system should be integrated into each respective process step. In addition, the resulting data and information of each step must be forwarded to the next and previous step/enterprise system.

Dimension – Cross-sectional technology criteria: This dimension focuses on assessing the extent to which technologies are used across all different fields of Industry 4.0. Based on the requirements, the respective fields are: Service-oriented architecture, Cloud computing, Big Data, and IT Security. In addition, the level of support that enterprise systems can provide for these fields should be evaluated in this dimension.

B.2 Stages of SIMMI 4.0

SIMMI 4.0 is divided into five stages. Additionally, key activities for each stage, which must be conducted in order to be able to achieve a higher stage, are briefly specified.
This five-stage division is justified by the fact that in the middle of this stage-model, in the third stage, the implementation of an intelligent factory (Smart factory) is completed. This foundation for Industry 4.0 should be and must be implemented in each company before stable, robust, and versatile value networks can be realized. By implementing an intelligent factory, a company can gain operating experience before the company and its systems are connected to other companies [15].

Stage 1 – Basic digitization level:
The company has not addressed Industry 4.0. Requirements are not or only partially met.
The enterprise systems along the enterprise's value chain support only their respective fields of activity. When integration is achieved, it is with specially implemented and complex interfaces. In addition, the processes are not or are only partially digitized. Product prototypes are designed in a costly way because of product development activities are not digitized. The company does not pursue service-oriented and cloud-based approaches.
The data of the enterprise systems are aggregated only for strategic decisions. In addition, the confidentiality of the data is not provided. The company's data is not protected against industrial espionage for example, incurring enormous damage annually. Anytime and continuous availability of data is not ensured. Sometimes, users cannot receive the data when they request it or access is not provided.

Activities:
- *Start of engagement with focus on Industry 4.0*
- *First explorations of service-oriented approaches*

Stage 2 – Cross-departmental digitization:
The company is actively engaged with Industry 4.0 topics. Digitization has been implemented across departments, and the first Industry 4.0 requirements have been implemented throughout the company.
Information can be (partially) exchanged automatically among different departments and business areas. This level of integration no longer contains data islands within the company. In addition, several production plants are connected but instead through cloud solutions they are

connected through the exchange of information in other ways (paper-based, email, FTP, etc.). Production and product development is supported by several enterprise systems. However, data and information exchange is not automatized. Therefore, the previous and following steps are not optimized. The company starts to implement an SOA. Legacy systems are broken down, and their functionalities are encapsulated into services. New systems are implemented directly following the SOA principles. Thus, initial processes can be built as services. In addition, an enterprise service bus (ESB) is implemented to replace enterprise application integration principles and to enable direct connection between new systems.

Activities:

- *Implementing an SOA*
- *Achieving cross-departmental integration*
- *First approaches for an IT security model*
- *First developments of mobile applications*

Stage 3 – Horizontal and vertical digitization:

The company is horizontally and vertically digitized. The requirements of Industry 4.0 have been implemented within the company, and information flows have been automated. The product development is consistently supported by enterprise systems. Information from the respective process steps can be forwarded to the next or previous process step.

The company has established an SOA. All the functionalities of the integrated systems are provided as services. The (semi-) products are part of this SOA and provide services themselves.

To exchange information within the enterprise, cloud principles are applied. Services are available company-wide and can be accessed anywhere. Employees are able to retrieve information everywhere through mobile devices. In addition, machines and (semi-) products are displayed on the mobile devices as soon as they come into the device's range. With this feature, the devices can display additional information about the machines (e.g., current processing step, maintenance status, etc.).

Various data from the production plants will be aggregated and processed together. Using this data and information gained from production, production itself can be optimized in real time and can be adapted to prevailing or changing conditions when necessary.

IT security is increased through the use of an advanced security model. Access to data is continuously protected, and data is transmitted in an encrypted state within the enterprise. The data's confidentiality, availability, and integrity are completely guaranteed.

Activities:

- *Connection with other companies to build value networks*
- *Development of a cloud-based platform to offer services across the company border*

Stage 4 – Full digitization:

The company has been completely digitized, even beyond corporate borders, and integrated into value networks. Industry 4.0 approaches are actively followed and anchored within the corporate strategy.

Consequently, the level of integration can be described as enterprise-wide and cross-corporate horizontal and vertical integration. In order to optimize processes, the product development steps automatically pass information to previous and following production steps.

The company has established a service-oriented and cloud-based platform that offers services in the value network in order to exchange information along the supply chain in real time. Machines can be maintained globally, regardless of their location (in terms of their software). Data is aggregated and processed company-wide as well as provided via entire value networks. The production floor in general is at a highly optimized level.

In addition to enterprise-wide data encryption, encryption is also used within the value networks. Users can access data anywhere by using established authentication measures.

Activities:

- *Beginning collaborations with companies within the value networks for end-to-end solutions and the optimization of information flows*

Stage 5 – Optimized full digitization:

The company is a showcase for Industry 4.0 activities. It collaborates strongly with its business partners and therefore optimizes its value networks. Through these collaborations, new business models and new end-to-end solutions are developed and enabled. During this development process each step inside and outside the company is digitized.

Within the value networks physical value and information flows can also be represented digitally, so the entire added value can be simulated in real time. Thus, it is possible to automatically perform necessary adjustments for all companies of the value network.

Furthermore, the IT security adjusts promptly to new risks. Occurring security problems are immediately solved. Encryption is optimized in cooperation with the partners the along the value networks.

III. Summary and Future Aspects

The aim of our research is to provide a maturity model for the classification of a company's IT system landscape in the context of the Industry 4.0 requirements. Through a systematic literature review [10], we could demonstrate that no maturity model currently exists that meets the needs of Industry 4.0 in terms of a company-wide and even a cross-corporate IT system landscape. However, due to the drastic changes produced by the digitalization of businesses and society itself, it becomes necessary for enterprises to assess their IT system landscape in a realistic way. Therefore, an

easy-to-handle tool could provide adequate support for assessment.

With this in mind, we designed a new maturity model (SIMMI 4.0 – System Integration Maturity Model Industry 4.0) for assessing the readiness of a company's IT system landscape in terms of Industry 4.0. However, this design process is not described in detail in this paper but can be found in [10].

Within this paper we present the first version of our maturity model SIMMI 4.0. Thus, the model's development is not yet fully complete. The next steps include: (1) conducting several expert interviews and model adjustments based on the interviews if necessary (2nd iteration); (2) group interviews with companies to test the model's practicability (3rd iteration). After these steps, evaluation of the maturity model will follow. This should be based on the concrete application of the model within several companies. The resulting design decisions based on the iteration steps, the transfer and evaluation in terms of the model's dimensions and stages, more detailed evaluation steps, and the model's scientific as well as practical contributions will be addressed in subsequent papers.

Beyond the development of SIMMI 4.0 (here primarily based on the literature review in [10], the comparison of existing maturity models), we identified additional links and needs for further research. For example, some maturity models already exist for the field of Industry 4.0 that deal with organizational aspects or system-specific aspects in detail. A mapping of these maturity models would be necessary to combine their different points of view. For example, different maturity level assignments and dimensions between these models should be developed to enable companies to fully classify themselves in terms of Industry 4.0 requirements in all levels of their enterprise. With this work, companies would be able to determine their overall maturity in the field of Industry 4.0.

A further aspect to investigate in the future is the data quality within various enterprise systems along the supply chain. Since companies in an Industry 4.0 environment must exchange data in large amounts and on an automated basis, a certain data quality is necessary to ensure efficient company-wide and cross-corporate business processes. Therefore, those companies should implement adequate master data management and data quality management. On this topic, two questions arise: (1) What design elements and components should be part of master data management and data quality management in the context of Industry 4.0? (2) How can master data management be integrated in maturity models addressing the IT systems landscape of Industry 4.0 companies? We will address those two questions in further research projects.

To conclude this contribution, some limitations must be recognized. Currently, SIMMI 4.0 has not been evaluated or tested. It is a maturity model that was derived from the literature by combining aspects of IT-related maturity models with Industry 4.0 requirements. In this respect, the development process of SIMMI 4.0 must continue. In the next iteration steps, we will clarify and review the model's components based on expert and company assessment. Additionally, SIMMI 4.0 must prove its practicability and usefulness in an enterprise environment. We will address both aspects of the model's limitations in our research project's future steps focusing the field of Industry 4.0.

IV. References

[1] M. Pagani, "Digital Business Strategy and Value Creation: Framing the Dynamic Cycle of Control Points," *MIS Q.*, vol. 37, no. 2, pp. 617–632, 2013.

[2] D. Straub and R. Watson, "Transformational Issues in Researching IS and Net-Enabled Organizations," *Inf. Syst. Res.*, vol. 12, no. 4, pp. 337–345, 2001. doi: 10.1287/isre.12.4.337.9706.

[3] B. Wheeler, "NEBIC: A Dynamic Capabilities Theory for Assessing Net-Enablement," *Inf. Syst. Res.*, vol. 13, no. 2, pp. 125–146, 2002. doi: 10.1287/isre.13.2.125.89.

[4] J. Schlick, P. Stephan, M. Loskyll, and D. Lappe, "Industrie 4.0 in der praktischen Anwendung," in *Industrie 4.0 in der Produktion, Automatisierung und Logistik*, T. Bauernhansl, M. ten Hompel, and B. Vogel-Heuser, Eds. Wiesbaden: Springer, 2014, pp. 56–84. doi: 10.1007/978-3-658-04682-8_3.

[5] O. A. El Sawy, A. Malhotra, YoungKi Park, and P. A. Pavlou, "Seeking the Configurations of Digital Ecodynamics: It Takes Three to Tango," *Inf. Syst. Res.*, vol. 21, no. 4, pp. 835–848, 2010. doi: 10.1287/isre.1100.0326.

[6] T. Schäffer and H. Beckmann, *Trendstudie Stammdatenqualität 2013: Erhebung der aktuellen Situation zur Stammdatenqualität in Unternehmen und daraus abgeleitete Trends. Steinbeis-Edition (Schriftenreihe Wirtschaftsinformatik)*. Stuttgart, 2014.

[7] C. Leyh, K. Bley, and T. Schäffer, "Digitization of German Enterprises in the Production Sector – Do they know how 'digitized' they are?," in Proc. of the 22nd *Americas Conferenence on Information Systems (AMCIS 2016)*, 2016.

[8] H. Kagermann, W. Wahlster, and H. Helbig, *Umsetzungsempfehlungen für das Zukunftsprojekt Industrie 4.0*. Abschlussbericht des Arbeitskreises Industrie 4.0, Frankfurt am Main, 2013.

[9] C. Lemke and W. Brenner, *Einführung in die Wirtschaftsinformatik: Band 1: Verstehen des digitalen Zeitalters*, 2015th ed. Heidelberg: Springer-Verlag, 2014. doi: 10.1007/978-3-662-44065-0.

[10] C. Leyh, T. Schäffer, and S. Forstenhäusler, "SIMMI 4.0 – Vorschlag eines Reifegradmodells zur Klassifikation der unternehmensweiten Anwendungssystemlandschaft mit Fokus Industrie 4 . 0," *Proc. zur Multikonferenz Wirtschaftsinformatik*, pp. 1651–1662, 2016.

[11] Industrie 4.0, *Industrie 4.0 – Whitepaper FuE-Themen*. Veröffentlichung der Plattform Industrie 4.0 in Zusammenarbeit mit dem Wissenschaftlichen Beirat, 2014.

[12] H. Schöning and M. Dorchain, "Data Mining und Analyse," in *Industrie 4.0 in Produktion, Automatisierung und Logistik*, T. Bauernhansl, M. ten Hompel, and B. Vogel-Heuser, Eds. Wiesbaden: Springer, 2014, pp. 543–554. doi: 10.1007/978-3-658-04682-8_27.

[13] M. Kappes, *Netzwerk-und Datensicherheit: Eine praktische Einführung*, 2nd ed. Wiesbaden: Springer-Vieweg, 2013. doi: 10.1007/978-3-8348-8612-5.

[14] H. Krcmar, *Einführung in das Informationsmanagement*, 2nd ed. Berlin, Heidelberg: Gabler, 2015. doi: 10.1007/978-3-662-44329-3.

[15] L. Forstner and M. Dümmler, "Integrierte Wertschöpfungsnetzwerke-Chancen und Potenziale durch Industrie 4.0," *e i Elektrotechnik und Informationstechnik*, vol. 131, no. 7, pp. 199–201, 2014. doi: 10.1007/s00502-014-0224-y.

[16] D. Wegener, "Industrie 4.0 – Chancen und Herausforderungen für einen Global Player," in *Industrie 4.0 in der Produktion, Automatisierung und Logistik*, T. Bauernhansl, M. ten Hompel, and B. Vogel-Heuser, Eds. Wiesbaden: Springer, 2014, pp. 343–358. doi: 10.1007/978-3-658-04682-8_17.

V. APPENDIX

TABLE I.
OVERVIEW OF SIMMI 4.0

Dimension **Vertical Integration**	Dimension **Horizontal Integration**	Dimension **Digital Product Development**	Dimension **Cross-sectional technology criteria**
Stage 5 – Optimized full digitization: The company is a showcase for Industry 4.0 activities. It collaborates strongly with its business partners and therefore optimizes its value networks.			
Continuous cross-corporate integration that is constantly optimized.	Continuous cross-corporate integration and collaboration in value networks.	Product development is processed digitally inside and outside the company (digitized end-to-end solution).	Simulation and optimization of value and information flows in real-time within the value network. IT security adjusts promptly to new risks. Occurring security problems are immediately solved. Encryption is optimized along the value networks.
Stage 4 – Full digitization: The company is completely digitized even beyond corporate borders and integrated into value networks. Industry 4.0 approaches are actively followed and anchored within the corporate strategy.			
Continuous cross-corporate integration.	Continuous cross-corporate integration in value networks.	Product development information are digitally forwarded.	Service-oriented cloud-based platform. Services are offered for the partners in the value networks. Information and data are exchanged in real-time along the supply chain. Optimization of the entire production through Big Data solutions. Access to data is protected. Cross-corporate encryption of data and authentication for global access.
Stage 3 – Horizontal and vertical digitization: The company is horizontally and vertically digitized. Requirements of Industry 4.0 have been implemented within the company, and information flows have been automated.			
Complete internal/enterprise-wide integration of all enterprise systems and machines.	Complete internal/enterprise-wide integration of all enterprise systems and machines.	Product development is continuously digitally supported.	SOA has been established. All functions are provided as services. (Semi-) products and their functionalities are available as services. To exchange information within the enterprise, cloud principles are applied. Production is adjusted and optimized in real-time. IT security is increased through the use of an advanced security model. Access to data is continuously protected, and data is transmitted in an encrypted state within the enterprise.
Stage 2 – Cross-departmental digitization: The company is actively engaged with Industry 4.0 topics. Digitization is implemented across departments and first Industry 4.0 requirements are implemented throughout the company.			
Cross-departmental integration	Cross-departmental integration	Production and product development is supported by several enterprise systems. Data and information exchange is not automatized.	Implementation of first services (SOA with an enterprise service bus (ESB)). First experience with Big Data and its applications. Development of the first IT security models
Stage 1 – Basic digitization level: The company has not addressed Industry 4.0. Requirements are not or only partially met			
Integration of enterprise systems only departmental-specific. The enterprise systems along the enterprise's value chain support only their respective fields of activity	Integration of enterprise systems only departmental-specific. The enterprise systems along the enterprise's value chain support only their respective fields of activity	Product development is not digitally supported	No service-oriented or cloud-based approaches. Data and information flows are not used for product improvement/optimization. Confidentiality, availability and integrity of the data are not guaranteed.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1303–1307
DOI: 10.15439/2016F330
ACSIS, Vol. 8. ISSN 2300-5963

Maturity of IT systems supporting communication processes in HCM in a modern organization

Andrzej Sołtysik
University of Economics in Katowice
ul. 1 Maja, 40-287 Katowice,
Poland
Email:
andrzej.soltysik@ue.katowice.pl

***Abstract*—The aim of this paper is to analyze available maturity models in the context of assessment of the maturity of IT systems that support communication processes in HCM. The paper presents theoretical issues connected with the evolution of information systems in context of support Human Capital Management (HCM) in a modern organization. Selected problems connected with assessment of maturity were presented, and examples of models for maturity assessment were analyzed in the context of their use for evaluation of HCM Information Systems. As a conclusion, the paper indicates necessity to create a new dedicated method for assessing maturity of the systems analyzed.**

I. Introduction

MODERN organizations still search for tools that would be able to ensure them advantage over their competitors. A necessary condition for an efficient functioning of an organization is the use of all kinds of solutions that will enable optimization and improvement of processes taking place in an organization. In the context of the development of an organization, it is especially important to support basic processes aimed at knowledge acquisition and processing as well as its skilful use in practice.

Employees of an organization may turn out to be the basic factor guaranteeing its development and achievement of advantage. For performance of tasks assigned by managers, employees use available knowledge. Knowledge can be acquired, processed and distributed as a result of creation and maintenance of efficient communication channels, both within an organization and between an organization and its environment. The use of mature IT systems supporting the performance of communication processes makes it easier to fulfill this task. As a result of these changes, traditional organizations turn into knowledge-based organizations and concentrate their activities on human resources, that are the most important of immaterial resources at the disposal of an organization.

The first part of the paper will discuss the evolution of the role fulfilled by employees in a modern organization, from human resources to emergence of human capital. Further, the paper will discuss tools supporting the performance of communication processes in HCM in an organization, with particular reference to IT systems.

Next, it will analyze preliminary results of research on IT systems designed to support processes of human capital management in Polish organizations. Next part will feature a review of several models in the context of evaluation of the maturity of IT systems designed to support communication processes in HCM.

II. Evolution of the Role of Employees in a Modern Organization

Not so long ago, employees were treated as one of the resources that an organization acquired and had at its disposal. For those managing an organization, it was important that the individual resources, including human ones, were at a specified level [1]. Management of human resources was concentrated mainly on the quantitative and economic aspect of managing staff - the so-called "hard HRM" [2]. HR managers mainly dealt with administration of remuneration, handled all the issues connected with labour law, management of an organization and working time [3]. Attention was paid to employees' skills - including even the definition of the level of competencies required for employment at a specific - but their use was optimized by managers by means of traditional methods and techniques covering acquisition, development and use [4], [5].

Most modern theories of HR management definitely depart from treating employees like objects. This new approach is based on the concept of the so-called soft HRM proposed in work [3]. The soft HRM makes employees committed to their work by encouraging them to identify with the objectives and mission of the organization and by involving them in defining further tasks [1]. A HR strategy more and more often recognizes and takes into account employees' talents, becoming the foundation of a modern view of the role of people in an organization. Human capital plays a key role in every organization, determining the differences between organizations and constituting the actual basis for competitive advantage [6]. Appropriate people can be the fundamental success factor, but inappropriate ones may contribute to failure of an organization. Summing up: human capital mainly involves

the knowledge of an organization's members, their skills, experience, ability to solve problems, willingness to create and introduce innovations [7], qualifications, creativity and loyalty towards an organization. The human capital at the disposal of an organization should be efficiently managed, which requires the use of reliable tools supporting the basic processes related with this task. Human capital management (HCM) is a set of practices related to people resource management. These practices are focused on the organizational need to provide specific competencies and are implemented in three categories: workforce acquisition, workforce management and workforce optimization.

III. TOOLS DESIGNED TO SUPPORT THE PERFORMANCE OF COMMUNICATION PROCESSES IN HCM IN AN ORGANIZATION

In the free-market economy, the impact of the external environment and conditions of operation are similar for all organizations. Basically, all organizations competing on the market have free access to employees and supporting technology. In such a situation, the main factor determining an organization's competitive advantage is acquisition, creation and appropriate use of its human capital. In order to be able to fully use employees' potential, it is necessary to not only ensure their development, but also to try to make the most valuable of them stay in the organization. We should, however, bear in mind that all organizations can support employees using generally available solutions for that purpose.

Efficient performance of basic processes connected with human capital management requires the use of various available solutions and technologies. Particular emphasis should be placed on systems designed to support broadly understood communication processes. As a response to this demand, more and more IT solutions are created to support the performance of processes connected with human capital management. The use of the latest IT technologies supports learning processes, and the traditional HR processes implemented by an organization are transformed into knowledge-based HCM processes, which leads to the emergence of a new model of an employee, the so-called knowledge worker. Employees' knowledge, skills, abilities, motivation and values are becoming increasingly important and constitute the basic element determining the role of people and related human capital in a modern organization [8]. Without IT support, process improvement may take too long to perform traditional activities connected with administration of a large amount of paper documentation, training materials, instructions describing the performance of assigned tasks, reports or current HR and payroll documents. Implemented IT systems are expected to unify HR processes, systematize the work of the HR department, and support managers in managing subordinate teams. [9]

Employees of an organization more and more often have basic or extended knowledge of modern IT technologies. This knowledge is necessary to fully take advantage of the potential offered by the different technologies. Although there is growing awareness of the role fulfilled by these technologies in supporting communication processes connected with HCM, organizations should take actions aimed at improvement of their employees' competencies connected with the use of such technologies.

This thesis is confirmed e.g. by results of a survey concerning applications of selected modern IT technologies in the different processes connected with acquisition, creation and maintenance of human capital in a knowledge-based organization. The quantitative studies were conducted at the beginning of 2015 by the technique of paper questionnaire interviews (PAPI), on a randomly selected sample of 196 people at different ages, with different place of residence, education and employment relationship. The group consisted of students and graduates of I degree and II degree full-studies and post-graduate studies. Although the studies were conducted mainly in the territory of Silesia, the respondents came from or lived in different regions of the country and abroad.

Results of the studies show that the use of solutions designed to support communication processes is quite widespread in Poland. Of 196 respondents, 121 people (61.73% of the sample) were employed at the moment the survey was conducted. Of those employed, as many as 117 (96.69 % of all those employed) declared that their employer had a corporate website or a different service supporting information exchange. Slightly fewer, i.e. 92 employers (76%) used a computer network or advanced IT systems. 56 respondents (46.28% of all employees) declared that their organization had mechanisms in place to enable exchange of knowledge with its environment. 32 of them (62.75% of all employees) claimed that their employer had or planned to implement in the near future procedures or agreements the subject of which was acquisition, creation or exchange of knowledge with other enterprises or external institutions. 51 respondents (42.15% of all employees) declared that their organization acquired, shared and exchanged knowledge with its environment apart from performing basic business processes. Of this group, 28 people (23.14% of all employees) definitely declared that their organization was focused on knowledge management processes, noticing at the same time the need to use IT solutions to support these activities.

The same studies also revealed which of the IT systems had the largest potential in supporting processes connected with HCM. Processes connected with HCM in an organization were divided into three areas:

- acquisition of human capital
- creation, development and improvement of human capital
- maintenance of human capital.

The studies conducted confirmed the dominating role of solutions supporting processes connected with internal and external communication in all three areas of HCM.

In processes connected with acquisition of human capital, local and wide area computer networks as well as solutions whose operation is directly based on them (Internet) (77.78%) are most significant according to respondents. Business social networks such as Goldenline or Linked in, as well as discussion groups (77.78%) were rated as the most important. Slightly less useful (75.56) are, according to respondents, traditional recruitment portals and social networks (Facebook, Twitter,...) as well as interface agents which are often encountered on various services in the form of virtual advisers. Respondents were equally interested in agent solutions used to search for information (62.22%) and agent solutions used to search the Internet (66.67%).

In processes connected with creation of human capital, a particularly important role was played by IT systems that used computer networks (82.22%). Of similar importance are traditional social networks, followed by discussion groups (77.78%), which in this context enable exchange of knowledge between their users, and interface agents that support training processes.

In the third area, significant importance of social networks, or virtual advisers (71.11%) was highlighted in processes connected with maintenance of human capital. Social networks facilitate knowledge distribution and exchange with the environment, whereas interface agents support users using knowledge possessed by an organization Respondents showed slightly lower trust in discussion groups supporting information exchange (68.89%).

Currently, the situation on the market of IT solutions designed to support the performance of processes connected with HCM is increasingly better. Producers deliver an increasing number of less or more advanced IT systems that use all available IT technologies. However, only "the best", most tailored IT systems can meet an organization's expectations. It is thus very important to possess reliable tools for evaluation of applied software to identify its strengths and weaknesses. Well adjusted tools can fully support processes connected with HCM.

IV. Evaluation of the Adequacy of IT Systems Used to Support Communication Processes in HCM

The key to achievement of effective HCM is to ensure efficient knowledge distribution; it is necessary to make communication channels available and to involve employees in cooperation using these channels. Recognizing the necessity to improve the performance of the different processes connected with reliable communication channels for information exchange as part of HCM processes, companies often allocate large budgets for implementation of IT systems, but are unable to find out whether the solutions used actually support the performance of processes to a sufficient extent. A lot of money is spent on training courses, but managers are unable to determine the actual increase in competencies of employees that participated in them. Thus, it is often argued that investments in IT systems that support communication processes and in training courses do not result in the improvement of their performance, and training courses completed by employees do not lead to acquisition of competencies relevant to a company.

In order to achieve the expected performance and the quality of the functioning of the new methods of software development organizations it is necessary to apply new tools offering the ability to manage this process. One of such methods of evaluation is to examine the maturity level. The term of maturity is defined in the Dictionary of the Polish Language as: "...the state of having taken on the final form, achievement of the final stage of the development or process of shaping... " [10]. The concept of maturity was initially used in psychology to refer to "achievement by an individual of a certain desired mental or emotional state" [11]. In management studies, it appeared in the 1970s. [12]. It was accepted in the theory of management that apart from the extreme states of immaturity and maturity, there is also a certain number of intermediate states [13]. In the broadest sense, the concept of maturity can be examined in relation to an organization as a whole. In the context of management of processes performed within an organization, maturity can be viewed as managerial maturity, process maturity in the area of technology, quality, knowledge or culture as well as praxeology. Integrated maturity of an organization includes responsibility, reliance on trust in the business activity and striving after perfection. Specialized maturity takes into account process and technical approaches, quality, culture, management of knowledge, intellectual capital and, above all, management of human capital [14]. In the context of IT systems, the concept of maturity is usually associated with the field of software engineering [15].

In practice, this concept is most often examined in the following contexts:

- process maturity [16] [17], [18],
- project maturity [19],
- quality maturity,
- implementation maturity,
- different combinations, e.g. process and project maturity

As was already mentioned, in a modern organization, the performance of processes of HCM is inseparably connected with the use of appropriate IT systems that support communication processes, enable efficient management, automation, monitoring and optimization of these processes, allowing thereby higher levels of process maturity to be achieved. At the same time, IT systems, the process of their creation and degree of their use in an organization may also be evaluated [12]

For maturity assessment, flexible tailor-made models are used. Models designed to evaluate the level of maturity allow an organization to assess its methods and processes in accordance with the best practices of management and based on clearly defined external reference values. There is a range

of models designed to evaluate the maturity level of models, from Crosby Quality Management Maturity Grid developed in the 1970s [20] to models dedicated to maturity evaluation developed based on internal assumptions of specific organizations. Models differ from each other in the scope of maturity level measurement which can be performed in various areas of an organization's activity. Many of them were taken into account in their research and described in their works by [21], [13].

Maturity modelling for the purpose of management and control of processes taking place in an organization is based on the method of evaluation of an organization. Its level of maturity can be evaluated on a scale from 0 (lack) to 5 (optimal). This approach is based on the maturity model developed in 1991 by Software Engineering Institute (SEI) - a model for assessing maturity of development processes and potential (capability) of IT system (software) development (Capability Maturity Model for Software) CMM [22], [23].

HCM is a specific area, as it deals with processes connected with the most spontaneous and unpredictable of the resources at the disposal of an organization - its employees. The dynamics accompanying processes taking place in the HR sphere justifies creation of detailed models designed to assess the maturity level of IT systems connected with acquisition, creation and maintenance of human capital. Although there are no models that directly refer to maturity of such systems, there is a range of models that focus on the one hand on assessment of the maturity of IT systems, while on the other hand - on assessment of process maturity. There are also models designed to measure maturity of skills offered to an organization by its employees [24].

An example of such a model in the area of human capital management is PCMM. The model was developed to support processes connected with improvement of employees' skills which are one of the most important premises for creation of an organization's human capital and are regarded as key factors leading to success achieved by knowledge-based organizations. PCMM allows an organization to identify necessary activities to support employees' development depending on the current level of an organization. The model indicates competencies that are of key importance for proper operation of an organization in a changing environment, increasing effectiveness of HCM [25].

In available sources, researchers and practitioners give a lot of attention to the problem of maturity of IT systems. There are also a few models designed to assess maturity of systems connected with HCM in an organization.

One of more interesting examples is HCM applications implementation maturity model proposed by K. Jones [25] describing five levels of progressively more mature HCM applications implementation capability.

Presented model describes four levels of progressively more mature HCM applications implementation capability

1 st., Level: Technology-Centric - focused strictly on technology, with lacks a compelling business case to implement new software supporting HCM,

2 nd. Level: Process Automation – implementation process of information systems supporting HCM in organization providing a team implementing new software without ongoing and future software upgrades,

3 rd. Level: HR Customer-Centric – where implemented systems have inconsistent communication with stakeholders & audiences,

4 th., the highest Level: HR Customer-Centric – where software changes are driven through documentation of preexisting business, with project consolidation and rationalization of business processes and continual communication and engagement of stakeholders and audiences,

The above-described examples of the models, as well as many other presented in sources, enable assessment of the maturity of processes taking place in HCM, assessment of the maturity of "people capability", maturity of IT systems and implementation of HCM Information Systems (HCMIS) in an organization. These are, however, general models which treat HCM in a comprehensive way.

Effective HCM embraces an integrated approach that requires concepts and practices that are tested and proven. Practical management tools are required to help HR leaders diagnose problems quickly, then prioritize and implement reforms along an HCM maturity model in an environment that has many and, at times, conflicting sources of priorities.

The aim of the author was to attempt to match a method to assessment of process maturity of IT systems designed to support processes connected with communication processes taking place in HCM in an organization.

Although many general and specialized models designed to assess maturity in various contexts can be found, there are no detailed works that directly address research on maturity of IT systems that support the performance of communication processes as part of HCM in an organization. In order to fill this gap in methodology, it is necessary to develop a model and new method that enable assessment of maturity of such solutions. The basic task in this situation is to develop tools that will make it possible to determine maturity levels, and then to find out what is the maturity level of an examined solution.

V. Conclusion

The aim of the paper was to indicate tools that would enable organizations, an assessment of the adequacy of IT systems used to support communication processes in HCM. IT systems are a necessary element required to support processes connected with information exchange in HCM (e.g. knowledge distribution). Only the use of reliable solutions will allow an organization to create human capital which will give it a chance to gain competitive advantage. For an organization to be able to use appropriate solutions, it

needs a reliable method for their assessment. The paper was an attempt to relate known methodologies for examining a system maturity to IT systems designed to support communication processes. It analyzed three selected models supporting assessment of maturity in three aspects connected with the selected research subject. This, however, does not exhaust the subject, although their use at a further stage of research creates significant chances in this aspect. My research shows that such systems function in almost every organization, with many organizations attaching great importance to them. However, given the lack of direct references in the academic literature to the issue of maturity of IT systems used to support communication processes in HCM, a few initial assumptions have to be made. The area of using IT systems to support HCM is nowadays intensively explored both in research and practice. There are already applications that support the different HCM processes, and over time they will fulfill an increasingly important role. Therefore, it is very important to ensure tools that will allow the usefulness of such solutions to be verified.

Given the volume of knowledge and literature, it is difficult to discuss all the aspects impacting the use the model to assess the maturity of HCMIS. None of the available maturity models takes into account all the conditions accompanying the introduction of modern communication technologies into HCM. It is thus necessary to create a new model that is based on the models available on the market, but, above all, takes into account all the conditions connected with implementation of HCMIS in an organization. Despite a large number of available models, not a single one that would enable a comprehensive assessment could be indicated. Necessary features were showed by a few models. It is thus justified to conduct appropriate research and propose a new model that would enable such assessment, defining maturity levels based on e.g. the level of knowledge representation. Although research is conducted, unexplored areas still exist. For instance, there are no solutions designed to assess maturity of IT systems that support communication processes and assist employees in their careers. However, with increasing availability of necessary tools, the possibilities to assess IT systems in their support of communication processes will increase, and the number of their successful applications in the area of HCM will undoubtedly grow.

An article is a background for my future researches focused on bridging the gap in methodology and creating new model that enable assessment of maturity of IT systems supporting communication processes in HCM.

ACKNOWLEDGMENT

The issues presented constitute a beginning part of the authors research into the aspect of improving communication processes in Human Capital Management

REFERENCES

[1] K. Legge, "*Human Resource Management: Rhetorics and Realities*", Basingstoke: Macmillan, 1995, pp 66-67
[2] E. Vaughan, "*The trial between sense and sentiment: a reflection on the language of HRM*", Journal of General Management, 19, 3, 1994, pp. 20-32.
[3] "*New Perspectives on Human Resource Management*". John Storey Ed., Routledge. London. Distributed by The Law Book Company Limited, 1989
[4] J. Drucker, G. White, A. Hegewisch, L. Mayne "*Between hard and soft HRM: human resource management in the construction industry*", Construction Management and Economics, 14, 1996, pp. 405-416.
[5] T. Keenoy, P. Anthony, "*HRM: metaphor, meaning and morality*", In: P. Blyton and P. Turnbull Eds., Reassessing Human Resource Management, London: Sage, 1994, pp. 233-255
[6] J.L. Chatzkel, "*Human capital. The rules of engagement are changing*", in *Lifelong learning in Europe*, vol. 9, nr 3, 2004, pp. 139-145.
[7] W. Kotarba (red.), "*Ochrona wiedzy a kapitał intelektualny organizacji*", Polskie Wydawnictwo Ekonomiczne, Warszawa 2006, pp. 18-19.
[8] A. Sołtysik, "*Zarządzanie Kapitałem Ludzkim z wykorzystaniem SAP HCM*" In: Wybrane Zagadnienia Wykorzystania Systemu SAP ERP w organizacji. M. Żytniewski Ed., Wydawnictwo Uniwersytetu Ekonomicznego w Katowicach, Katowice, 2015: pp. 219-258;
[9] A. Sołtysik, "*Wspieranie procesów pozyskiwania, kreowania i utrzymania kapitału ludzkiego w organizacji opartej na wiedzy. Wstępne wyniki badań*". In: Technologie Wiedzy W Zarządzaniu Publicznym, Wydawnictwo Naukowe UE w Katowicach, 2015
[10] Słownik Języka Polskiego SJP.PL
[11] T. E. Moffitt, "*Adolescence-Limited and Life-Course-Persistent Antisocial-Behavior – A Developmental Taxonomy*", Psychological Review, no. 100/1993, pp. 674–701
[12] W. Flieger, "*Odkrywanie procesów jako składowa dojrzałości procesowej urzędów administracji samorządowej*", In: Roczniki Kolegium Analiz Ekonomicznych, Szkoła Główna Handlowa, Warszawa, Zeszyt 33/2014
[13] D. Hillson, "*Assessing Organizational Project Management Capability"*, Journal of Facilities Management, no. 2/ 2003.
[14] E. Skrzypek, "*Dojrzałość jakościowa organizacji w świetle teorii i doświadczeń organizacji*", [http://www.marketingirynek.pl/files/1276809751/file/konferencja_5_2014.pdf]
[15] B. Begier, "*Inżynieria oprogramowania ? problematyka jakości*", Wyd. Politechniki Poznańskiej, Poznań 1999
[16] P. Grajewski, "*Przesłanki podejścia procesowego do projektowania i zarządzania organizacją*", In: J. Lichtarski Eds, Nowe kierunki w zarządzaniu przedsiębiorstwem – wiodące orientacje, Wydawnictwo UE we Wrocławiu, Wrocław, 2014.
[17] B.W. Cieśliński, "*Doskonalenie procesowej orientacji przedsiębiorstw: model platformy treningu procesowego*", Wydawnictwo UE we Wrocławiu, Wrocław, 2011, p.39
[18] G. Jokiel, "*Podejście procesowe w zarządzaniu – geneza i kierunki rozwoju koncepcji*", In: Podejście procesowe w organizacjach, S. Nowosielski Ed., Wydawnictwo UE we Wrocławiu, Wrocław, 2009.
[19] J.R. Meredith, S.J. Mantel, "*Project Management*", John Wiley& Sons, New York, 2000
[20] P. Crosby, "*Quality is free*", McGraw Hill, New York 1979, pp. 32-33.
[21] P. Wyrozębski, M. Juchnowicz, W. Metelski, "*Wiedza, dojrzałość, ryzyko w zarzadzaniu projektami*", Oficyna Wydawnicza SGH, Warszawa 2012, pp. 127–128
[22] M.B. Chrissis, M. Konrad, S. Shrum, "*CMMI for Development: Guidelines for Process Integration and Product Improvement*", 3, Addison Wesley, 2011
[23] M. C. Paulk, B. Curtis, M. B. Chrisis, "*Capability Maturity Model for Software, Version 1.1,*" Software Engineering Institute Technical Report, CMU/SEI-93-TR, February 24, 1993
[24] B. Curtis, W.E. Hefley, S. Miller. The People Capability Maturity Model: Guidelines for Improving the Workforce. Reading, MA: Addison Wesley Longman, 2002.
[25] K. Jones "*http://www.bersin.com/Lexicon/details.aspx?id= 12843*"

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1309–1316
DOI: 10.15439/2016F72
ACSIS, Vol. 8. ISSN 2300-5963

An Information Security Framework for Ubiquitous Services in e-Government Structures: A Peruvian Local Government Experience

Manuel Tupia
Pontificia Universidad Católica del Perú. Engineering Department. Av. Universitaria 1801 San Miguel, Lima, Perú. Email: tupia.mf@pucp.edu.pe

Mariuxi Bruzza
Pontificia Universidad Católica del Perú. Engineering Department. Av. Universitaria 1801 San Miguel, Lima, Perú. Email: a20146472@pucp.pe

Flavio Rodriguez
Pontificia Universidad Católica del Perú. Engineering Department. Av. Universitaria 1801 San Miguel, Lima, Perú. Email: flavio.rodriguez@pucp.edu.pe

***Abstract*—This paper describes a framework designed to establish vital conditions of information security for *ubiquitous* services (U-Government) both in district and province municipalities (departments' capitals) within the Peruvian electronic (e-government) government structures. The framework contains current regulations concerning information security, data privacy, business continuity, and natural disasters management based on good international practices, including but not limited, ISO 27001, ISO 27002, ISO 22301 standards. The aim is to help implement security controls in the use of mobile services which are part of the e-government services catalogue. The framework structure is closely related to the COBIT 5.0 process model.**

I. Introduction

At present, electronic government structures include a solid component of services oriented to the use of devices and mobile solutions [1]. These services are intended to take advantage of the widespread use of this type of devices by citizens and their knowledge of mobile applications [2]. Most local (municipalities) electronic government initiatives in Peru are focused on this type of services rather than web services as displayed in Figure 1:

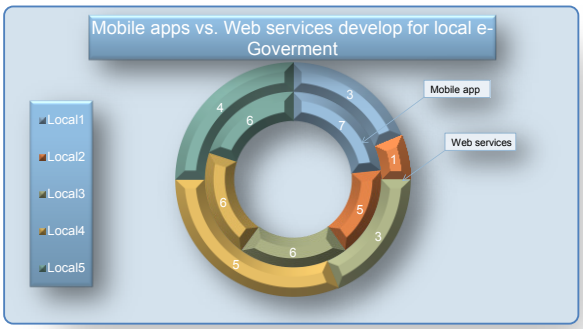


Fig. 1 Comparative table taken from http://www.movil.softwarepublico.gob.pe/ (2015)[1]

New technological developments and innovations are setting the stage for a higher demand of U-government services. This will lead, in turn, to new services for the government which will be forced to increase their number and take into account pertinent restrictions, e.g., information security for regulatory compliance [3], [4], [5], [23].

As there are doubts about the information security in the mobile applications (apps) and due to the huge amount of information to be protected by governmental institutions rendering services through them [6], [22], it is mandatory to develop a framework that facilitates the implementation of controls and the follow-up of good international practices on information security both in the implementation projects of U-Services (ubiquitous services) in Peruvian local governments and the assurance of the already existing apps. The proposed framework must comply with current domestic regulatory framework.

This paper discusses the general structure of a security framework based on COBIT 5.0 and above all its process reference model (PRM) [7] in ISO 27001, ISO 27002 and ISO 31000 standards.

Likewise, the current Peruvian regulations including the Personal Information Protection Law[2] and Peruvian Technical Standards on security will be referred to[3].

II. Ubiquitous Government

A. Ubiquitous computing

Ubiquitous technology (ubicomp, English acronym ubiquitous computing) is an extension of mobile informatics based on access technologies through mobile or handheld devices [8]. It consists of re-orienting ITs to propose mobile technology-based solutions for access, consumption and exploitation of information in an effort to render services as a support to business processes; thus paving the road on which companies and organizations would modify their strategies to meet the needs of clients and stakeholders.

[1] Local label represents a Peruvian public institution

[2] Personal Data Protection Law N°29733

[3] NTP ISO/IEC 17799 and NTP ISO/IEC 27001

B. U-Government

Ubiquitous computing has been closely linked to the electronic government on the premise that it will assure availability and multichannel connections to electronic services offered by the government to the citizen by means of mobile or web-based solutions. Ubiquitous government or u-government mirrors the new manner of interaction and transaction between the government and citizens and other stakeholders in such a way that access to these services is available through entry points (multi-purpose single windows) anywhere and anytime, from different types of mobile electronic devices [9], [13]. Precisely, the main concern of U-services implementation relates to security around the involved transactions and sending of information.

C. Information Security requirements in U-services

Most e-government implementations entail the design of web pages and services offered to citizens from these portals [10]. There are three types of said services: statics information, data consultation (mild interaction that may not imply personal data entry) and transactions (advanced interaction encompassing business transactions such as payments and personal data entry) [11], [20].

Below are the chief considerations on the security related regulatory compliance [14], [21]:

- Is there personal data involved in the service?
- Are digital signatures and certificates being used to authenticate the citizens for transactions?
- Are transactions encrypted? [12]
- Are there VPN infrastructures between local government (Website) and citizens to conduct the transactions involved in the services?
- Do services require the intervention of a payment gateway?
- Do services require the intervention of external payment methods (e.g. PayPal)?

D.Regulatory Requirements of Peruvian Local Governments

Peruvian local governments (province and district municipalities) set forth the following regulations with respect to information security, e-government, and data privacy:

- Mandatory use of the Peruvian Technical Standard "ISO NTP/IEC 27001:2014 Information Technology. Security Techniques. Information Security Management Systems. Requirements 2nd Edition",
- Open Government Action Plan (Open Data Plan AGA) 2015-2016
- Compliance of the Electronic Government Domestic Politics 2013 – 2017
- Personal Data Protection Law N° 29733
- Digital Signatures and Certificates Law N° 27269
- Law N° 29985 of Electronic Money as a financial inclusion tool

Local governments are bound by the current regulations to:

- Formulate the electronic government plan
- Establish the e-government structures needed to render U-services
- Implement the controls needed for U-services that require digital signatures and certificates for citizen authentication.
- Identify personal data sources held and managed by local governments, and comply with data privacy regulations.
- Set up an information security management system (ISMS) in accordance with ISO 27001 and ISO 27002 standards.

E. Applications involved in U-government in Peru

The nature of mobile applications involved in local government U-services in Peru is basically intended to load and report tax information, property tax, and information about traffic offenses. Only a few are focused on *payment transaction* of taxes, local duties, traffic offenses and the like. However, there exist regulations aimed at gradually increasing transactions by using, for example, electronic money in which mobile devices will serve as payment wallets for these procedures.

All these mobile services imply use, storage, handling and traffic of data of a personal nature, all of them regulated in Peru since 2013. Personal data-handling organizations are responsible for implementing – and stating their status as such – a series of ISO 27002 security controls, whereas governmental organizations must establish ISMS based on ISO 27001.

Nevertheless, the methodological gap identified shows lack of a suitable implementation guide for controls and good information security practices required for said cases [15], [16].

The proposed framework intends to fill this gap for U-services, which are part of more complex e-government organizational structures.

III. PROPOSED FRAMEWORK

A. Framework's General Structure

The framework has been divided into four parts:

- Stakeholders' needs
- Alignment matrix between business goals and information security goals for U-services supporting business goals.
- Information Security Process reference model (ISPRM) containing the relevant information security domains for the applications included in the U-services, and for the respective compliance, in light of the previous regulations.

- Current process implementation guide in the ISPRM.

The following paragraphs detail each of these components.

B. Stakeholders' Needs

In this case, by making an analogy with the COBIT 5.0 presentation, we can establish up to five related business objectives:

i. Optimize investments
ii. Optimize risks
iii. Comply with current and competent regulations
iv. Optimize resources
v. Satisfy citizens

Corporate and information security goals will be added to the above objectives in the following alignment matrix.

C. Alignment Matrix

The alignment matrix is introduced in two tables whose dimensions are those of the Balance Score Card. Table 1 shows the framework's corporate goals, whereas Table 2 lists the information security goals regarding development and delivery of U-services, which are in line with the business objectives. Letters P and S within the cells containing objectives mean that the goal is directly related to such objective.

Table I provides the 11 corporate goals related to e-Government and, intrinsically, to delivery of U-services. These goals have been adapted from [7]:

1. Value for stakeholders from the investments in e-Government
2. Portfolio of U-services
3. Risks of managed business
4. Compliance of laws and external rules with regard to e-Government
5. Orientation towards citizens
6. Continuity of U-services
7. Optimization of costs in delivery of U-services
8. Optimization of business processes involved in delivery of e-Government services
9. Compliance of internal policies and procedures related to delivery of e-Government services
10. Trained and motivated staff
11. Innovation culture in e-Government services, especially in the U-services

On the other hand, Table II shows the 12 proposed security goals, which have been adapted from [17]:

1. Alignment of the information security towards the business.
2. Contribution of information security to the compliance of e-government related regulations
3. Business risks regarding information security and compliance managed.
4. Top Management's commitment to decision-making related to information security of U-services.
5. Delivery of secure U-services according to business requirements
6. Adequate use of applications and information and other technologies for U-services development.
7. Design of appropriate U-services to citizens' needs
8. Optimization in the use of information, resources and capabilities of ITs in e-government services, especially U-services.
9. Adequate delivery of U-services to meet citizens' needs and compliance of the business requirements.
10. Compliance of security policies and current regulations in U-services
11. Skilled and motivated staff responsible for information security.
12. Knowledge, awareness and training in information security as part of the innovation process in delivery of e-government services.

The *so-called goals cascade* proposed by COBIT in order to conduct the respective alignment between business and security [18] will depend on each particular company and on their services provided as a part of its e-government. The goal cascade is not included in our research for an organization, however we do recommend it.

D. Information Security Process reference model (ISPRM)

As the next component, our research has put forward a series of processes both for government and information security management specific for development, acquisition, and maintenance of U-services.

After a first division by government processes and management processes, four domains of security management have been determined: planning and organization, acquisition and implementation of U-government, delivery of U-services and monitoring of U-services. Below are the definitive processes listed in Table II.

Government Processes:

1. Establish and maintain over time government structures for information security
2. Make sure risk optimization includes information security risks in e-government services
3. Secure resources optimization needed for establishing the information security government and its continuity over time

Management processes of U-Services' information security:

Planning and organization domain:

4. Manage and maintain the information security management framework
5. Prepare and maintain the information security strategy
6. Define the Ubiquitous architecture for services corresponding to e-government.

TABLE I.
CORPORATE GOALS IN EUFRAME-SECURITY

Balance Score Card Dim.	Corporate Goal	Obj i.	Obj ii.	Obj iii.	Obj iv.	Obj v.
Financial	1. Value for stakeholders from e-Government Investments	P			P	
	2 . U-services portfolio	S			P	P
	3. Managed risks		P	S	S	
	4. Compliance with laws and regulations related to external e-Government			P		
Client	5. Citizen orientation			S		P
	6. U-services continuity	S		P	P	S
	7. Cost optimization providing U-services	P		S	P	S
Internal	8. Optimization of business processes involved in provision of e-Government			S	P	S
	9. Compliance with internal policies and procedures related to provision of e-Government			P	S	S
Learning and knowledge	10. Personal prepared and motivated		P	S	P	
	11. Culture of innovation in e-Government services and in particular in the U-services		P	S	P	S

TABLE II
GOALS RELATED TO INFORMATION SECURITY IN EUFRAME-SECURITY

Balance Score Card Dim.	Information Security Goal	Obj i.	Obj ii.	Obj iii.	Obj iv.	Obj v.
Financial	1. Alignment of the information security towards the business	P			P	
	2. Contribution of information security to the compliance of e-government related regulations		S	P	S	
	3. Business risks regarding information security and compliance managed		P	P	S	
	4. Top Management commitment to decision-making related to information security of U-services	S			P	S
Client	5. Delivery of secure U-services according to business requirements		S	S	P	P
	6. Adequate use of applications and information and other technologies for U-services development		S	S	P	P
Internal	7. Design of appropriate U-services to citizens' needs		S	S	P	P
	8. Optimization in the use of information, resources, and capabilities of ITs in e-government services, especially U-services.	S			P	S
	9. Adequate delivery of U-services to meet citizens' needs and compliance of business requirements.			S	P	P
	10. Compliance of security policies and current regulations in U-services		S	P	S	
Learning and knowledge	11. Skilled and motivated staff responsible for information security		P	S	P	
	12. Knowledge, awareness and training in information security as part of the innovation process in delivery of e-government services		P	S	P	S

7. Manage U-services portfolio
8. Manage service level agreements (SLA) of U-services
9. Manage information security risks in U-services
10. Manage risks of non-compliance in U-services
11. Manage information security in the process of designing, developing, acquiring, and maintaining U-services within the e-government.

services based on this type of aplications. The list of procedures is as follows:

Governance procedures for the implementation of government processes[4].

- Define information security policies
- Determine the people responsible for the information security across all the organization

TABLE III
INFORMATION SECURITY PROCESS REFERENCE MODEL (ISPRM)

Government	1. Alignment of the information security towards the business	Monitoring U-services
	2. Contribution of the information security to the compliance of the e-government related regulations	19. Assess performance of U-services
	3. Business risks regarding information security and compliance managed	20. Assess compliance of information security regulations of U-services
Planning and Organization	4. Manage and maintain the information security management framework	Acquisition and implementation of U-government
	5. Prepare and maintain the information security strategy	12. Manage U-services implementation projects
	6. Define the Ubiquitous architecture for the services corresponding to the e-government	13. Define information security requirements at e-government structure level with emphasis on U-services
	7. Manage U-services portfolio	14. Manage availability and capacity of U-services
	8. Manage service level agreements (SLA) of U-services	15. Manage information assets taking part in U-services
	9. Manage information security risks in U-services	Delivery de los U-services
	10. Manage risks of non-compliance in U-services	16. Manage appropriate operation of U-services
	11. Manage information security in the process of designing, developing, acquiring, and maintaining U-services within the e-government	17. Manage problems and incidents of security with U-services
18. Manage continuity of U-services operations		

Acquisition and implementation of U-government domain:
12. Manage U-services implementation projects
13. Define information security requirements at e-government structure level with emphasis on U-services
14. Manage availability and capacity of U-services
15. Manage information assets taking part in U-services

Delivery of U-services Domain:
16. Manage appropriate operation of U-services
17. Manage problems and incidents of security with U-services
18. Manage continuity of U-services operations

Monitoring of U-services Domain
19. Assess performance of U-services
20. Assess compliance of information security regulations of U-services.

E. Implementacion Guide

The guide provides a series of procedures to implement the above listed processes. Our proposal focuses on transaction mobile applications dealing with information of a personal nature (Personally identifiable information – PII o Sensitive Personal Information - SPI), that is why proposed procedures lay emphasis on security and compliance of

- Conduct a risk management methodology at an organizational level, including information security risks.
- Determine the people in charge of conducting the risk analysis including information security risk analysis.
- Incorporate the resources deemed necessary into the budget to establish and maintain the information security government
- Prepare the electronic government plan
- Define e-government structures necessary for delivery of U-services
- Adjust e-government structures to comply with the Peruvian National Police of the 2013 – 2017 Electronic Government.

Below are the procedures for management processes within the security Planning and Organization domain:

- Define a framework for the information security management within the e-government structures.
- In the business strategy and information technology plans define information security

[4] Each procedure may include a series of related projects to be executed and in doing so achieve a complete related process(s)

strategies for e-government structures and services.
- Define U-services as a part of the e-government plan.
- Set up a management mechanism for IT services including U-services.
- Conduct the risk analysis including information security risks of U-services.
- Define Peru's Open Government Action Plan (Open Data Plan AGA) for 2015-2016
- Design an information security management system (ISMS) in accordance with ISO 27001 and ISO 27002 standards in order to comply with the mandatory use of the Peruvian Technical Standard "ISO NTP/IEC 27001:2014 Information Technology. Security Techniques. Information Security Management Systems. Requirements 2nd Edition"

Below are the procedures for management processes within the Acquisition and Implementation of U-government domain:
- Define a framework for projects management of acquisition, implementation, and deployment of U-services.
- Define information security requirements for outsourcing, which supply service assets of U-services or complete delivery.
- Identify personal data sources held and managed, and comply with data privacy regulations.
- Define security requirements and those responsible for compliance of Personal Data Protection Law N° 29733.
- Implement security controls needed for U-services requiring use of signatures and digital certificates for citizen's authentication.
- Define security requirements and those responsible for compliance of Digital Signatures and Certificates Law N° 27269.
- Define security requirements and those responsible for compliance of Law N° 29985 of Electronic Money as a financial inclusion tool.
- Maintain service level related to availability and capacity of U-services.
- Mantain service assets involved in delivery of U-services.

Below are the following procedures for management processes within the Delivery of U-services domain:
- Deliver U-services.
- Set up a help desk for incidents and problems management of the information security of e-government services, including U-services.
- In the business continuity plans and information technologies continuity plans include procedures to maintain continuity of U-services deemed critical.

Finally, below are the procedures for management processes within the domain Monitor U-services,:
- Set up and maintain an internal control system.
- Define metrics and performance indicators of U-services.
- Define metrics and satisfaction indicators of citizens in the use of U-services.
- Conduct measurements of all metrics and indicators of U-services.
- Identify non-compliance of information security within U-services.

F. Good Practices in the Implementation Guide

The guide is based on a series of good practices in which ISO 27000 standards are the most important ones. Table 4 shows the mapping among the above proposed procedures for each domain in line with clauses of the ISO 27002 standard [19].

IV. CONCLUSIONS

The lack of frameworks for the implementation of information security governments results in non-compliance of relevant regulations by local and municipal governments, above all in data privacy matters.

The proposed eUframe-security framework provides a basic guide for consolidating the information security government, thus filling the procedural gap to address the U-government needs.

Next step (which is part of the future work of this paper) is to establish a testing mechanism to validate the model.

Table IV
Mapping ISPRM procedures versus ISO 27002

Domains	Implementación Guide Procedures	Clauses ISO 27002
Government	· Define information security policies.	**5**
	· Determine the people responsible for the information security across all the organization.	**6**
	· Conduct a risk management methodology at an organizational level, including information security risks.	**6**
	· Define the people in charge of conducting the risk analysis including information security risk analysis.	**6**
	· Incorporate the resources deemed necessary into the budget to establish and maintain the information security government	**6, 7, 8**
	· Prepare the electronic government plan	**6**
	· Define e-government structures necessary for delivery of U-services	**6**

	· Adjust e-government structures to comply with the Peruvian National Police of the 2013 – 2017 Electronic Government.	**6**
Planning and Organization	· Define a framework for information security management within e-government structures.	**5, 6**
	· In the business strategy and information technology plans define the information security strategies for e-government structures and services.	**6**
	· Define U-services as a part of the e-government plan.	**6**
	· Set up a management mechanism for IT services including U-services.	**6**
	· Conduct the risk analysis including information security risks of U-services.	**6**
	· Define Peru's Open Government Action Plan (Open Data Plan AGA) for 2015-2016	**6**
	. Design an information security management system (ISMS) in accordance with ISO 27001 and ISO 27002 standards in order to comply with the mandatory use of the Peruvian Technical Standard "ISO NTP/IEC 27001:2014"	**5-18**
Acquisition and implementation	· Define a framework for projects management of acquisition, implementation and deployment of U-services.	**8, 14, 15**
	· Define information security requirements for outsourcing which supply service assets of U-services or complete delivery.	**7, 14, 15**
	· Identify personal data sources held and managed, and comply with the data privacy regulations.	**8, 9, 10**
	· Define security requirements and those responsible for compliance of Personal Data Protection Law N° 29733.	**18**
	· Implement security controls needed for U-services requiring use of signatures and digital certificates for the citizen's authentication.	**18**
	· Define security requirements and those responsible for the compliance of Digital Signatures and Certificates Law N° 27269.	**18**
	· Define security requirements and those responsible for compliance of Law N° 29985 of Electronic Money as a financial inclusion tool.	**18**
	· Maintain service level related to availability and capacity of U-services.	**12**
	. Maintain service assets involved in delivery of U-services.	**8, 12**
Delivery	**· Deliver U-services.**	**12**
	· Set up a help desk for incidents and problems management of the information security of e-government services, including the U-services.	**12, 16**
	. In the business continuity plans and information technologies continuity plans include the procedures to maintain continuity of the U-services deemed critical.	**17**
Monitoring	**· Set up and maintain an internal control system.**	**18**
	· Define metrics and performance indicators of U-services.	**18**
	· Define metrics and satisfaction indicators of citizens in the use of U-services.	**18**
	· Conduct measurements of all metrics and indicators of U-services.	**18**
	. Identify non-compliance of information security within U-services.	**18**

REFERENCES

[1] H. Ranaweera, "Perspective of trust towards e-government initiatives in Sri Lanka", *SpringerPlus*., vol. 5, no 22, pp. 1-11, 2016. http://dx.doi.org/10.1186/s40064-015-1650-y

[2] J. Batlle-Montserrat, J. Blat, and E. Abadal, "Local e-government Benchlearning: Impact analysis and applicability to smart cities benchmarking," *Information Polity*., vol. 21, pp. 43-59, 2016. http://dx.doi.org/10.3233/IP-150366

[3] P. Pitchay Muthu Chelliah, R. Thurasamy, A. I. Alzahrani, O. Alfarraj, and N. Alalwan, "E-Government service delivery by a local government agency: The case of E-Licensing Telematics and Informatics", vol. 33, pp. 925-935, 2016. http://dx.doi.org/10.1016/j.tele.2016.02.003

[4] A. Ramtohul, and K. M. S. Soyjaudah, "Information security governance for e-services in southern African developing countries e-Government projects", *Journal of Science and Technology Policy Management*., vol. 7, pp. 26-42, 2016. http://dx.doi.org/10.1108/JSTPM-04-2014-0014

[5] R. Kennedy and H. J. Scholl, "E-regulation and the rule of law: Smart government, institutional information infrastructures, and fundamental values", *Information Polity*., vol. 21, pp. 77-98, 2016. http://dx.doi.org/10.3233/IP-150368

[6] L. G. Anthopoulos and C. G. Reddick, "Understanding electronic government research and smart city: A framework and empirical evidence", *Information Polity*., vol. 21, pp. 99-117, 2016. http://dx.doi.org/10.3233/IP-150371

[7] ISACA, COBIT 5.0 *For Information Security*. ISACA Publishing, USA, 2012.

[8] J. Krumm, *Ubiquitous Computing Fundamentals*. Chapman and Hall/CRC, USA, 2009.

[9] A. Anttiroiko, "Towards Citizen-Centered Local e-Government – The Case of the City of Tampere", *Idea Group Publishing*, vol. 6, pp. 370–372, 2004. http://dx.doi.org/10.4018/978-1-59140-259-6.ch021

[10] J. Joo and A. Hovav, "The influence of information security on the adoption of web-based integrated information systems: an e-government study in Peru", *Information Technology for Development*, vol. 22, pp. 94-116, 2016. http://dx.doi.org/10.1080/02681102.2014.979393

[11] B. W. Wirtz and O. T. Kurtz. "Determinants of Citizen Usage Intentions in e-Government: An Empirical Analysis", *Public Organization Review*, pp. 1-20, 2016. http://dx.doi.org/10.1007/s11115-015-0338-7

[12] G. Sangeetha and L. Manjunatha Rao, "Modelling of E-governance framework for mining knowledge from massive grievance redressal data", *International Journal of Electrical and Computer Engineering*, vol. 6, pp. 367-374, 2016. http://dx.doi.org/10.11591/ijece.v6i1.9019

[13] A. Djeddi and I. Djilali, "A user centered ubiquitous government design framework", *ACM International Conference Proceeding Series*, 2015. http://dx.doi.org/10.13140/RG.2.1.4890.6000

[14] B. Schneir, "Ubiquitous Surveillance and Security [Keynote]", IEEE Technology and Society Magazine, vol. 34, no. 7270448, pp. 39-40, 2015. http://dx.doi.org/10.1109/MTS.2015.2461232

[15] A. Asquer. "E-government, M-government, L-government: Exploring future ICT applications in public administration", Public Affairs and Administration: Concepts, Methodologies, Tools, and Applications, 2015, vol. 4, pp. 2155-2168. http://dx.doi.org/10.4018/978-1-4666-8358-7.ch11

[16] K. Malladi, S. Sridharan and L. T. Jayprakash, "Architecting a large-scale ubiquitous e-voting solution for conducting government elections", International Conference on Advances in Electronics, Computers and Communications, ICAECC 2014, no. 7002445, 2015. http://dx.doi.org/10.1109/ICAECC.2014.7002445

[17] E. Mello and J. Souza Neto, "A Governance and Management Model for the Public Sector Shared Services Center Based on COBIT 5". COBIT Focus, ISACA Publishing, no. 3, on site http://www.isaca.org/COBIT/focus/Pages/a-governance-and-management-model-for-the-public-sector-shared-services-center-based-on-cobit5.aspx?utm_campaign=ISACA+Main&cid=sm_1202172&utm_content=1460055833&utm_source=facebook&utm_medium=social&appeal=sm, 2016.

[18] K. Maes, P. De Bruyn, G. Oorts and P. Huysmans, "On the Imperative Solicitude for Evolvability Evaluation in Value Management", International Journal of IT/Business Alignment and Governance (IJITBAG), vol. 5, pp. 70-87, 2014. http://dx.doi.org/10.4018/ijitbag. 2014070104

[19] nternational Organization for Standardization, ISO/IEC 27002:2013, Information technology – Security techniques – Code of practice for information security management, Switzerland, 2013.

[20] S. Alghamdi, N. Beloff, "Exploring Determinants of Adoption and Higher Utilisation for E-Government: A Study from Business Sector Perspective in Saudi Arabia", 10th Conference on Information Systems Management, ISM 2015, vol. 5, pp. 1469 – 1479, 2015. http://dx.doi.org/10.15439/2015F257.

[21] P. Chatzoglou, D. Chatzoudes, S. Symeonidis, "Factors affecting the intention to use e-Government services", 10th Conference on Information Systems Management, ISM 2015, vol. 5, pp. 1489–1498, 2015. http://dx.doi.org/10.15439/2015F171

[22] S. Alghamdi, N. Beloff, "Towards a Comprehensive Model for E-Government Adoption and Utilisation Analysis: The Case of Saudi Arabia", 9th Conference on Information Systems Management, ISM 2014, vol. 2, pp. 1217–1225, 2014. http://dx.doi.org/10.15439/2014F146

[23] G. Wangen, E. A. Snekkenes, "A Comparison between Business Process Management and Information Security Management", 1st Workshop on Emerging Aspects in Information Security, EAIS 2014, vol. 2, pp. 901 – 910, 2014. http://dx.doi.org/10.15439/2014F77

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1317–1327
DOI: 10.15439/2016F469
ACSIS, Vol. 8. ISSN 2300-5963

PEQUAL - E-commerce websites quality evaluation methodology

Jarosław Wątróbski
West Pomeranian University of Technology in Szczecin, Żołnierska 49, 71-210 Szczecin, Poland
Email: jwatrobski@wi.zut.edu.pl

Paweł Ziemba
The Jacob of Paradyż University of Applied Science in Gorzów Wielkopolski, Chopina 52, 66-400 Gorzów Wielkopolski, Poland
Email: pziemba@pwsz.pl

Jarosław Jankowski
Department of Computational Intelligence, Wrocław University of Technology, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland
Email: jjankowski@wi.zut.edu.pl

Waldemar Wolski
University of Szczecin, Mickiewicza 64, 71-101 Szczecin, Poland
Email: wwolski@wneiz.pl

Abstract— **Website quality evaluation is an important research task. Evolution and a growing set of available methods are observed. The article presents the authors' evaluation methodology of the quality of websites named PEQUAL. The formal foundation of the proposed methodology is the broadening of the classical EQUAL method with aspects of preference modelling and evaluation aggregation used in Multi-Criteria Decision Analysis (MCDA). Its empirical verification has been carried out for top e-commerce websites. The conducted research has revealed significant practical possibilities of analysis and interpretation of obtained final rankings.**

I. Introduction

ELECTRONIC commerce is one of the most prominent areas of e-business with increasing sales year by year and estimated 28.3 trillion dollars of worldwide retail sales in 2018 [1]. Estimation of digital buyers shows that 47.3 percent of global Internet users will purchase products online in 2018 what creates an increasing interest in this area [2]. E-commerce is dependent on the development of new technologies and with the growth of infrastructure, improvements of hardware and software accessibility and has changed its character since first applications [3]. The dynamic development of online sales platforms [4] and dedicated user interfaces [5] together with supporting technologies like personalization engines [6], recommending systems [7], online payment systems [8] and online marketing systems dedicated to electronic commerce [9] is observed.

The constant growth of a number of Internet stores intensifies competition between entities that offer goods and services online [10]. To improve results and maximize profits entrepreneurs use sophisticated analytic software [11], web mining techniques [12] or conversion maximization systems [13]. Together with the market growth more and more important is identification of factors affecting the performance of electronic commerce systems and customer loyalty [14]. Key elements of strategies are based on building trust [15], improving the quality of the systems [16], levels of security and privacy [17], their accessibility [18], development of international versions [19], solving cultural issues [20] and implementing new features towards consumer satisfaction and web usability [21].

For studying the quality and usability of e-commerce services and similarly for studying various types of Internet services, different types of methods based on the identification of key factors [76] influencing the perception of a given service by users are employed i.e. [22]. The methods differ especially in terms of assessment criteria and theoretical foundations on which they are based [74] [75]. Since evaluation of websites is a multiple-criteria problem, in the literature one can see attempts of using Multi-Criteria Decision Analysis (MCDA) methods for evaluating websites.

The objective of this article is to construct a quality assessment model of the most popular world e-commerce websites. An attempt to combine selected classical methods for evaluating the quality of websites (e.g. eQual) with a formal background used in the MCDA methodology constitutes a methodological research basis, which at the same time is the authors' contribution. However, it is assumed that additional application of the MCDA methodology will make it possible to carry out a wide analysis and verification of, obtained in the research, website rankings, and of users' preferences. This issue is of great importance and website quality evaluation methods used nowadays allow conducting this type of analyses only to a limited extent. Paper is organized as follows: Section II includes literature review, Section III presents methodological framework of proposed approach, Section IV presents results from the empirical study with conclusions within the Section V.

II. Literature review

A. Website evaluation methods

Website evaluation methods described in the literature employ different quality models, consequently, they differ in criteria used as well their quantity and structure [23]. In order to obtain an opinion on websites, they most often use questionnaires, and grades are expressed on an n-degree Likert scale [24]. Among website quality evaluation methods one can distinguish presented with references and key characteristics in the Table I : eQual, Web Portal Site Quality, Ahn method, SiteQual, Website Evaluation Questionnaire, Website Quality Model, E-S-QUAL and E-RecS-Qual,WAES.

The eQual method was constructed on the basis of Quality Function Deployment which is a structured process ensuring means of identification and providing users' opinion on the quality of a product on subsequent stages of its manufacturing process [25]. The eQual method was successfully used to evaluate: e-commerce [26], e-government [25] [27] [28], university [29] and WAP [30] websites.

Web Portal Site Quality came into existence on the basis of a Technology Acceptance Model. The TAM is to explain the influence of perceiving, by the user, information system characteristics on his or her acceptance of the given system. It is based on two quality dimensions, that is, perceived usefulness and perceived ease of use [31]. The Model of Information Systems Success by DeLone and McLean includes information quality and system quality [32][33]. The WPSQ method is used in evaluating portals delivering broadly defined information and services [34].

The Ahn method, similarly to Web Portal Site Quality, was devised with the use of Technology Acceptance Model [35]. The first version of the Ahn method was to study the influence of trust to bank websites on the acceptance by users [36]. When working on the method, the original TAM model was extended with subsequent elements which were important from the perspective of the Internet: information quality, system quality and service quality. These elements were borrowed from an extended Model of Information Systems Success of DeLone and McLean [37][38]. Also, quality characteristics regarding trade: the quality of a product and its delivery were added [39].

The SiteQual method [40] came into being as a combination of the SERVQUAL [41] and Data Quality [42] models. The SERVQUAL model was to reflect service quality, whereas Data Quality was to be responsible for information quality. This model was constructed on the basis of questionnaires concerning music e-commerce websites [43].

When preparing the Website Evaluation Questionnaire method, criteria used in the Website User Satisfaction (WUS) model were used [44]. As in WUS, in every characteristic there is one negative criterion, which is used to verify reliability evaluation [45]. This method came into existence in order to examine e-government websites, but it can also be employed to assess other types of websites which are to provide their users with knowledge and information [46].

The E-S-QUAL and E-RecS-Qual methods stem from the SERVQUAL method used for studying and evaluating service quality [47]. They are a result of adjusting the SERVQUAL scale to the needs of service quality assessment on the Internet. Here, some evaluation criteria in the SERVQUAL model were kept and new criteria essential for determining e-service quality were introduced. The E-S-QUAL method contains the core of the e-SERVQUAL scale, that is criteria perceived by customers who do not have questions and problems related to e-services. On the other hand, the E-RecS-QUAL method comprises additional criteria which are vital when the user encounters problems when using services. These methods were used to evaluate service quality on bank websites [48] as well as e-commerce [49] websites.

While preparing the Website Quality Model method [50], Kano's quality model was used, in which there are defined three levels of customers' expectations with regard to the quality of a product or a service: basic, performance, and exciting [51]. The evaluation of news websites, among other things, CNN.com [52], was carried out by means of this method.

The WAES (Website Attribute Evaluation System) method is designed for assessing office and administration websites. It consists of two groups of characteristics describing transparency and interactivity of a website. An expert's evaluation on a binary scale is employed in the method [53].

In Table I one can see characterized individual quality evaluation methods of websites. For methods using questionnaires it is assumed that the number of users evaluating a website should at least amount to 30 [54].

The most interesting method, out of all analyzed ones, seems to be eQual, which is characterized by the highest formalization level and which in many cases proved to be highly universal. The method is based on 22 criteria in the form of questionnaire questions. When evaluating, a Linkert scale, which ranges from 1 to 7, is used. Weights of individual criteria are determined in the same way. Apart from criterial evaluation, respondents also provide overall evaluation of a website. On the basis of this assessment, the reliability of partial opinions of every user is verified [27]. When a collection of questionnaire results has been gathered, an analysis of the questionnaires is conducted with regard to reliability and internal cohesion. To determine the reliability of results of a questionnaire in the eQual method, Cronbach's alpha is employed. It is assumed that the reliability of results is appropriate, if the value of coefficient alpha amounts to at least 0.6 [28]. In the method the result of evaluation is the European Quality of Government Index (EQI) calculated on the basis of the formulas (1), (2), (3) and (4):

$$EQI = \sum_{k=1}^{m} EQI_k / m \tag{1}$$

$$EQI_k = \left(Score_k / Max_k\right) \cdot 100\% \tag{2}$$

$$Score_k = \sum_{i=1}^{n} (o_i(k) \cdot w_i(k)) / n \tag{3}$$

$$Max_k = \sum_{i=1}^{n} (7 \cdot w_i(k)) / n \tag{4}$$

where: m – the number of criteria, n – the number of polled users, $o_i(k)$ – the evaluation of a website with regard to the n-th criterion, given by the i-th user, $w_i(k)$ – the weight of the k-th criterion given by the i-th user.

The problem related to a practical use of the method is to gain weights of criteria by means of questionnaires, because explicit declaration of users' preference may generate errors in the research [55]. This is also confirmed by the authors' research, in which it was demonstrated that weights of crite-

TABLE I.
CHARACTERISTICS OF SELECTED METHODS OF WEBSITE QUALITY ASSESSMENT

Method	Application	No of criteria	Method determining weights of criteria	Assessment scale	Method of examining websites	No of evaluators	Theoretical basis of method	Verification of solution	Reference
eQual	e-commerce, e-government, university websites, WAP websites	22	Questionnaires	1-7	Questionnaires	min. 30	Quality Function Deployment	Consistency reliability of questionnaires (Cronbach's Alpha)	[26], [25], [29], [27], [28], [30]
Ahn	e-banking, e-commerce	54	-	1-7	Questionnaires	min. 30	Technology Acceptance Model, Model of Information Systems Success	Consistency reliability of questionnaires (Cronbach's Alpha)	[39]
SiteQual	e-commerce	28	-	1-9	Questionnaires	min. 30	SERVQUAL, Data Quality	Consistency reliability of questionnaires (Cronbach's Alpha)	[40], [43]
WEQ	e-government	18+8 (negative)	-	1-5	Questionnaires	min. 30	Website User Satisfaction	Negative criteria	[45], [46]
WPSQ	information services	19	-	1-5	Questionnaires	min. 30	Technology Acceptance Model, Model of Information Systems Success	Complex reliability tests (i.a. convergence evaluation, discriminant analysis)	[34]
WQM	information services	32	Questionnaires	1-3	-	-	Kano quality model (levels of customers' expectations)	-	[52], [51]
E-S-QUAL/ RecS-Qual	e-banking, e-commerce	22+11	-	1-5	Questionnaires	min. 30	SERVQUAL	-	[49], [48]
WAES	e-government	40	-	0-1	Expert evaluation	min. 1	-	-	[53]

ria received by means of questionnaires lead to incorrect decision solutions [56][57].

B. Evaluation of websites with the use of MCDA methods

Apart from "classical" methods, discussed in part A, in the literature there are also attempts at employing MCDA methods for evaluation. It is justified since assessment of websites is a multi-criteria problem, in which one needs to take into consideration many dimensions of quality [58]. For instance, Lee and Kozar [59] used the AHP method to evaluate e-tourist and e-commerce websites. Chmielarz widely uses his original scoring method to asses a wide range of websites, i. a. e-commerce as well as e-banking [60][61][62]. Sun and Lin [63] evaluated e-commerce websites with the use of the fuzzy TOPSIS method. Del Vasto-Terrientes et al. [64] evaluated tourist destination websites by means of a new ELECTRE-III-H method. Furthermore, in the works of Lin [65] as well as Kong and Liu [66] in the fuzzy AHP method was used to determine the significance of quality evaluation criteria of e-learning and e-commerce websites. To assess websites, hybrids of various MCDA methods are also used. In the paper by Bilsel et al. [67] determining the weights of criteria was conducted by means of the AHP method, whereas a ranking of hospital websites was constructed with the use of the fuzzy Promethee method. Similarly, Kaya [68] employed the fuzzy AHP method to define weights of criteria, and used the fuzzy TOPSIS method to construct an e-commerce website ranking. A combination of MCDA methods was also used by Huang et al. [69], where solutions were compared with the use of, among other things, Simple Additive Weighting, Multiplicative Exponent Weighting, TOPSIS, concordance and discordance analysis methods. Weights of criteria in above-mentioned were determined by means of the OWA method.

The analysis of application of MCDA methods in website evaluation indicates that most of them used questionnaires to collect assessments of websites. As for determining weights of criteria, pairwise comparison matrices and the AHP method are most often used for this purpose. Since a significant number of such comparisons might be problematic, a limited number of criteria are usually used. It should be emphasized that for constructing a model of criteria only a few papers used theoretical bases identifying the need for presenting both specific quality measures and criteria. Moreover, only in some papers the sensitivity/robustness analysis of results were carried out. However, applying MCDA methods to evaluate websites has a greater potential than just constructing a ranking. This can be proved by a model of a decision process defined by Guitouni [70] wherein critical steps are exploitation and recommendation stages. On the operation stage, one can conduct the analysis of an obtained solution, such as examining its stability [77] [78] or the analysis of decision-makers' preference.

III. PEQUAL Methodological Framework

A. Selection of an MCDA method for evaluating websites

Every decision problem can be attributed to the problematics the decision problem deals with. The problematics result from the aim which is expected from the decision process [79]. In the problematics of description (P.δ), preparing a description of potential actions and the identification of a criterion or a family of criteria pose a problem. In the problematics of choice (P.α), supporting the decision-maker is concentrated on selecting a small number of "good" variants. The problematics of sorting (P.β) is concentrated on attributing a variant to one of classes available. Finally, in the problematics of ranking (P.γ), a ranking of decision variants according to defined criteria is prepared [80].

The MCDA method, which is used in evaluating websites, should especially take into consideration indifference and preference relations, which will make it possible to differentiate the quality of evaluated websites. Moreover, it should not allow an indifference relation to appear, since it is essential that the website ranking is total. Taking into account acceptable compensation criteria, it is reasonable to assume that certain website elements can convince users to use it, even though in some respects it falls short of expectations. Therefore, compensation of low-marked criteria by high-marked ones seems to be legitimate. Measuring data, on which the method will work, cannot be determined as reliable, since these are subjective users' opinions expressed in questionnaire on a quantitative scale. Nevertheless, such data unreliability may be expressed by defining a proper value of an indifference threshold. The problematics considered by individual MCDA methods is of importance, because a method ought to consider, first of all, the problematics of a ranking. It allows putting websites in order according to their synthesized quality, expressed on a quantitative scale. What is more, one can consider methods comprising also the problematics of description, what allows analyzing the obtained solution in a broader way. For the reason that quality is assessed by many users, a method should also offer a group evaluation mechanism. The analysis of characteristics and abilities of individual MCDA methods [71][72][73] with relation to the requirements discussed points out to the fact that the Promethee II [88] method along with its group development, i.e. Promethee GDSS, can be used in evaluating websites.

B. Framework of website evaluation

The authors' methodology of website quality evaluation named PEQUAL (Promethee - eQual) is based on the eQual method, which has its foundations in Quality Function Deployment. To do empirical research at first, questionnaires were collected from 41 users. In the research sample, there were computer literate users who are experienced in doing the shopping online. All of them evaluated 10 e-commerce websites: Alibaba, Amazon, Apple, BestBuy, eBay, Macy's, Rakuten, Staples, Target, and Walmart. The reason for selecting the e-commerce websites was the result of analysis of valid rankings of top e-commerce websites presented, among other things, in [81], [82], [83], [84], [85]. Thus, 410 questionnaires were collected which then were verified in terms of consistency reliability and Cronbach's alfa was determined. Questionnaire evaluation was conducted with the use of criteria and an evaluation scale of the eQual method and the results of the questionnaires were aggregated with the use of the Promethee method. Also, on the basis of this method, the broad analysis of the obtained solution was carried out. The research took into consideration two scenarios of aggregation of partial evaluations in a overall ranking. The input data had been obtained in the questionnaires. In the first scenario, partial evaluations were averaged, and next, the aggregation of mean criterial evaluations into a overall evaluation, with the use of the Promethee II method, was conducted. The second scenario consisted in determining individual rankings by means of the Promethee II meth-

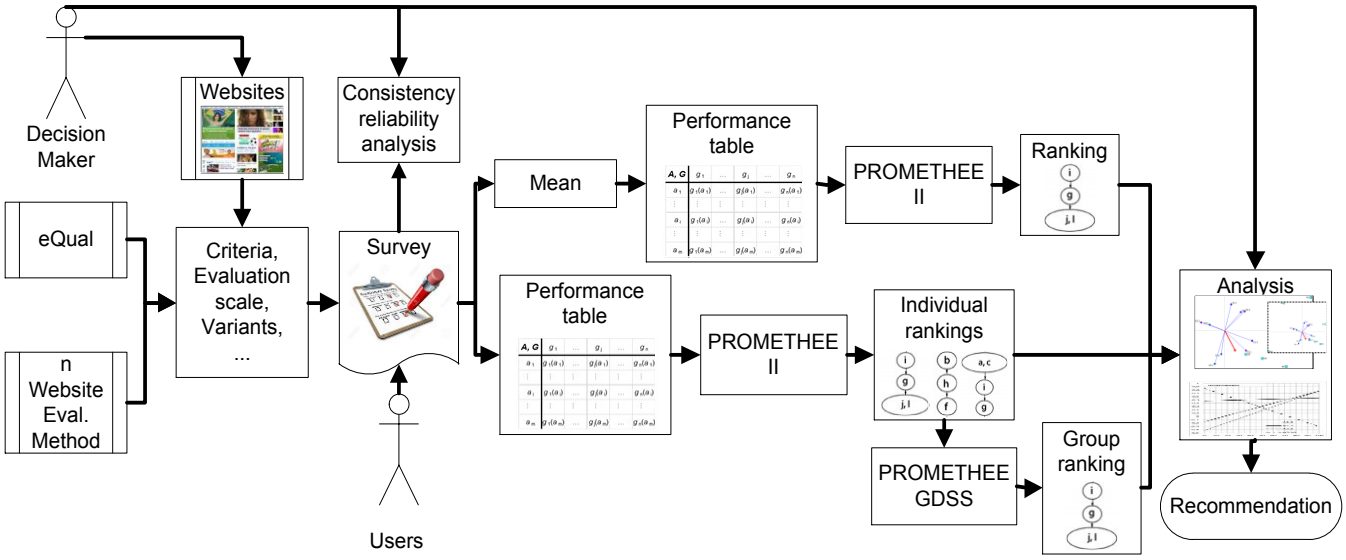


Fig. 1 PEQUAL methodological framework of website evaluation

od on the basis of partial evaluations and then aggregating individual rankings in a group ranking by means of the method Promethee GDSS. After generating rankings, the analysis of the solution obtained was carried out in every scenario with the use of the GAIA method and the analysis of ranking robustness to the changes of weights of criteria, which are an integral part of the Promethee method. On the grounds of the transparency of the conducted analysis, it was assumed that the weights of all criteria are equal. Moreover, it was assumed that the partial evaluations obtained in the questionnaires can be characterized by some degree of uncertainty. Therefore, an application of various indifference variants was considered, and consequently the influence of minor errors in partial evaluations on the obtained results was eliminated. The presented practical approach is depicted in Figure 1. What is more, at first, the aggregation of questionnaire results into a final evaluation with the use of the eQual method was carried out. This was aimed at conducting a comparative analysis of obtained results with reference to the results of eQual.

IV. Results

A. Empirical research

After gathering the results of the questionnaires, their consistency reliability analysis was carried out. The value of Cronbach's alpha obtained in the research amounted to 0.95. All scores of the consistency reliability analysis, including the value of Cronbach's alpha for individual groups of criteria, are presented in Table II. The values of Cronbach's alpha indicate high reliability of the conducted questionnaire, since it exceeds the boundary value of 0.6 [28].

Next, in accordance with the eQual method a overall value of eQual Index and values obtained for individual criteria and their groups were determined. The scores of the total value and groups of criteria are depicted in Table III.

In another research the scores of questionnaires were averaged and calculations were made with the use of the Promethee II method. Average criterion evaluations of questionnaire scores, which constitute a performance table, are depicted in Table IV (average values are in accordance with values used for the eQual method). In the Promethee II method for each criterion a preference model with a preference V-shape function, for which an indifference threshold q=0 and a preference threshold p=7, was used. A preference direction was maximized. The selected preference model was assumed in order to reflect, as accurately as possible, the model used in the eQual method. A website ranking obtained according to the Promethee II method, with the use of the given preference model, is presented in Table V. The sequence of variants in the obtained ranking is in accordance with the sequence in the eQual method.

TABLE II.
CRONBACH'S ALPHA SCORES

Cluster of criteria	Group of criteria	Criterion	α if item deleted	α for group of criteria
Usability	Usability	C1	0.9454	0.9379
		C2	0.9453	
		C3	0.9449	
		C4	0.9453	
	Site design	C5	0.9461	0.8722
		C6	0.9454	
		C7	0.9459	
		C8	0.9451	
Information quality	Information quality	C9	0.9455	0.8854
		C10	0.9455	
		C11	0.9521	
		C12	0.9452	
		C13	0.9450	
		C14	0.9456	
		C15	0.9449	
Service interaction	Trust	C16	0.9447	0.9038
		C17	0.9445	
		C18	0.9455	
		C22	0.9464	
	Empathy	C19	0.9465	0.767
		C20	0.9473	
		C21	0.9472	

B. Graphical analysis of Promethee solution

After determining the ranking, its analysis, based on the GAIA methodology, was done. Figures 2-4 depict the scores of this analysis separate for: clusters (Figure 4), groups (Figure 3) and individual criteria (Figure 2).

TABLE III.
ASSESSMENT RESULTS OF WEBISITES ACCORDING TO EQUAL METHOD

Website	Evaluation Quality Index									
	Alibaba	Amazon	Apple	BestBuy	eBay	Macy's	Rakuten	Staples	Target	Walmart
Usability	70.30%	78.66%	79.62%	70.56%	82.93%	71.17%	70.47%	69.08%	68.38%	72.65%
Site design	69.25%	75.09%	83.01%	62.46%	69.34%	68.21%	62.28%	65.68%	60.80%	67.16%
Information quality	71.33%	78.00%	77.15%	72.97%	78.55%	70.08%	69.34%	69.34%	71.33%	68.74%
Trust	68.12%	81.62%	85.37%	64.72%	79.88%	65.85%	61.85%	67.42%	60.54%	70.30%
Empathy	60.05%	70.85%	70.15%	55.87%	61.09%	57.26%	54.24%	55.98%	52.96%	59.23%
Overall	68.64%	77.27%	79.20%	66.79%	75.53%	67.42%	64.84%	66.46%	64.41%	68.15%
Rank	4	2	1	7	3	6	9	8	10	5

TABLE IV.
PERFORMANCE TABLE FOR PROMETHEE II BASED ON MEAN VALUES OF CRITERION EVALUATIONS

Group of criteria	Criterion	Website									
		Alibaba	**Amazon**	**Apple**	**BestBuy**	**eBay**	**Macy's**	**Rakuten**	**Staples**	**Target**	**Walmart**
Usability	C1	4.902	5.610	5.683	5.000	6.024	5.049	4.976	4.927	4.854	5.049
	C2	4.951	5.707	5.415	4.878	5.951	4.976	5.098	4.927	4.756	5.220
	C3	5.000	5.317	5.610	5.000	5.610	4.854	4.805	4.829	4.683	4.829
	C4	4.829	5.390	5.585	4.878	5.634	5.049	4.854	4.659	4.854	5.244
Site design	C5	4.829	5.024	5.976	4.341	4.683	4.707	4.268	4.512	4.220	4.927
	C6	5.098	5.488	6.024	4.561	5.341	5.049	4.707	4.927	4.707	4.805
	C7	4.829	5.366	5.829	4.537	4.878	4.756	4.439	4.732	4.415	4.805
	C8	4.634	5.146	5.415	4.049	4.512	4.585	4.024	4.220	3.683	4.268
Information quality	C9	5.000	5.537	5.049	5.073	5.634	4.780	4.805	4.780	4.756	4.537
	C10	4.902	5.537	5.902	5.098	5.683	4.902	5.024	4.805	4.902	4.805
	C11	5.585	5.268	5.488	5.122	5.415	5.512	5.488	5.146	5.561	5.317
	C12	4.951	5.463	5.341	5.268	5.537	4.902	4.732	4.854	5.049	4.610
	C13	4.732	5.537	5.561	5.244	5.512	4.878	4.756	4.707	4.902	4.976
	C14	4.854	5.488	5.171	5.098	5.220	4.634	4.659	4.854	5.024	4.488
	C15	4.927	5.390	5.293	4.854	5.488	4.732	4.512	4.829	4.756	4.951
Trust	C16	4.927	5.829	5.927	4.244	5.878	4.512	4.415	4.488	4.195	4.927
	C17	4.732	5.805	6.000	4.537	5.659	4.512	4.293	4.927	4.317	4.951
	C18	4.732	5.610	5.805	4.707	5.561	4.659	4.390	4.780	4.220	4.902
	C22	4.683	5.610	6.171	4.634	5.268	4.756	4.220	4.683	4.220	4.902
Empathy	C19	3.951	4.927	4.878	3.537	4.049	3.976	3.659	3.756	3.366	3.951
	C20	3.878	4.683	4.293	3.366	3.488	3.439	3.463	3.610	3.146	3.756
	C21	4.780	5.268	5.561	4.829	5.293	4.610	4.268	4.390	4.610	4.732

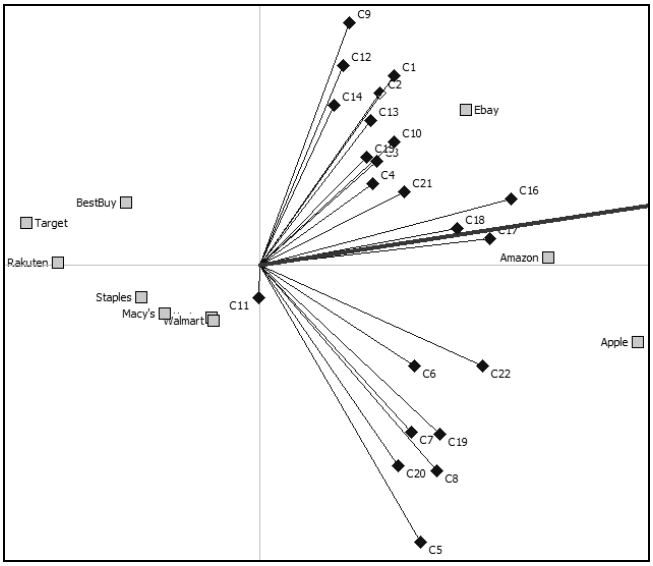


Fig. 2 GAIA analysis for criteria

The analysis of Figure 2 demonstrates that almost all criteria, except C11, support three leading variants in the ranking, i.e. Apple, Amazon and eBay. This observation is confirmed by a detailed analysis of numbers contained in Table IV. Moreover, a criterion C11 is in conflict with C9, C12, C14, C1, C2, C13 and also partially with subsequent criteria placed in the first quarter of the system of coordinates. It means that variants which are highly evaluated with regard to the criterion C11 get lower evaluation in terms of other criteria mentioned. Furthermore, the length of C11 vector points out that this criterion has the least influence on the final website ranking.

An analysis of Figure 3 allows finding out that users evaluate Usability and Information Quality of individual websites in a similar way. In other words, if an examined website gets high marks for criteria in the Usability group, it is usually highly marked with regard to criteria in the Information Quality group. However, evaluations of criteria in the Usability and Site Design groups are independent of each other. This piece of information is important, because criteria in the groups belong to one cluster of criteria, therefore, their evaluations should usually be similar to one another. Similarly, criteria evaluations of Empathy and Trust, which also belong to one cluster, are independent of each other. Moreover, one can state that the most significant influence on the final ranking have criteria belonging to the Site Design and Usability groups, since their vectors are longest. Furthermore, the final ranking of websites most strongly overlaps with the criterial evaluations of the Trust group.

The presentation of clusters of criteria (Figure 4) on the GAIA plain indicates that evaluations of variants with regard to criteria belonging to the Usability and Service Interaction clusters and they are independent of the Information Quality cluster. It may seem contradictory to the conclusion drawn

when analyzing Figure 3 which expresses the similarity of evaluations with regard to the Usability and Information Quality criteria groups. However, one needs to bear in mind that the Usability cluster contains the Usability and Site Design criteria groups. When considering the impact of individual criteria clusters on the final ranking, it should be noted that this ranking is, to the highest degree, dependent on the criteria belonging to the Information Quality and later Service Interaction clusters. The criteria belonging to the Usability cluster have the lowest influence on the ranking.

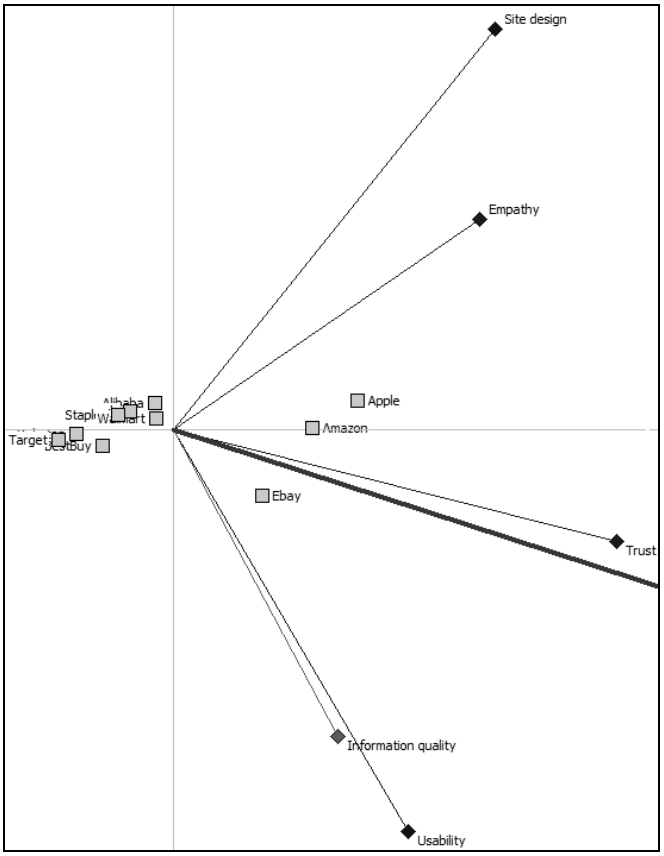


Fig. 3 GAIA analysis for groups of criteria

C. Robustness analysis of solution

Apart from the GAIA analysis, also, the robustness analysis, taking into consideration changes in weights of criteria, of the ranking was carried out. Figure 5 depicts, one by one, the scores of the analysis for weight changes of subsequent criteria clusters, i.e. Usability, Information Quality and Service Interaction. Also, Figure 5 presents the robustness analysis of the ranking for changes of weights of criteria belonging to the Usability cluster. It must be explained that the weight changes regarded all clusters, for instance, when a weight of Usability was 0, the criteria in other clusters obtained a weight of 7.14%, whereas when a weight of Usability was 100%, all of its criteria obtained a weight of 12.5%. Analogically, weights for the robustness analysis of the Information Quality and Service Interaction clusters were determined. The results of the robustness analysis indicate that three top positions in the ranking are very stable, because only increasing weights of criteria in the Information Quality cluster above 80% (that is over 11.5 for each criteria of the cluster) may cause changes on these positions. Therefore, one can assume with the high level of probability that, independent of weights of criteria, the obtained ranking is correct.

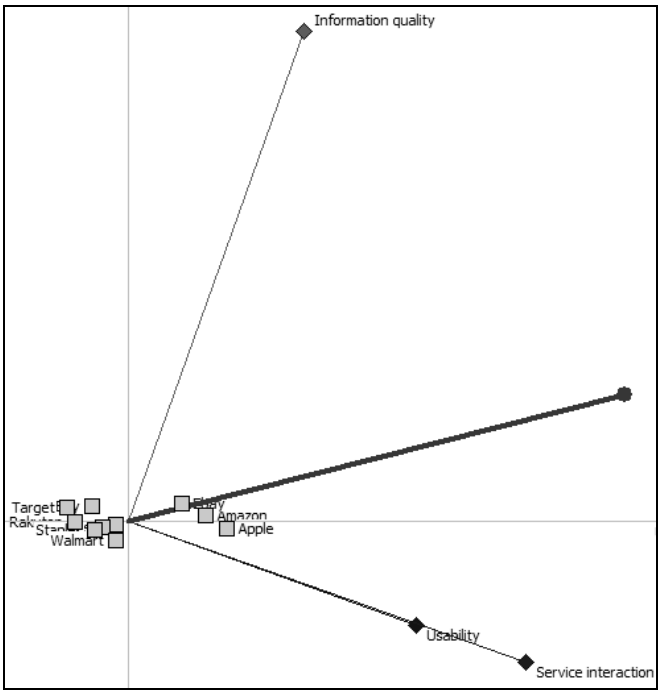


Fig. 4 GAIA analysis for clusters of criteria

D. Uncertainty analysis

The next step in the conducted analysis was to verify the influence of uncertainty of partial evaluations on the sequence of variants in the ranking. Therefore, a new ranking of variants was determined on the basis of a modified preference model, in which a preference function V-shape with an indifference area, where the indifference q=1 and preference p=7 thresholds were used, was applied. Using the indifference threshold was to eliminate the influence of potential mistakes in users' evaluation, which consists in considering one website slightly better than another one. It should be noted that the threshold q=1 for averaged values provides a

TABLE V.
RANKING OF WEBSITES BASED ON PROMETHEE II AND AVERAGED CRITERIA EVALUATIONS

Website	Alibaba	Amazon	Apple	BestBuy	eBay	Macy's	Rakuten	Staples	Target	Walmart
ϕ_{net}	-0.0137	0.0822	0.1037	-0.0343	0.0629	-0.0272	-0.0559	-0.0380	-0.0607	-0.0191
Rank	4	2	1	7	3	6	9	8	10	5

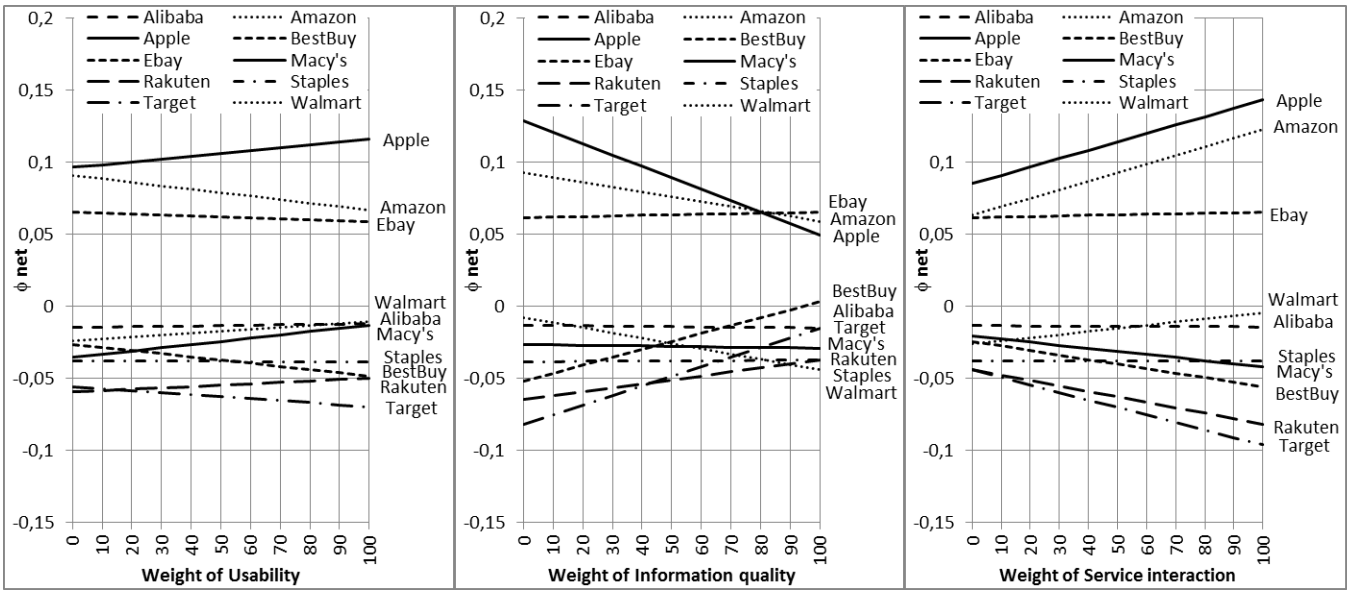


Fig. 5 Robustness analysis of criteria clusters

significant error of margin, and with regard to many criteria almost all variants are considered indifferent. However, the ranking obtained has variant shifts only on further positions. To be more specific, there was a change of websites on positions 4 and 5 (Alibaba a Walmart) as well as 7 and 8 (Staples and BestBuy). Therefore, it can be assumed that the basic ranking which was obtained with the use of the Promethee II method is reliable. The ranking obtained with the used of the described preference model is depicted in Table VI.

E. Comparison of averaged ranking with group ranking

Another part of the conducted research consisted in conducting, by means of the Promethee II method, an aggregation of partial evaluations for individual questionnaires and determining individual rankings. They were generated next to the preference model, which had been used for averaged evaluations, i.e. with the preference function V-shape and thresholds q=0, p=7. Later, the individual rankings were aggregated into a group ranking with the use of the Promethee GDSS method. The ranking is presented in Table VII and its analysis indicates that the obtained sequence of websites is the same as in the ranking obtained before, which was based on averaged evaluations, presented in Table V. The values of evaluations ϕ net are also similar.

The next step was to conduct a GAIA analysis for the group ranking. For the reason of clarity, Figure 6 depicts the GAIA plane for 10 respondents.

The projection of decision-makers' preferences on the plane shows that everybody, except DM6, supports to some extent five best websites in the ranking. However, it should be noted that evaluations of users DM7 and DM8 are contradictory, similarly to users DM4 and DM6. Moreover, the highest influence on the final ranking, out of 10 presented users, have DM8, DM9 and DM10, whose vectors are longest. As far as the respondents' individual rankings are concerned, their analysis was carried out analogically to the analysis conducted for averaged evaluations.

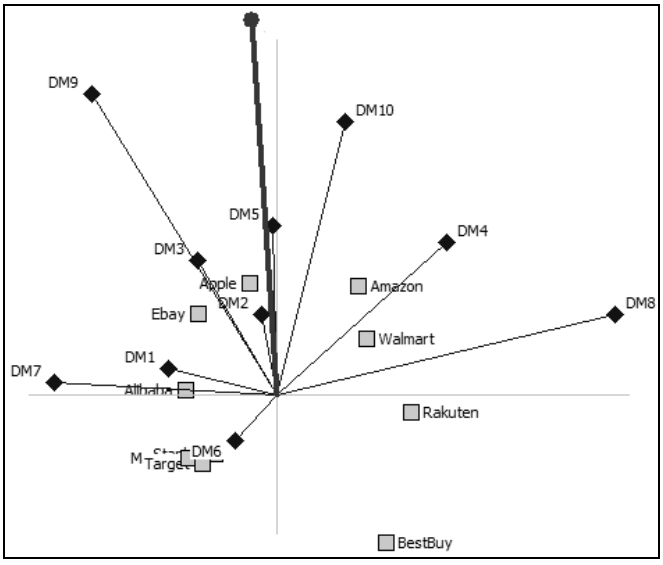


Fig. 6 GAIA analysis for clusters of criteria

TABLE VI.
RANKING OF WEBSITES BASED ON PROMETHEE II AND AVERAGED CRITERIA EVALUATIONS AND THE THRESHOLD Q=1

Website	Alibaba	Amazon	Apple	BestBuy	eBay	Macy's	Rakuten	Staples	Target	Walmart
ϕ_{net}	-0.001	0.0076	0.0175	-0.0054	0.0038	-0.0027	-0.0069	-0.0028	-0.0093	-0.0008
Rank	5	2	1	8	3	6	9	7	10	4

TABLE VII.
RANKING OF WEBSITES BASED ON PROMETHEE GDSS

Website	Alibaba	Amazon	Apple	BestBuy	eBay	Macy's	Rakuten	Staples	Target	Walmart
ϕ_{net} **GDSS**	-0.0078	0.0455	0.0574	-0.0192	0.0347	-0.0134	-0.0313	-0.0213	-0.0339	-0.0108
Rank GDSS	4	2	1	7	3	6	9	8	10	5

V. CONCLUSION

In the proposed approach the multi-stage construction of the model was realized with regard to the criteria taken from the eQual method with the use of the Promethee method (PEQUAL). It extends earlier approaches by introducing MCDA based multi stage evaluation and analyses. In the article, 10 most popular world e-commerce websites were evaluated. On the basis of the presented research, one can state that e-commerce websites most highly valued by users are: Apple, Amazon and eBay. The conclusions were confirmed by verifying the obtained ranking with the use of the analysis of robustness to changes of weights of criteria and examining the influence of evaluations on the final ranking.

Furthermore, the use of the Promethee GDSS method and the GAIA analysis, which is an integral part of the Promethee method, made it possible to indicate users' individual preferences. Also, the GAIA analysis allowed examining mutual dependences between individual groups and clusters of criteria on the basis of graphic data. The interpretation of the GAIA plane is less time-consuming and easier than the analysis of number values of evaluations, and the conclusions drawn on its basis are equally essential [86] [87].

The research framework of the quality of websites presented in the article can be the basis for their evaluation along with the correctness verification of obtained evaluations and preferences of the respondents. As it has been demonstrated in the presented research, this solution is functionally richer than classical MCDA-based methods of website evaluation methods which have been used in the literature to date.

ACKNOWLEDGMENT

This work was partially supported by the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no. 316097 [ENGINE].

REFERENCES

[1] eMarketer, "Retail ecommerce sales worldwide from 2013 to 2018," http://www.emarketer.com/Article.aspx?R=1011765, 2014. date accessed 25.03.2016

[2] Statista, "Digital buyer penetration worldwide from 2011 to 2018," http://www.statista.com/statistics/261676/digital-buyer-penetration-worldwide, 2014. date accessed 25.03.2016

[3] A. Chaudhury and J.P. Kuilboer, *E-business and E-commerce Infrastructure: Technologies Supporting the E-business Initiative*. McGraw-Hill Higher Education, 2001.

[4] P. De, Y. Hu, and M.S. Rahman, "Technology usage and online sales: An empirical study," *Management Science*, vol. 56, no. 11, pp. 1930-1945, 2010.

[5] Y. Purwati, "Standard features of e-commerce user interface for the web," *Journal of Arts, Science & Commerce*, vol. 2, pp. 77-87, 2011.

[6] F. Abbattista, M. Degemmis, O. Licchelli, P. Lops, G. Semeraro, and F. Zambetta, "Improving the usability of an e-commerce website through personalization," in *Proc. 2nd International Conference on Adaptive Hypermedia and Adaptive Web Based Systems, RPeC'02*, Malaga, Spain, 2002, pp. 20-29.

[7] B. Xiao and I. Benbasat, "E-commerce product recommendation agents: Use, characteristics, and impact," *Mis Quarterly*, vol. 31, no. 1, pp. 137-209, 2007.

[8] S. Özkan, G. Bindusara, and R. Hackney, "Facilitating the adoption of e-payment systems: theoretical constructs and empirical analysis," *Journal of enterprise information management*, vol. 23, no. 3, pp. 305-325, 2010.

[9] R.R. Yager, "Targeted e-commerce marketing using fuzzy intelligent agents," *IEEE Intelligent Systems and their Applications*, vol. 15, no. 6, 42-45, 2000.

[10] C. Forman, A. Ghose, and A. Goldfarb, "Competition between local and electronic markets: How the benefit of buying online depends on where you live," *Management Science*, vol. 55, no. 1, pp. 47-57, 2009.

[11] L. Hasan, A. Morris, and S. Probets, "Using Google Analytics to evaluate the usability of e-commerce sites," in *Human centered design*, Berlin Heidelberg: Springer, pp. 697-706, 2009.

[12] Y. Jyoti and B. Mallick, "Web Mining: Characteristics and Application in Ecommerce," Intl. *J. of IJECSE*, vol.1, no. 4, 2011.

[13] X. Wu and A. Bolivar, "Predicting the conversion probability for items on C2C ecommerce sites," in *Proc. of the 18th ACM conference on Information and knowledge management*, ACM, 2009, pp. 1377-1386.

[14] S.S. Srinivasan, R. Anderson, and K. Ponnavolu, "Customer loyalty in e-commerce: an exploration of its antecedents and consequences," *Journal of Retailing*, vol. 78, no. 1, pp. 41-50, 2002.

[15] D.W. Manchala, "E-commerce trust metrics and models," *IEEE Internet Computing*, vol. 4, no. 2, pp. 36-44, 2000.

[16] M. Cao, Q. Zhang, and J. Seydel, "B2C e-commerce web site quality: an empirical examination," *Industrial Management & Data Systems*, vol. 105, no. 5, pp. 645-661, 2005.

[17] A.K. Ghosh, *E-commerce Security and Privacy*, New York: Springer, 2012.

[18] O. Sohaib and K. Kang, "The Importance of Web Accessibility in Business to-Consumer (B2C) Websites," in *Proc. 22nd Australasian Software Engineering Conference (ASWEC 2013)*, 2002.

[19] T.R. Lituchy and R.A. Barra, "International issues of the design and usage of websites for e-commerce: Hotel and airline examples," *Journal of Engineering and Technology Management*, vol. 25, no. 1, pp. 93-111, 2008.

[20] O. Sohaib, "Usability and Cultural Issues in Global E-Commerce," *Journal of engineering and technology management*, vol. 25, no. 1, pp. 156-166, 2012.

[21] D. Belanche, L.V. Casaló, and M. Guinalíu, "Website usability, consumer satisfaction and the intention to use a website: The moderating effect of perceived risk," *Journal of retailing and consumer services*, vol. 19, no. 1, pp. 124-132, 2012.

[22] P. Ziemba, J. Jankowski, J. Wątróbski, W. Wolski, and J. Becker, "Integration of Domain Ontologies in the Repository of Website Evaluation Methods," *Annals of Computer Science and Information Systems*, vol. 5, pp. 1585-1595, 2015. http://dx.doi.org/10.15439/2015F297

[23] P. Ziemba, M. Piwowarski, J. Jankowski, and J. Wątróbski, "Method of Criteria Selection and Weights Calculation in the Process of Web Projects Evaluation," *Lecture Notes in Artificial Intelligence*, vol. 8733, pp. 684-693, 2014. http://dx.doi.org/10.1007/978-3-319-11289-3_69

[24] P. Ziemba, J. Wątróbski, J. Jankowski, and W. Wolski, "Construction and Restructuring of the Knowledge Repository of Website

Evaluation Methods," *Lecture Notes in Business Information Processing*, vol. 243, pp. 29-52, 2016. http://dx.doi.org/ 10.1007/978-3-319-30528-8_3

[25] S.J. Barnes and R. Vidgen, "Measuring Web site quality improvements: a case study of the forum on strategic management knowledge exchange," *Industrial Management & Data Systems*, vol. 103, no. 5, pp. 297-309, 2003.

[26] S.J. Barnes and R. Vidgen, "The eQual Approach to the Assessment of E-Commerce Quality: A Longitudinal Study of Internet Bookstories," in *Web Engineering: Principles and Techniques*, W. Suh, Ed. London: Idea Group Publishing, 2005, pp. 161-181.

[27] S.J. Barnes and R. Vidgen, "Data Triangulation in action: using comment analysis to refine web quality metrics," in *Proc. 13th European Conference on Information Systems*, 2005.

[28] S.J. Barnes and R. Vidgen, "Data triangulation and web quality metrics: A case study in e-government," *Information & Management*, vol. 43, no. 6, pp. 767-777, 2006.

[29] S.J. Barnes and R. Vidgen, "WebQual: An Exploration of Web-site Quality," in *Proc. 8th European Conference on Information Systems*, vol. 1, 2000, pp. 298–305.

[30] S.J. Barnes, K. Liu, and R. Vidgen, "Evaluating WAP News Sites: The WebQual/M Approach," in *Proc. 9th European Conference on Information Systems*, 2001.

[31] H.P. Shih, "Extended technology acceptance model of Internet utilization behavior," *Information & Management*, vol. 41, no. 6, pp. 719-729, 2004.

[32] W.H. DeLone and E.R. McLean, "Information Systems Success: The Quest for the Dependent Variable," *Information Systems Research*, vol. 3, no. 1, pp. 60-95, 1992.

[33] P.B. Seddon, "A Respecification and Extension of the DeLone nad McLean Model of IS Success," *Information Systems Research*, vol. 8, no. 3, pp. 240-253, 1997.

[34] Z. Yang, S. Cai, Z. Zhou, and N. Zhou, "Development and validation of an instrument to measure user perceived service quality of information presenting Web Portals," *Information & Management*, vol. 42, no. 4, pp. 575-589, 2005.

[35] T. Ahn, S. Ryu, and I. Han, "The impact of Web quality and playfulness on user acceptance of online retailing," *Information & Management*, vol. 44, no. 3, pp. 263-275, 2007.

[36] B. Suh and I. Han, "Effect of trust on customer acceptance of Internet banking," *Electronic Commerce Research and Applications*, vol. 1, no. 3-4, pp. 247-263, 2002.

[37] W.H. DeLone and E.R. McLean, "The DeLone and McLean Model of Information Systems Success: A Ten-Year Update," *Journal of Management Information Systems*, vol. 19, no. 4, pp. 9-30, 2003.

[38] S.M. Jafari, N.A. Ali, M. Sambasivan, and M.F. Said, "A Respecification and Extension of DeLone and McLean Model of IS Success in the Citizen-centric e-Governance," in *Proc. IEEE International Conference on Information Reuse and Integration*, 2011, pp. 342-346.

[39] T. Ahn, S. Ryu, and I. Han, "The impact of the online and offline features on the user acceptance of Internet shopping malls," *Electronic Commerce Research and Applications*, vol. 3, no. 4, pp. 405-420, 2004.

[40] H.W. Webb and L.A. Webb, "Business to consumer electronic commerce Website quality: integrating information and service dimensions," in *Proc. 7th Americas Conference on Information Systems*, 2001.

[41] G.J. Udo, K.K. Bagchi, and P.J. Kirs, "Using SERVQUAL to assess the quality of e-learning experience," *Computers in Human Behavior*, vol. 27, no. 3, pp. 1272-1283, 2011.

[42] R.Y. Wang and D.M. Strong, "Beyond Accuracy: What Data Quality Means to Data Consumers," *Journal of Management Information Systems*, vol. 12, no. 4, pp. 5-33, 1996.

[43] H.W. Webb and L.A. Webb, "SiteQual: an integrated measure of Web site quality," *Journal of Enterprise Information Management*, vol. 17, no. 6, pp. 430-440, 2004.

[44] S. Muylle, R. Moenaert, and M. Despontin, "The conceptualization and empirical validation of web site user satisfaction," *Information & Management*, vol. 41, no. 5, pp. 543-560, 2004.

[45] S. Elling, L. Lentz, and M. de Jong, "Website Evaluation Questionnaire: Development of a Research-Based Tool for Evaluating Informational Websites," *Lecture Notes in Computer Science*, vol. 4656, pp. 293-304, 2007.

[46] S. Elling, L. Lentz, M. de Jong, and H. van den Bergh, "Measuring the quality of governmental websites in a controlled versus an online setting with the 'Website Evaluation Questionnaire'," *Government Information Quarterly*, vol. 29, no. 3, pp. 383-393, 2012.

[47] A. Parasuraman, V.A. Zeithaml, and L.L. Berry, "SERVQUAL: A Multiple-item Scale for Measuring Consumer Perceptions of Service Quality," *Journal of Retailing*, vol. 64, no. 1, pp.12-40, 1988.

[48] S. Akinci, E. Atilgan-Inan, and S. Aksoy, "Re-assessment of E-S-Qual and E-RecS-Qual in a pure service setting," *Journal of Business Research*, vol. 63, no. 3, pp. 232-240, 2010.

[49] A. Parasuraman, V.A. Zeithaml, and A. Malhotra, "E-S-QUAL A Multiple-Item Scale for Assessing Electronic Service Quality," *Journal of Service Research*, vol. 7, no. 10, pp. 1-21, 2005.

[50] G.M. von Dran, P. Zhang, and R. Small, "Quality Websites: An Application of the Kano Model to Website Design," in *Proc. 5th Americas Conference on Information Systems*, 1999, pp. 898-900.

[51] P. Zhang and G. von Dran, "User Expectations and Rankings of Quality Factors in Different Web Site Domains," *International Journal of Electronic Commerce*, vol. 6, no. 2, pp. 9-33, 2002.

[52] P. Zhang and G. von Dran, "Expectations and Rankings of Website Quality Features: Results of Two Studies on User Perceptions," in *Proc. 34th Hawaii International Conference on System Sciences*, 2001.

[53] C.C. Demchak, C. Friis, and T.M. La Porte, "Webbing Governance: National Differences in Constructing the Face of Public Organizations," in *Handbook of Public Information Systems*, G.D. Garson, Ed. New York: Marcel Dekker, 2000, pp.179-196.

[54] A. Holzinger, "Usability Engineering Methods for Software Developers," *Communications of ACM*, vol. 48, no. 1, pp. 71-74, 2005.

[55] A. Zenebe, L. Zhou, F. Norcio, "User preferences discovery using fuzzy models," *Fuzzy Sets and Systems*, vol. 161, no. 23, pp. 3044-3063, 2010.

[56] P. Ziemba, J. Jankowski, J. Wątróbski, and M. Piwowarski, "Web Projects Evaluation Using the Method of Significant Website Assessment Criteria Detection," *Transactions on Computational Collective Intelligence*, no. 22, pp. 167-188, 2016. http://dx.doi.org/10.1007/978-3-662-49619-0_9

[57] P. Ziemba and M. Piwowarski, "Procedure of Reducing Website Assessment Criteria and User Preference Analyses," *Foundations of Computing and Decision Sciences*, vol. 36, no.3-4, pp. 315-325, 2011.

[58] S. Kim and L. Stoel, "Dimensional hierarchy of retail website quality," *Information & Management*, vol. 11, no. 2, pp. 109-117, 2004. http://dx.doi.org/10.1016/S0969-6989(03)00010-9

[59] Y. Lee and K.A. Kozar, "Investigating the effect of website quality on e-business success: An analytic hierarchy process (AHP) approach," *Decision Support Systems*, vol. 42, no. 3, pp. 1383-1401, 2006.

[60] W. Chmielarz and M. Zborowski, "The Application Of A Conversion Method In A Confrontational Pattern-Based Design Method Used For The Evaluation Of It Systems," *Annals of Computer Science and Information Systems*, vol. 2, pp. 1227-1234, 2014. http://dx.doi.org/10.15439/2014F198

[61] W. Chmielarz and M. Zborowski, "Comparative Analysis of Electronic Banking Websites in Selected Banks in Poland in 2014," *Annals of Computer Science and Information Systems*, vol. 5, pp. 1499-11504, 2015. http://dx.doi.org/10.15439/2015F43

[62] W. Chmielarz, "Evaluation of selected mobile applications stores from the user's perspective," *Online Journal of Applied Knowledge Management*, vol. 3, no. 1, pp. 21-36, 2015.

[63] C.C. Sun and G.T.R. Lin, "Using fuzzy TOPSIS method for evaluating the competitive advantages of shopping websites," *Expert Systems with Applications*, vol. 36, no. 9, pp. 11764-11771, 2009. http://dx.doi.org/10.1016/j.eswa.2009.04.017

[64] L. Del Vasto-Terrientes, A. Valls, R. Słowiński, and P. Zielniewicz, "ELECTRE-III-H: An outranking-based decision aiding method for hierarchically structured criteria," *Expert Systems with Applications*, vol. 42, no. 11, pp. 4910-4926, 2015. http://dx.doi.org/10.1016/j.eswa.2015.02.016

[65] H.F. Lin, "An application of fuzzy AHP for evaluating course website quality," *Computers & Education*, vol. 54, no. 4, pp. 977-888, 2010.

[66] F. Kong and H. Liu, "Applying fuzzy analytic hierarchy process to evaluate success factors of e-commerce," *International Journal of Information and System Sciences*, vol. 1, no. 3-4, pp. 406-412, 2005.

[67] R. U. Bilsel, G. Buyukozkan, and D. Ruan, "A Fuzzy Preference-Ranking Model for a Quality Evaluation of Hospital Web Sites," *International Journal of Intelligent Systems*, vol. 21, pp. 1181-1197, 2006.

[68] T. Kaya, "Multi-attribute Evaluation of Website Quality in E-business Using an Integrated Fuzzy AHPTOPSIS Methodology," *International Journal of Computational Intelligence Systems*, vol. 3, no. 3, pp. 301-314, 2010. http://dx.doi.org/10.1080/18756891.2010.9727701

[69] J. Huang, X. Jiang, and Q. Tang, "An e-commerce performance assessment model: Its development and an initial test on e-commerce applications in the retail sector of China," *Information & Management*, vol. 46, no. 2, pp. 100-108, 2009.

[70] A. Guitouni, J.M. Martel, and P. Vincke, "A Framework to Choose a Discrete Multicriterion Aggregation Procedure," *Defence Research Establishment Valcatier (DREV)*, 1998.

[71] J. Wątróbski and J. Jankowski, "Knowledge Management in MCDA Domain," *Annals of Computer Science and Information Systems*, vol. 5, pp. 1445-1450, 2015. http://dx.doi.org/10.15439/2015F295

[72] J. Wątróbski and J. Jankowski, "An Ontology-Based Knowledge Representation of MCDA Methods," *Lecture Notes in Artificial Intelligence*, vol. 9621, pp. 54-64, 2016. http://dx.doi.org/10.1007/978-3-662-49381-6_6

[73] J. Wątróbski and J. Jankowski, "Guideline for MCDA Method Selection in Production Management Area," in *New Frontiers in Information and Production Systems Modelling and Analysis*, Intellingent Systems Reference Library 98, Berlin Heidelberg: Springer, 2016, pp. 119-138. http://dx.doi.org/10.1007/978-3-319-23338-3_6

[74] J. Jankowski, J. Wątróbski, P. Ziemba, "Modelling the impact of visual components on verbal comunication in online advertising," in *Computational Collective Intelligence. ICCCI 2015, Part II. LNAI,* vol. 9330, Heidelberg: Springer, 2015, pp. 44-53.

[75] J. Jankowski, J. Watrobski, and M. Piwowarski, "Fuzzy Modeling of Digital Products Pricing in the Virtual Marketplace," in *Proceedings of 6th International Conference on Hybrid Artificial Intelligent Systems*, LNCS, vol. 6678. Heidelberg: Springer, 2011, pp. 338-346.

[76] J. Jankowski, K. Kolomvatsos, P. Kazienko, J. Wątróbski, "Fuzzy Modeling of User Behaviors and Virtual Goods Purchases in Social Networking Platforms," *Journal of Universal Computer Science*, vol. 22, no. 3, pp. 416-437, 2016.

[77] P. Ziemba, J. Wątróbski, J. Jankowski, and M. Piwowarski, "Research on the Properties of the AHP in the Environment of Inaccurate Expert Evaluations," in *Selected Issues in Experimental Economics*, K. Nermend and M. Łatuszyńska, Ed. Switzerland: Springer, 2016, pp. 227-243.

[78] P. Ziemba and J. Wątróbski, "Selected Issues of Rank Reversal Problem in ANP Method," in *Selected Issues in Experimental Economics*, K. Nermend and M. Łatuszyńska, Ed. Switzerland: Springer, 2016, pp. 203-225.

[79] A. Ishizaka and P. Nemery, *Multi-Criteria Decision Analysis. Methods and Software*. Chichester: Wiley, 2013.

[80] B. Roy, *Multicriteria Methodology for Decision Aiding*. Dordrecht: Springer, 1996.

[81] http://www.alexa.com/topsites/category;3/Business/E-Commerce. date accessed 25.04.2016

[82] http://www.mbaskool.com/fun-corner/top-brand-lists/13991-top-10-e-commerce-companies-in-the-world-2015.html?start=1. date accessed 25.04.2016

[83] http://www.dollarfry.com/worlds-top-10-ecommerce-sites-alexa-rank-basis/. date accessed 25.04.2016

[84] https://www.mitx.org/files/zmags-top100-web.pdf. date accessed 25.04.2016

[85] http://unctad.org/en/PublicationsLibrary/ier2015_en.pdf. date accessed 27.04.2016

[86] J. Jankowski, P. Ziemba, J. Wątróbski, P. Kazienko, "Towards the Tradeoff Between Online Marketing Resources Exploitation and the User Experience with the Use of Eye Tracking," *Lecture Notes in Artificial Intelligence*, ACIIDS 2016, vol. 9621, pp. 330-343, 2016. http://dx.doi.org/10.1007/978-3-662-49381-6_32

[87] J. Jankowski, J. Wątróbski, K. Witkowska, P. Ziemba, "Eye Tracking Based Experimental Evaluation of the Parameters of Online Content Affecting the Web User Behaviour," in *Selected Issues in Experimental Economics*, K. Nermend and M. Łatuszyńska, Ed. Switzerland: Springer, 2016, pp. 311-332.

[88] J. Wątróbski, P. Ziemba, and W. Wolski, "Methodological Aspects of Decision Support System for the Location of Renewable Energy Sources," *Annals of Computer Science and Information Systems*, vol. 5, pp. 1451-1459, 2015. http://dx.doi.org/ 10.15439/2015F294

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1329–1333
DOI: 10.15439/2016F112
ACSIS, Vol. 8. ISSN 2300-5963

Aspects of Mobility in e-Marketing from the Perspective of a Customer

Witold Chmielarz
University of Warsaw, Faculty of Management, in Warsaw
ul. Szturmowa 1/3, 02-678 Warsaw, Poland
Email: witold@chmielarz.eu

Marek Zborowski
University of Warsaw, Faculty of Management, in Warsaw
ul. Szturmowa 1/3, 02-678 Warsaw, Poland
Email: mzborowski@wz.uw.edu.pl

***Abstract*—The main aim of this article is to analyze the use of marketing mechanisms on the Internet. To meet this objective, the authors conducted a study limited to a selected group of individual users - students of the University of Warsaw. The paper presents the characteristics of the application of marketing tools as well as the users' opinion on the usefulness of these tools in broadly defined e-commerce. The article also presents the discussion and the findings of the present study.**

I. Introduction

THE MAIN objective of this paper is to analyze the possibilities of using the Internet on mobile and desktop devices for marketing purposes, under the circumstances of the dynamic development of both mobile devices and mobile applications running on them. This article aims to examine the situation where basic marketing tools are used by clients on the desktop (PC) or portable equipment (e.g. a laptop) as well as the devices combining the advantages of a phone and a computer, i.e. smartphones and tablets.

There are numerous definitions of e-marketing presented by scholars and academics. They are as follows: electronic marketing may be defined as activities of an organization which through applying new technologies, in particular the Internet, uses the potential of the market (Meng X., 2009). Internet marketing is a combination of communication efforts undertaken by organizations, operating both in the physical and virtual markets, to satisfy individual and collective needs in the electronic space, with the application of information technologies, especially the Internet, in order to make a profit (Chmielarz, 2007). It encompasses among other things: the study of the behavior of Internet users, the development of new products as well as e-promotions, e-distribution and e-service. In addition, the definition of e-marketing may be extended to include also the approach which states that it comprises all marketing activities aimed at fulfilling the operational goals via the Internet. From a theoretical point of view, it is a result of a combination of modern marketing, management and risk management theories as well as the application of modern communication technologies (Sun, 2011; Hasan, 2011).

Mobile marketing may be defined as a part of electronic marketing, one of the methods of direct marketing, practiced via mobile devices such as e.g.: cell phones, smartphones, tablets, PDA, MDA and notebooks. In m-marketing the messages with commercial or non-commercial content are communicated with the application of such technologies as SMS, MMS, NFC and others (Hovancakowa, 2011; Bernauer, 2008). It includes also advertising activities taking advantage of mobile phones' functionalities (Wikipedia, 2016). The Mobile Marketing Association defines mobile marketing as any marketing, advertising or promotional activity, targeted at clients and transmitted using a mobile channel (Salo J., Sinisalo J., Karjaluto, 2008).

The phenomena of e-marketing and m-marketing have been examined in numerous studies (ÅőwierczyÅĎska-Kaczor, 2012; Wielki, 2012; Kiba-Janiak M., 2014; Gao T., Sultan F., Rohm A. J., 2010; Roach G., 2009), including those carried out on a large scale (IAB, 2015; InternetStandard, 2015). Nevertheless, the majority of them took place before the most rapid development of smartphone and tablet applications or not takes into account this phenomenon. Fast, or rather rapidly growing, market of new mobile devices and the increasing pace of eliminating older devices from the market results in the fact that for the purpose of this article the authors have applied an additional division between traditional electronic marketing via the devices such as PC and laptop and a new mobile marketing, operated on smartphones and tablets. Unfortunately, in the present research the authors were not able to make a distinction between the traditional electronic marketing on the devices such as personal computers or laptops and mobile marketing used via these devices in browsers (browser marketing) and mobile marketing employed in applications (app-based marketing), because respondents are not always even aware of the distinction. Nevertheless, it allowed evaluating both forms of marketing used on smartphones and tablets.

The authors of the article hoped to indicate certain basic implications concerning the new phenomena impacting the directions of development of mobile marketing. Thus, they undertook a study whose main aim was to analyze the use of e-marketing and m-marketing among the users of different kinds of computer devices providing Internet access. The findings presented in the article constitute a summary report of the research examining a selected group of users in Poland at the beginning of 2016.

II. Research methodology

Due to few and fragmentary studies concerning the sphere of e-marketing and m-marketing applications as perceived by an individual client in national and foreign literature, the present research was based on the authors' own approach [Chmielarz, 2015) consisting of the following stages:

- analysis of a selected group of users on the basis of a quantitative and qualitative survey (CAWI (Computer Associated Web Interview method)),
- making an online version of a questionnaire available on the servers of the Faculty of Management at the University of Warsaw, and subsequent testing and verifying it,
- conducting user surveys,
- analysis and discussion of findings,
- drawing conclusions from the obtained findings on the current state of e-marketing and m-marketing and future directions of their development, based on the opinions of users.

The detailed scope of the survey, which consists of twenty-five questions, divided into five sections, is presented below:

- the place and role of the Internet in marketing,
- the assessment of e-marketing as a source of information about products/services,
- the assessment of the effectiveness of e-marketing media and their acceptance by the client,
- the assessment of the utility of e-marketing from the point of view of a client,
- assessment of m-marketing.

The article presents the findings of the analysis of completed questionnaires. The survey was carried out in March-April in 2016. The selection of the group was not quite random; it was the case of convenience sampling: the respondents were students of part-time and full-time BA and MA studies at the University of Warsaw. The choice of students groups were random. The surveys were distributed electronically, the level of responsiveness slightly exceeded 80%, despite the fact that students constitute a group which is particularly open to all kinds of innovations, especially those related to mobile devices. This group in Poland is the most active buyers in Internet.

A limitation in this sort of selection was an expected high participation of people owning smartphones, tablets, laptops and cell phones - not necessarily of high quality, but with a high usage time. The survey was completed by 130 people, with 106 respondents (81.54% of the sample) who completed the questionnaire correctly in its full form. Among the respondents, there were 65.09% women and 34.91% men. An average age of the respondent was 23.08 years, and the median value was 19 years of age. It is the typical age for the students of first years of part-time and full-time BA studies and the first years of MA studies, the interviewees who were asked to fill in the questionnaire. The oldest of the respondents was 51. Among the respondents, there were 65.09% of only students, 32.08% of working students and 23.89% professionals. Respectively, 66.98% declared having secondary education, and 27.36% of the respondents completed BA studies. Higher education and post-graduate studies were indicated by only 5.66% of the interviewees. Almost 36% of the respondents stated that they reside in cities of more than 500,000 inhabitants, over 12% came from cities of 100-500,000 inhabitants, more than 11% from cities with 50-100,000 residents, almost 21% of the sample were from towns with 50,000 inhabitants, and 19.81% from rural areas. The simplicity of the survey did not cause many distortions in the process of its completion.

III. Data analysis and discussion of findings

Respondents provided answers to twenty-five substantive questions. Here we concentrate on aspect of m-marketing in the frames of e-marketing, but survey had more wide context. The most important findings are listed below.

The first group of questions concerned generally the place and role of the Internet in marketing. Over 95% of respondents said that they use the Internet several times a day, 3.77% at least once a day, and only one respondent stated that he/she seldom uses the Internet.

The most popular device used by the respondents when accessing the Internet is a smartphone (33%), or a smartphone interchangeably with the laptop (32%). The laptop is used by slightly over 11% of students. Almost 8.5% of the survey participants use a personal computer for this purpose. This indicates a significant reorientation of Internet users towards the device which they have learned to use relatively early in life and use it for the longest time, that is, handheld mobile devices, especially a smartphone (the tablet itself is used by only by 3.77% of respondents).

Generally, the Internet is seen as a very good marketing medium by 52% of the respondents, or as a good marketing medium by 33% of the sample. Only 15% of the interviewees perceive the Internet as average or sufficient. Nobody has rated it as unsatisfactory. What lies at the core of such discrepancy between the level of using the Internet, making online purchases and a very good or good opinion on internet marketing? It appears that it stems not only from the specific shift of emphasis towards communication (as it is the main purpose of using smartphones) but also from the fact that the Internet has become the basic source of information on the reality, not only an economic one. The obtained information on products and services can also lead to making purchases in a traditional way, not only via websites. Also, one needs to consider the fact that the Internet, apart from its communication and information function plays a more and more important role as a means for providing entertainment (films, music, e-books, computer games, etc.). The last question in this section of the survey concerns the statement whether internet marketing is - in the opinion of the Internet users - better than traditional marketing. The majority of responses - i.e. 53% claimed that the best option is a combination of traditional and internet marketing (marketing mix). Interestingly, almost 40% of responses show that internet marketing is better than traditional marketing, and only 7.5% of the responses point out that it is just the opposite. The first thing which comes to the fore is the question why

internet marketing might be seen as better than its traditional form. The opinions concerning the reasons were divided. In almost 25% of responses of participants of the survey, they paid attention to its continuous accessibility and 23% to the possibility of gathering information about a particular product or service. The third place with the score of 22% was taken by the opportunity to use internet comparison engines, directly supporting the purchase process with special software. The fourth position was taken by the possibility of direct purchase after clicking on the link (19%). It should be noted that the difference between the evenly distributed attributes, in total, amounts to only six percentage points. Information about products and services on the Internet are mainly collected through search engines (almost 49% responses). Social media take the second position with the score of 28%, which is nearly 20 percentage points higher than the source which is in the first position. The blog entries and links to banners were at the level of 10% of responses. At the same time in order to verify the importance of the above information, the authors examined the percentage of people who perceive the Internet as the main source of information. Actually, in the examined group of respondents the Internet undoubtedly gained considerable recognition as a source of knowledge about products and services - 33% of responses. The next, second position was taken by information provided by friends (28% - word-of-mouth advertising). As the survey shows, television still enjoys a strong position (16% responses). The press and advertising materials in shops reached a similar level of about 8% of responses. However, the role of the radio and leaflets is markedly depreciated. The research also pointed to the low score of the opinions of the industry experts in the ranking.

Interestingly, information about the products and services is mainly used by the participants of the survey to make purchases in traditional stores (47% of responses), which would confirm the thesis concerning the importance of the Internet with regard to its informative function. Nevertheless, an almost equally large group of users (42% of responses) use the information to buy items in online shops and 10% on online auctions. This partly justifies the low percentage of Internet users declaring shopping in the electronic sphere - a considerable group use the information available on the Internet to make purchases in traditional stores.

The next section concerns the evaluation of the usage of e-marketing media and their acceptance by the client. For the Internauts, the presence in social media with the score of 30% of responses turned out to be the most significant factor. Nearly 22% of responses of students admitted that the most important factor inducing the purchase was an attractive website design. The least attention among the respondents assigning good scores was paid to e-mailing (nearly 5% of responses). In total, the presence in social media has obtained 46% of very good and good scores, and the clear and attractive website has gained 44%, which is a slightly lower score. Consequently - the greatest number of negative scores was obtained by e-mailing (more than 50%) and a newsletter 23% of responses.

The fourth section of the survey concerns the evaluation of the usage and the success of e-marketing from the point of view of a client. The ranking of marketing media inducing customers to make purchases corresponds with the previous results. In the last six months, the clients were most inclined to buy goods when attracted by the clarity and graphic design of the website (over 16%) and the presence in social media. The economic factor is not without significance - the third position with the score of 11% of responses was taken by the discounts offered when customers exceeded a specific value of the purchases. The worst in the ranking are brandmark and e-mailing (spam), running an internet forum and pop-up/under windows. Among other factors affecting purchases, the respondents indicated: discount codes/coupons available on the Internet.

The next element of the survey is the specification of a technical element - the tool which draws clients' attention to e-marketing to the greatest extent. Here the elements which are clearly visible on the screen, such as video - 27% of responses, or billboards - 23% of responses were decisively advantageous. Static and dynamic banners seem to be of lesser importance - 20% of responses. The respondents pay the least attention to buttons or pop-up/under windows.

The last, most comprehensive section of the survey examined m-marketing functioning on mobile devices such as a smartphone or a tablet. The questions concerned people who simultaneously use traditional devices (PC/laptop) and modern mobile equipment (smartphone/tablet). Generally speaking, if we consider the respondents' approach to electronic marketing it is nearly in one-third of responses (31%) negative. It appears that they prefer internet marketing on traditional hardware (40%). Only 15% answers of respondents indicated the preference for marketing on modern devices, and 14% of survey participants' responses claim that they like both types of marketing.

Over 65% answers of respondents claim that marketing on traditional hardware and modern mobile devices is different, in 28% of responses - have no opinion on the subject, and only in 7% answers - think that such differences do not actually exist.

Among the factors which are most disliked by customers in the manifestations of mobile marketing, the frequent comments concern greater problems with closing them (25% of responses) and their taking proportionally larger space on the screen (24%). Nevertheless, the statements related to problems encountered when communicating with the device are almost at the same level (23%). A half answers of the respondents still draws attention to the fact that displaying ads on mobile devices extends the time needed to roll the screens. Further positions are taken by the worse level of readability, connected with lower resolution, or excessive conciseness. Among other problems, the survey participants paid attention to the fact that older devices may be overloaded, and they tend to crash (see Fig.1).

Among the advantages of m-marketing, the respondents indicate the concise content (61% of responses), lower level of

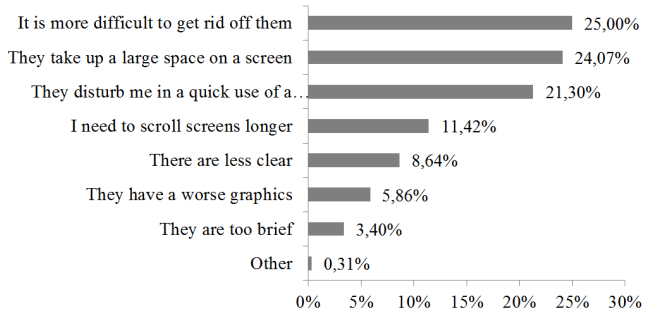


Fig. 1. Disadvantages of m-marketing in the respondents' opinions (%% of responses); n=314

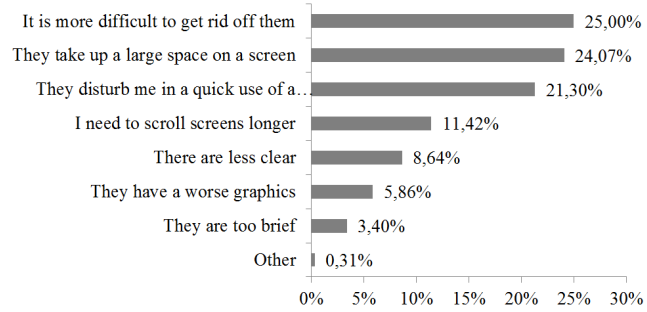


Fig. 2. The most important features of m-marketing (%% of responses); n=281

intrusiveness of the tools (25%) and the fact that, in general, they are easier to use (14%). The responses show that the feature which is indicated by some users as a disadvantage (e.g. excessive conciseness), may be seen as an advantage by others (content which is limited to the minimum, i.e. brevity).

Detailed analysis of the above responses was conducted through the prism of the most important features of mobile marketing. In over 27% of responses it was indicated the facts that m-marketing is always available, and it guarantees the precise and fast way to reach out to the target group (18%) are its most important features. The third position is taken by easy, dynamic and flexible interaction with the recipient (17% of responses). The high level of effectiveness of this type of marketing raises serious doubts (see Fig. 2).

The next question concerned the types of m-marketing tools which are most frequently used by the respondents. The greatest number of users indicated the use of mobile applications (28% of responses). The next positions are taken by SMS marketing - 26% of responses - and the use of geolocation and mobile navigation - 16%. The lowest scores are assigned to designing mobile websites and the use of NFC technology and QR codes (about 9% of responses). It appears that among the respondents few people used railway tickets (QR codes) or maps and guides (e.g. TatrzaÅĎski Park Narodowy). With regard to the question concerning the most visible forms of m-marketing the respondents pay the greatest attention to graphic advertising (73% of responses), next to application ads 12% and text ads 10%. In less than 5% answers of survey participants believe that the advertisements designed with the application of Google Web Designer technology prove to be most effective. Among m-marketing technologies operated on smartphones or tablets, the most efficient are graphic ads (43% of responses), then graphic application ads and the application ads (about 17% each). True View ads used to promote applications (about 1% of responses) and text ads (nearly 8%) are seen as less effective. The last question concerned the form of mobile marketing, which the respondents used most frequently in the last six months. The most popular forms were mobile applications (e.g. product placement or display ads), which amounted to 28% of responses. SMS marketing took the next position. The use of other forms of m-marketing was 5-6 percentage points lower in the ranking. The least popular forms were NFC (Near Field Communication) technology or QR (Quick Response) codes - generally at the level of 9% of responses.

IV. Conclusions

The conducted and presented studies point to the following conclusions:

- In the present study, all respondents of the survey were students, which was clearly indicated in the obtained findings. The higher the year of study, the less interest in completing the survey and its conclusions. It is probably caused by the greater effort required not only for the studies but also students' involvement with temporary or regular jobs (working students constitute more than 32% of the sample). On the other hand working students are very often used e-commers and have to do with e-marketing.
- The frequency of the use of the Internet in the examined sample was high - over 95% of respondents replied that they use the Internet several times a day, and nearly 4% at least once a day; they were connecting to the Internet only via their smartphone (1/3 of the sample), the second most popular option was laptop. It is the generation that has learned how to use a mobile phone/mobile device relatively early in their life and later on switched to smartphones or tablets; the change has been fairly easy and natural for them. The transition was supported by a relatively low price, in comparison to a laptop, and a large number of free applications available on all three operational systems (Android, iOS, Windows) running on smartphones,
- In the examined group fewer than 50% of respondents do their shopping on the Internet several times a month, up to a few times a day, which indicates a specific shift of the role played by the Internet from the economic role to a more communicative-informative one, i.e. the change in the main function of mobile phones at present is no longer focusing on business relations (only 6% of the sample use the Internet at work). The core aspect of the mobile sphere starts to prevail over the desktop one.

- At the same time, the Internet is generally perceived as a good and very good marketing medium - 85%. As far as the high level of internet purchases is concerned, the users declare that purchases made outside the Internet may be seen as the outlet for marketing activity,
- Simultaneously respondents claim that internet marketing is better than a traditional one due to its: continuous availability, a quick way to broaden the knowledge about a product or service and a possibility to use comparison engines - including mobile tools,
- Information about goods and services continue to be gathered by means of browsers (almost 49% of responses); social media take the second position, and the role of blog entries is gaining in importance, other sources of information are marginalized,
- Among the elements which respondents pay the greatest attention to with regard to e-marketing, we may list general graphic design and an advertised product or service,
- Among the technical elements of marketing, the Internet users see the advantage of the elements which clearly stand out on the screen: video (short film), or billboards. Static or dynamic billboards seem to be slightly less important,
- Marketing tools are positively perceived mainly in social media, on corporate blogs, on websites and in online shops. It emerges that in social media and on websites they are seen as both eagerly viewed and irritating - apparently to equal degree,
- A third of the "mobile generation" sees electronic marketing as a negative phenomenon, more people prefer to see it applied on traditional hardware rather than mobile devices,
- Over 65% of respondents believe that marketing via traditional equipment (PC, laptop) and modern devices (smartphone, tablet) are different,
- Among negative factors concerning m-marketing we experience greater problems with removing the ads from the screen, taking proportionally greater space and interfering in operating the device,
- Among the positive factors in m-marketing, the respondents mainly paid attention to concise contents, limited to a minimum, lower intrusiveness of the tool and the consequent greater comfort in using it,
- The most important feature of m-marketing is common availability, the guarantee of precise and fast access to the target group and easy, dynamic interaction with the recipient. However, there emerge certain doubts concerning the effectiveness of this form of marketing,
- The most popular type of m-marketing tool among the respondents were the ads available through mobile apps, SMS marketing as well as using geolocation and mobile navigation,
- The most noticeable forms of m-marketing are banners and much less (6 times less) noticeable app ads and text ads, and here lies, in the opinion of users, their greatest effectiveness,
- Most recently the respondents most frequently used the app-embedded ads (e.g. product placement and display) and SMS marketing.

The above conclusions constitute a proper basis for the continuation of the research as well as extending it to include the effects, consequences and results of using m-marketing in business activity.

REFERENCES

[1] Bernauer D., *Mobile Internet - Grundlagen, Erfolgsfaktoren und Praxisbeispie-le.* Vdm Verlag Dr. Müller.; 2008.

[2] Chmielarz W., *Marketing w sieci*, Chapter 3 in: Systemy elektronicznego biznesu, Difin, Warszawa, 2007, pp.139-168.

[3] Chmielarz W., *Study of Smartphones Usage from the Customer's Point of View*, Procedia Computer Science, Elsevier, Vol. 65, 2015, pp. 1085-1094, DOI: 10.1016/j.procs.2015.09.045.

[4] Gao T., Sultan F., Rohm A. J., *Factors influencing Chinese youth consumers' acceptance of mobile marketing*, Journal of Consumer Marketing 27/7, 2010, pp. 574-583, DOI: 10.1108/07363761011086326.

[5] Hasan J., *Analysis of E-marketing Strategies*, Studia commercialia Bratislavensia, Volume 4; Number 14 (2/2011), 2011, pp. 201-208, DOI: 10.2478/v10151-011-0006-z.

[6] Hovancakova D., *Mobile Marketing*, Studia commercialia Bratislavensia, Volume 4; Number 14 (2/2011), 2011, pp. 211-225, DOI: 10.2478/v10151-011-0007-y.

[7] IAB, http://iab.org.pl/badania-i-publikacje/raport-iabpwc-adex-2015-q1, accessed 2016-04-11.

[8] InternetStandard, http://www.internetstandard.pl/news/401894/Reklama.online.dominuje.rynek.html(eklamaonlinedominujerynek), accessed 2016-04-10.

[9] Kiba-Janiak M., *The Use of Mobile Phones by Customers in Retail Stores: a Case of Poland, Economics & Sociology*, Vol. 7, No 1, 2014, pp. 116-130, DOI: 10.14254/2071-789X.2014/7-1/11.

[10] Meng X., *Developing Model of E-commerce E-marketing*, Proceedings of the 2009 International Symposium on Information Processing (ISIP'09), Huangshan, P. R. China, August 21-23, 2009, pp. 225-228.

[11] Roach G., *Consumer perceptions of mobile phone marketing a direct marketing innovation*, Direct Marketing An International Journal Vol. 3 No. 2, 2009, pp. 124-138, DOI: 10.1108/17505930910964786.

[12] Salo J., Sinisalo J., Karjaluto H., *Intentionally developed business network for mobile marketing: a case study from Finland*, Journal of Business & Industrial Marketing 23/7, 2008, pp.497-506, DOI: 10.1108/08858620810901257.

[13] Sun S., *Innovation Mode and Strategy Research on Small and Medium-sized Enterprise E-marketing in Post Financing Crisis*, Contemporary Logistics 04, 2011, p. 13.

[14] Świerczyńska-Kaczor U., *e-Marketing przedsiębiorstwa w społeczności wirtualnej*, Difin, Warszawa, 2012.

[15] Wielki J., *Relacje organizacji z jej klientami*, Chapter 2.2.2., in: Modele wpływu przestrzeni elektronicznej na organizacje gospodarcze, Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu, Wrocław, 2012, pp. 87-104.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1335–1340
DOI: 10.15439/2016F534
ACSIS, Vol. 8. ISSN 2300-5963

The Role of ICT Solutions In the Intelligent Enterprise Business Activity

Monika Łobaziewicz
University College of Enterprise
and Administration in Lublin,
Poland
e-mail: ml@un.pl

***Abstract*— During the last years the number of innovative ICT systems, applications and tools has been growing still to support intelligent business performance. However, advanced ICT solutions are only means to the end of better process performance, not a substitute for it. Intelligent enterprises running their activity in increasingly more dynamic, complex and uncertain environment. The aim of paper is the discussion about the wide spectrum of ICT solutions used by the intelligent enterprise and their meaning in the management of intelligent organization. This is the first approach in this case that the author is going to continue in the advanced research.**

I. Introduction

There are many definitions of an intelligent enterprise. Undoubtedly, it has a high abilities to learn from experience, adapt to new situations, understand and handle abstract concepts, and use knowledge to manipulate the business environment. Problem solving, comprehending complex ideas, learning quickly, and learning from experience are crucial for the intelligent enterprise. Companies today are looking to boost operational productivity and performance while addressing the full range of information requirements throughout the extended enterprise. In this case, the intelligence in the enterprise management is more than just delivering reports from a data warehouse. It's about providing large numbers of people – executives, analysts, customers, partners, and everyone else – secure and simple access to the right information so they can satisfy their unique reporting or analysis requirements, then share that information accordingly.

The intelligent enterprise provides information for service-oriented purposes and for optimizing operational systems.

The enterprise can be intelligent in two ways:

- ✓ It can behave intelligently or/ and can „utilize" intelligence;
- ✓ For the enterprise to be intelligent, it needs to maximize the extend and utility of its intellectual capital;
- ✓ The intelligent enterprise is an organization which acts effectively in the present and is capable of dealing effectively with the challenges of the future. It meets its objectives both of the enterprise itself and those of its stakeholders and makes trade – offs between them.

Because, the role of ICT in business management is crucial and modern enterprises run their activity using advanced IT systems, applications, tools – the question concerning the differences between intelligent and traditional enterprise appears.

The aim of paper is the discussion about the wide spectrum of ICT solutions used in the intelligent enterprise and their meaning in the management of intelligent organization. This is the first approach in this case that the author is going to continue in the advanced research.

II. The Intelligent Enterprise

Nowadays, the concept of an intelligent enterprise has its source particularly in the following ideas: an organisation based on knowledge and information management, a learning organization, an organisation based on the intellectual capital. On the other hand, from the ICT point of view, the intelligent enterprise is correlated with following terms: business intelligence, artificial intelligence, enterprise intelligence systems. The results of a literature surveys conduct that a discussion about the intelligent organisation has been started in 90. of the 20'th century. The concept of the "intelligent enterprise" was first articulated by J. B. Quinn as follows "the self-sufficient enterprise is becoming anachronistic. Each organization is part of a matrix of merging and evolving ideas and opportunities. Leading companies focus less on positioning and more on patterns of people and institutions they work with – or against" [1] and now it has a special meaning because of the use of advanced tools based on ICT.

Ming & Feng [2], Hopkins Lavalle, Balboni [3], Kruschwitz & Shockley [4], Dayani [5], Quinn [6], Stubbs [7], Tan & Cao [8,9] note that the knowledge management, wireless networking technologies, mobile devices has prompted many modern enterprises to look for management information systems to remotely monitor and control of their company operations in order to increase their flexibility and competitiveness in the market. In other words, the intelligent enterprise operates with knowledge-

based technologies, especially on-line systems for remote work and activities improving the effectiveness of business processes and has important role in creating competitive advantage.

Szczerbicki [10] notices that a modern intelligent enterprise is able to convert intellectual resources using ICT solutions to the end product with a high level of added value.

One should be pointed that now intelligent enterprises use in their business a hybrid approach rather than using a single intelligent system or application to do activities and to make decisions. Modern IT environment includes various interfaces and components completely Web-based and uses XML extensively which can work like shared platform to be accessed by multiple users and decision makers [11]. Enterprises operate on B2B platforms with 'in built' EDI technologies that integrate ERP systems and special applications of business partners, use the workflow, CRM. All of them provide a lot of data and information in the integrated way. Thus, they act like knowledge management systems [12], [13]. Nowadays the main problem for the intelligent enterprise is not be the access to information but the ability to verify it and then to transform it into a useful operational and strategic resource.

Managing of the enterprises that function in uncertain 'information-rich' environment requires greater understanding of the role of information and in systems operation. To gain this understanding, the knowledge is required [14]. In the intelligent enterprise, employees know that ICT tools enable the knowledge sharing, not only fosters collaboration but also facilitates experience and knowledge discovery. Thannhuber emphasizes that IT systems supported by knowledge and intelligence paired together allow to adapt dynamically the enterprise to its environment, provide the framework for making optimal decisions [15]. Moreover, the intelligent enterprise applies automated analytics on data generated by systems and applications to better understand what resources are being used, how well they should be used to support the business processes. The intelligent enterprises create the high ability to measure past performance for the future purposes. ICT solutions deliver knowledge to the right people when and where it is needed, and keep in mind that timeliness is an issue. The issue is that not everybody needs all information, they need just the right information for themselves. In the intelligent enterprise, there is a situation when delivering reports from a data warehouse is an operational action. The intelligence of this organization is about providing large numbers of people – executives, analysts, customers, partners, and everyone else – secure and simple access to the right information so they can satisfy their unique reporting or analysis requirements, then share that information accordingly to the logic of business processes.

According to Dayal [16] the intelligent enterprise is characterized by being able to adapt quickly to changes in its operating environment. It monitors not only its own business processes but its interactions with customers, partners, suppliers and collaborators, as well. The intelligent enterprise understands how the exchange of information among all business participants relates to its business objectives and it acts to control and optimize its operations to meet its business objectives. In this enterprise decisions are made quickly and accurately to modify business processes on the fly, dynamically allocate resources, or change business partners (e.g., suppliers, service providers) and partnerships (e.g., establish new service level agreements).

III. BUSINESS MODELS OF THE INTELLIGENT ENTERPRISE

The results of literature surveys conduct that there are very little discussion about business models of the intelligent enterprise. Most of the conceptions are related to business intelligence (BI) or knowledge management(KM) models. There are a few concepts presented as white papers or presentations exposed at IT conferences.

Professor Larry Lucardie from Knowledge Value Institute defines the intelligent enterprise as lean, agile and learning, which a business model is based on the knowledge value (Fig. 1).

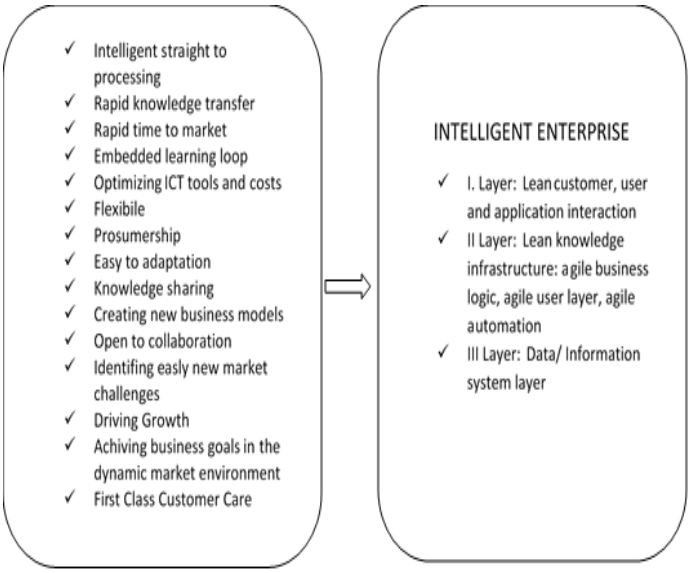


Fig. 1 Intelligent enterprise optimizing the knowledge driven organisation

Andrew Coleman from IBM notices that the business model of intelligent enterprise is based on prediction. The intelligent organisation is able to compare what is happening right now with past experience to predict the future so that it can anticipate the changes needed to proactively optimize the business. Therefore, the intelligent enterprise is a market game changer (Fig. 2).

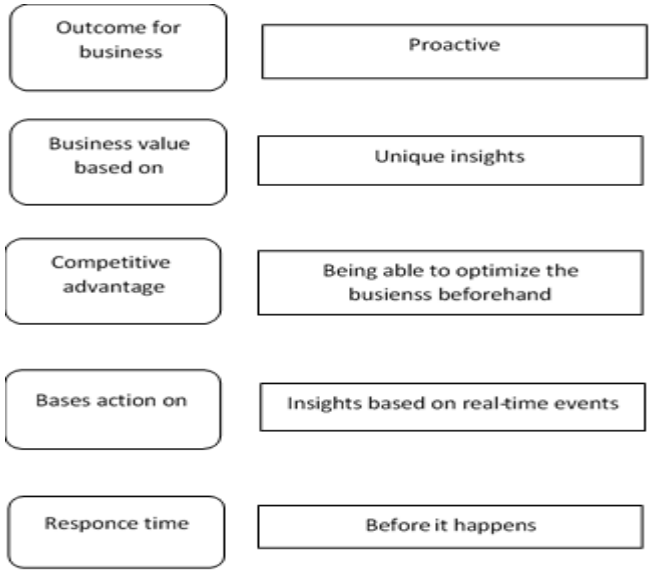


Fig. 2 Predictive intelligent enterprise model

The research conducted by IBM in a group of 225 business leaders worldwide, show that enterprises are operating with bigger blind spots and that they are making important decisions without access to the right information. They recognize that new analytics, coupled with advanced business process management capabilities, signal a major opportunity to close gaps and create new business advantage. Those who have the vision to apply new approaches are building intelligent enterprises and will be ready to outperform their peers [17]. IBM have pointed the essential characteristics that describe an enterprise ready to exploit advanced analytics and optimized performance (Fig. 3).

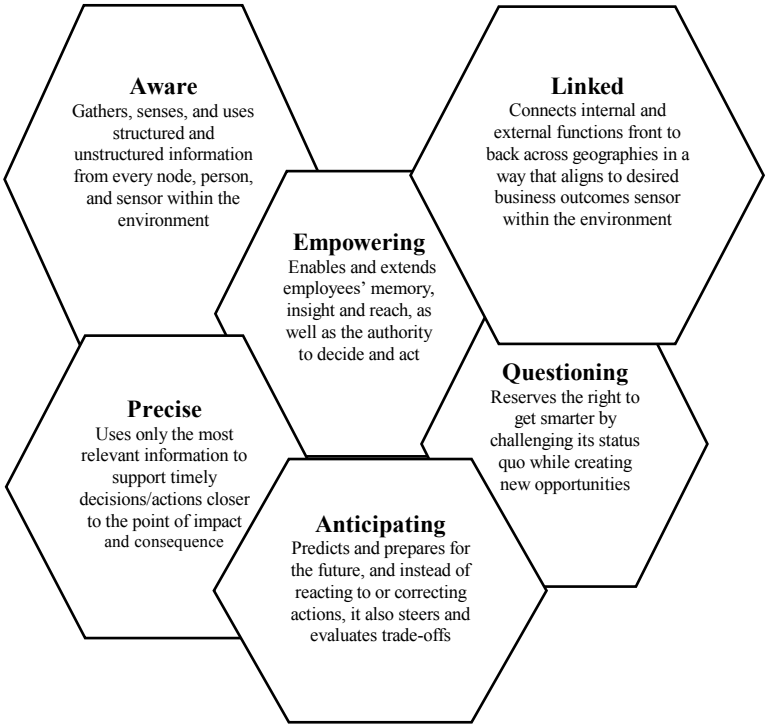


Fig. 3 Characteristics of the intelligent enterprise in IBM conception

Intelligent enterprises operate in increasingly complex IT systems what is the result of business processes complexity. Autonomous subsystems are still be interrelated and embedded in larger systems. Intelligent enterprises opposite to traditional organizations are able to integrate their strategy and the knowledge management with IT systems, applications and tools (Fig. 4).

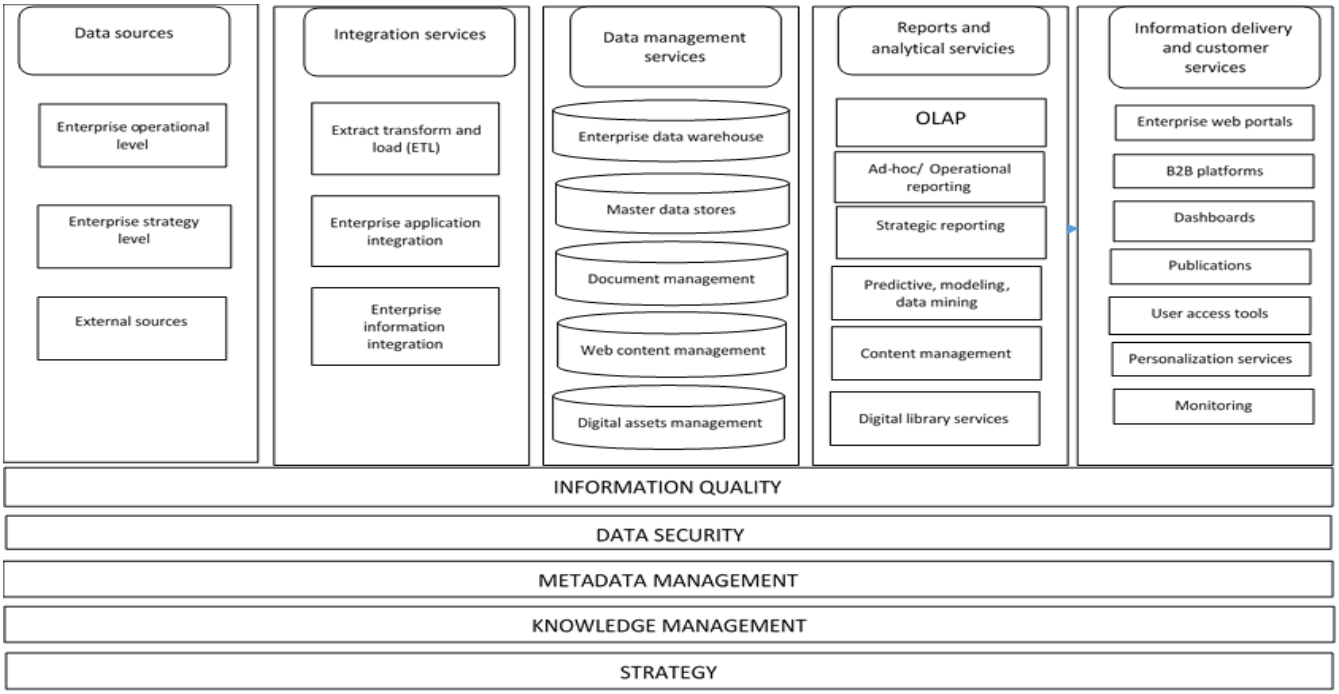


Fig. 4 Integrated intelligent enterprise model

Building the business model of the intelligent enterprise provides for a strategy and technology infrastructure that ensures that accurate and timely information is effectively incorporated into the decision making process so that the organizations can exploit this information through process, knowledge and visualization based technologies to manage their business effectively.

IV. THE RESEARCH OF INTELLIGENT ENTERPRISE

The study show that there is a lack of advanced surveys devoted to the intelligent enterprises.

Up to date, in Poland the only research concerning intelligent enterprises was carried out by Polish Agency for Enterprise Development (PARP) in 2010 on a group of 300 small and medium-sized enterprises (SMEs). One of the purpose of the research was finding an answer to a question what are the characteristics of the intelligent enterprise in Poland and whether they use ICT solutions more effectively than other organisations.

In the research carried out by PARP it was assumed that an intelligent organisation has following features:

- it has a long term strategy of development to achieve goals;
- it has an advanced human resources management (HRM) policy;
- it has a company website and intra network as well as it uses specialised ICT business management tools;
- it uses the knowledge management.

Surveys have shown that 26.5% of SMEs had a long term strategy, 31.6% had the HRM policy, 47% used developed ICT tools and 38% used the knowledge management. In contrast, 63% of big companies had both the strategy and the personnel management policy well developed. Therefore, the bigger organisations meet the criteria of intelligent organisation to a larger extent than SMEs.

In Poland, intelligent organizations do not have a clear innovative profile yet established. Now, when the Operational Programme Intelligent Development 2014-2020 started, it is known that a type of innovation is not a factor differentiating companies in terms of their willingness to implement solutions typical for intelligent organizations. More often are process innovations (28%), organizational innovations (24%) and product innovations (21%). The tendency to introduce the solutions adequate for intelligent organizations to the business practice increases with the size of company turnover. From the business sector point of view, intelligent organisations have the biggest share among industrial companies (14%), as well as trade and service companies.

The research indicate then a stronger focus on technological development among intelligent organizations, their better adaptation to the challenges of the knowledge based economy, the speed of access to knowledge and the possibility of its use are key competitive factors.

Intelligent organizations in Poland more often use ICT solutions to support management processes in comparison with other organizations. The most popular are e-workflow, databases and data warehouses management (83%), as well as Intranet (76%). The further are Customer Relationship Management (twice more often than organizations that do not meet criteria for intelligent organizations) and solutions supporting a team working, every fifth - HRM and every sixth - Business Intelligence (three times more often than other organizations).

The last problem concerning ICT solutions that support the management of intelligent organizations is their effectiveness assessment. The few critical comments were focused on low efficiency of databases and data warehouses. Very positively were evaluated Supply Chain Management (78%) and Customer Relationship Management (70%). As far as the effectiveness of various ICT tools by intelligent organizations is concerned, it is worth to emphasize that generally ICT tools are assessed as less effective by small businesses than by middle sized and large. This is due to the specific nature of these tools, which do not necessarily have to be effective in organizations with a low developed organizational structure and not very complicated business processes [18].

In 2010 MIT Sloan Management Review and the IBM Institute for Business Value conducted a research among nearly 3,000 executives, managers and analysts working across more than 30 industries and involved intelligent organizations of various sizes in more than 100 countries. There were also interviewed academic experts and subject matter experts from a number of industries and disciplines to understand the practical issues facing intelligent organizations [19].

As a result, there are following results:

- Intelligent enterprise is focused on the highest value uses each business opportunity, starting with questions, not data opposite to traditional organizations. It should first define the insights and questions needed to meet the big business objective and then identify data needed for targets. They can target specific subject areas, and use readily available data in the initial analytic models;
- Intelligent enterprise drives actions and delivers value. This means that new methods and tools to embed information into business processes, ICT analytics solutions, optimization, workflows and simulations are making insights more understandable and actionable;
- Intelligent enterprise develops existing capabilities adding new ones. To do this, they use sophisticated modelling and visualization tools based on ICT, but that does not mean that spreadsheets and charts should go away. On the contrary, new tools should supplement earlier ones, or continue to be used side by side, as needed;

- Intelligent enterprise uses an information agenda to do plan for the future. Big data is getting bigger. Information is coming from interconnected supply chains today. Strategic information arrives through unstructured digital channels: social media, smart phone applications and an ever-increasing stream of emerging Internet-based gadgets. The information agenda identifies foundational information practices and tools while aligning IT and business goals through enterprise information plans and financially justified deployment road maps. This agenda helps establish necessary links between those who drive the priorities of the organization by line of business and set the strategy, and those who manage data and information. A comprehensive agenda also enables managers to keep pace with changing business goals. It provides a vision and high-level road map for information that aligns business needs to growth. [3]

Summing up, intelligent enterprises are combining the new systems and tools based on ICT with expertise in business process management. They are able to extract the precise information they need – highly relevant and contextualized – and predict the most likely outcomes of key decisions and events. They are able to shape their own futures.

V.ICT DRIVERS EMPOWERING THE INTELLIGENT ENTERPRISE

The results of theoretical study and the research conducted by PARP and MIT Sloan Management Review and the IBM Institute for Business Value became the background to develop a scientific discussion about the intelligent enterprise and the role of ICT in this organization. Taking into account the wide spectrum of ICT solutions used in the intelligent enterprise management, there are some areas to analyse.

5.1 Mobile workforce integration

The access to professional knowledge is critical in the intelligent enterprises and mobile connections to operating systems, applications, platforms are important, especially in the fast – paced business environment. Mobile technologies drive technical innovation to improve networks, ensure employees remain fully integrated with their company and clients wherever they are. Thus, in the intelligent organization its coherency is determined by the intelligence of its network that becomes the organization with wireless tentacles spreading from it to embrace location-aware services.

5.2Smart virtual workplace

As the approaches to virtualization of IT infrastructure, networks and storage devices continue to mature, infrastructures become software-driven. Smart virtual workplace provides end to end desktop virtualization allowing employees to access applications, data safely over any network from the device of any choice. New trends show that business will increasingly turn to hybrid cloud solutions to enable scalable business processes. Hybrid clouds can quickly scale to a company's needs and services can be paid for as needed. They combine the best of two worlds, offering true benefits to intelligent enterprises aiming to stay ahead in their markets.

5.3 E- Collaboration

E-collaboration is the standard for business communication today, nearly eliminating the need to meet face to face. While knowledge sharing increases, formal and informal groups become e- collaborative communities to reach organizational goals. Intelligent enterprises continue to integrate these into their business processes and reinvent their customer engagement models.

In the intelligent enterprise ICT tools allow disparate teams to work together in real-time, enabling multiple individuals to interact as efficiently and effectively with co-workers, clients, and suppliers.

5.4 Business flexibility

The intelligent enterprise can be called as the designer of changes where the business flexibility is crucial. ICT solutions are very helpful in this case. They must now be highly flexible and resilient in order to seamlessly communicate and interoperate with disparate technologies and systems. The IaaS model becomes very useful. It makes it easier the internetworking and deploying servers and endpoints from multiple sources.

5.5Scalability and customization

Intelligent enterprises align their IT infrastructure capabilities with business requirements. Modularity of systems, applications allow companies to have only what is needed at present, trimming up-front costs and leaving open the possibility of expanding or incorporating new technologies in the future. With the increase of consolidation, intensive virtualization, the traditional data center will transform to the 'hyperscale' data center. It requires a fundamentally different approach than that taken with typical enterprise IT systems. Rather than building 'monolithic' platforms, distributed architecture design is implemented around distributed processing frameworks. That requires software and ICT tools that automate node deployment, recover from failure (rerouting of workloads), and other management and monitoring tools.

5.6Business continuity

It is obvious that intelligent companies need to have 24 hours a day access to their data. Data digitalization and rapidity of their processing require more accurate, reliable

and sophisticated ICT tools converting all data into intelligence for better business outcomes. On the other hand, managers need them to be not complicated in their use. Moreover, for a high level of operational uptime, infrastructure components must be fault tolerant with the ability to recover from complex failures and data storage must be secure.

5.7 Converting data into business intelligence

Advanced ICT solutions enable extracting from huge amounts of data collected from the real cyberspace. Intelligent enterprises are able to manage Big Data to drive better business processes, product development, and customer service. The important is the fact that they enable to use effectively unstructured data captured from different systems, mobile devices, social media, log files, emails to perform real-time context analytics to understand received information, its content to make right decisions in the right time.

Therefore, intelligent enterprises are not only the users of advanced tools based on ICT technologies to optimize business practices, drive workforce engagement and create a competitive edge, but they are also able to leverage and to create value from the date and information generating by ICT solutions.

VI. CONCLUSIONS

The discussion about intelligent enterprise is at the beginning stage. There are more theoretical disputes then surveys in this case. There are no in depth research devoted business models of these organizations or effectiveness, strategy or management in this organizations, especially with use of ICT solutions.

In the paper, there were presented two surveys that can be the good start for the future research about intelligent enterprises. Poland is at the stage of an intensive investing in the research and development, therefore Polish companies are still learning how to create the intelligence and how to be intelligent organizations. This is especially a challenge for companies from the SMEs sector.

The research conducted by MIT Sloan Management Review and the IBM Institute for Business Value show that for the intelligent enterprise, the new reality is this: personal experience and insight are no longer sufficient. New analytics capabilities are needed to make better decisions.

The above premises encourage the author to continue the scientific discussion about intelligent enterprise with a special attention to ICT solutions. The advanced research will be continued.

REFERENCES

[1] “The Intelligent Enterprise: A New Paradigm, Academy of Management Executive, 6(4), 1992

[2] Y. H. Ming, and D. X. Feng: "Research on the Intelligent Enterprise Based on Intelligent Behavior", Proceedings of the 7th International Conference on Innovation and Management, Vols. I and II, 2010, pp. 2094–2099.

[3] M. S. Hopkins, S. Lavalle, and F. Balboni: "10 Insights: A First Look at The New Intelligent Enterprise Survey", Mit Sloan Manage Rev, 2010, 52, (1), pp. 22.

[4] N. Kruschwitz, and R. Shockley: "First Look: The Second Annual New Intelligent Enterprise Survey", Mit Sloan Manage Rev, 2011, 52, (4), pp. 87-89.

[5] B. Dayyani: "The Intelligent Enterprise: Knowledge-Driven Category Management", Proceedings of the 7th International Conference on Intellectual Capital, Knowledge Management and Organisational Learning, 2010, pp. 138-145.

[6] J. B. Quinn: "The intelligent enterprise a new paradigm", Acad Manage Exec, 2005, 19, (4), pp. 109-121.

[7] E. Stubbs: ‘The Intelligent Enterprise’, Wiley Sas Bus Ser, 2014, pp. 79-93

[8] "The intelligent enterprise and the changing role of computer information systems in strategic planning", Information Resources Management Journal, 1991, 4, (1), pp. 21-29.

[9] B. Tan, X. W. Cao, and D. Ahpak: "Achieving Competitive Advantage through Building an Intelligent Organization with Technological Innovation: The Case Of A Chinese Enterprise", Proceedings of 2010 International Conference on Innovation, Management and Service, 2010, pp. 24-29.

[10] E. Szczerbicki: "Intelligent enterprise management", Cybernet Syst, 2001, 32, (7), pp. 697-699.

[11] S. G. Wang., R. Liu, and X. T. Liu: "An enterprise intelligent system development and solution framework", Icemi 2007: Proceedings of 2007 8th International Conference on Electronic Measurement & Instruments, Vol Iv, 2007, pp. 115-118.

[12] J. N. Gupta, and S.K. Sharma: "Creating knowledge based organizations", Igi Global, 2004.

[13] G. H. Stonehouse, and J. D. Pemberton: "Learning and knowledge management in the intelligent organisation", Participation and Empowerment: An International Journal, 1999, 7, (5), pp. 131-144.

[14] E. Szczerbicki, Z. Gomółka, "Management of Information in Complex Systems: Perspectives for New Millenium", Intelligent Processing and Manufacturing of Materials, IPMM '99. Proceedings of the Second International Conference on Vol. 2,1999, DOI: 10.1109/IPMM.1999.791491.

[15] M. J. Thannhuber: "The Intelligent Enterprise", Springer, 2005.

[16] U. Dayal, „Managing the intelligent enterprise”, e-Commerce Technology, 2004. CEC 2004. Proceedings. IEEE International Conference on 2004, DOI: 10.1109/ICECT.2004.1319711.

[17] “Business analytics and optimization for the intelligent enterprise”, IBM Institute for Business Value, IBM Global Business Services Executive Report, USA 2009.

[18] P. Kordel, J. Kornecki, K. Pylak, J. Wiktorowicz, A. Kowalczyk, K. Krawczyk, “Inteligentne organizacje – zarządzanie wiedzą i kompetencjami pracowników”, Warszawa 2010.

[19] S. LaValle, E. Hopkins, R. Lesser, M.S. Shockley, N. Kruschwitz, “Big Data, Analytics and the Path From Insights to Value, Research Feature, December, 2010.

22nd Conference on Knowledge Acquisition and Management

KNOWLEDGE management is a large multidisciplinary field having its roots in Management and Artificial Intelligence. Activity of an extended organization should be supported by an organized and optimized flow of knowledge to effectively help all participants in their work.

We have the pleasure to invite you to contribute to and to participate in the conference “Knowledge Acquisition and Management". The predecessor of the KAM conference has been organized for the first time in 1992, as a venue for scientists and practitioners to address different aspects of usage of advanced information technologies in management, with focus on intelligent techniques and knowledge management. In 2003 the conference changed somewhat its focus and was organized for the first under its current name. Furthermore, the KAM conference became an international event, with participants from around the world. In 2012 we’ve joined to Federated Conference on Computer Science and Systems becoming one of the oldest event.

The aim of this event is to create possibility of presenting and discussing approaches, techniques and tools in the knowledge acquisition and other knowledge management areas with focus on contribution of artificial intelligence for improvement of human-machine intelligence and face the challenges of this century. We expect that the conference&workshop will enable exchange of information and experiences, and delve into current trends of methodological, technological and implementation aspects of knowledge management processes.

TOPICS

- Knowledge discovery from databases and data warehouses
- Methods and tools for knowledge acquisition
- New emerging technologies for management
- Organizing the knowledge centers and knowledge distribution
- Knowledge creation and validation
- Knowledge dynamics and machine learning
- Distance learning and knowledge sharing
- Knowledge representation models
- Management of enterprise knowledge versus personal knowledge
- Knowledge managers and workers
- Knowledge coaching and diffusion
- Knowledge engineering and software engineering
- Managerial knowledge evolution with focus on managing of best practice and cooperative activities
- Knowledge grid and social networks
- Knowledge management for design, innovation and eco-innovation process
- Business Intelligence environment for supporting knowledge management
- Knowledge management in virtual advisors and training
- Management of the innovation and eco-innovation process
- Human-machine interfaces and knowledge visualization

EVENT CHAIRS

- **Hauke, Krzysztof,** Wroclaw University of Economics, Poland
- **Nycz, Malgorzata,** Wroclaw University of Economics, Poland
- **Owoc, Mieczyslaw,** Wroclaw University of Economics, Poland
- **Pondel, Maciej,** Wroclaw University of Economics, Poland

PROGRAM COMMITTEE

- **Andres, Frederic,** National Institute of Informatics, Tokyo, Japan
- **Chmielarz, Witold,** Warsaw University, Poland
- **Christozov, Dimitar,** American University in Bulgaria, Bulgaria
- **Helfert, Markus,** Dublin City University, Ireland
- **Jan, Vanthienen,** Katholike Universiteit Leuven, Belgium
- **Jelonek, Dorota,** Faculty og Management of Czestochowa University of Technology
- **Kania, Krzysztof,** University of Economics in Katowice, Poland
- **Kayakutlu, Gulgun,** Istanbul Technical University, Turkey
- **Korczak, Jerzy,** Wrocław University of Economics, Poland
- **Ligęza, Antoni,** AGH University of Science and Technology, Poland
- **Mach-Król, Maria,** University of Economics in Katowice, Poland
- **Mercier-Laurent, Eunika,** University Jean Moulin Lyon3, France
- **Nalepa, Grzegorz J.,** AGH University of Science and Technology, Poland
- **Sobińska, Małgorzata,** Wroclaw University of Economics, Poland

- **Surma, Jerzy,** Warsaw School of Economics, Poland and University of Massachusetts Lowell, United States
- **Vasiliev, Julian,** University of Economics in Varna, Bulgaria
- **Zhelezko, Boris,** Belorussian State Economic University, Belarus
- **Zhu, Yungang,** College of Computer Science and Technology, Jilin University, China

ORGANIZING COMMITTEE

- **Marciniak, Katarzyna,** Wroclaw University of Economics, Poland

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1343–1348
DOI: 10.15439/2016F001
ACSIS, Vol. 8. ISSN 2300-5963

Knowledge Acquisition at the Time of Big Data

Francis Rousseaux, Stéphane Cormier
Reims Champagne Ardenne University, France
Email: {first name.name}@univ-reims.fr

***Abstract*—What is exactly 'Big Data', and for what purpose and application is it really efficient? Between the commercial promises made by the industrial actors and the Cassandra's cautions from some whistle-blowers, we propose a singular Big Data field to investigate with Inductive Data-Driven Algorithms: developing collections. Last but not least, we investigate the innovative possibility to curate 'figural' collections, characterized by Jean Piaget as follows: *"A figural collection composes a figure, through the spatial relationships between its elements, whereas non-figural collections and classes are free of any figure"*. Thus, we incidentally disclose some important Abstract Truth of Big Data.**

***Index Term*—predictive analysis; inductive data-driven algorithms; big data; figural collections; digital societies; search by similarity.**

I. Introduction

THE classical processes of Information Technology (IT) – including Artificial Intelligence approaches – were traditionally based on deductive knowledge modelling and simulation, so that explanation and theory, based on validations or refutations [13], were never far from engineering.

Things turn different within Inductive Data-Driven Algorithms and Big Data, because practitioners do not need anymore domain theories and explanation-based processes to succeed in providing useful results.

Those new features and paradigmatic evolutions introduce some important epistemological breakthrough in the IT classical environment, with important consequences.

At least, our digital societies are about to revisit an old utopia a human dream, close to our day live: the management of our overabundant and oversized collections, as lighten by contemporary art.

II. Predictive Analysis through Big Data: a Tsunami in IT

A. What changes has the arrival of Inductive Data-Driven Algorithms and Big Data made in the data ecosystem?

The title of the document is placed in appropriate frame on the top of the document and is formated with style `Title`. Do not capitalize short words (a, and, the, and so on).

In the world of computing, we are currently living through a period in which the epistemological nature of data is being thoroughly changed, as Big Data seems to be revolutionizing most of the methods and approaches of digital decision-making tools and processes. Indeed it is as if the Curse of Dimensionality and the pitfall of the Combinatorial Explosion that limited computer science in the 20th century have been left behind us. Natural allies of Data Mining and Machine Learning, the tendency is now towards Big Data and Data-Driven Intelligent Predictive Systems (DDIPS).

These systems are called Data-Driven because they mobilize Big Data to make connections or analogies, aiming to create certain configurations or to anticipate situations that might cause delays or predictive challenges. They are called Predictive because, unlike 20th century computer systems, their performances are measured more by their ability to predict or to discover rather than to understand, explain or theorize.

They are called Intelligent because they frequently make use of input from supervised or unsupervised automatic learning techniques, data mining, and even Deep Learning based on Convolutional Neural Networks, which contribute to performances that far outstrip the preceding generation's.

Because of this epistemological turnaround, and with the help of DDIPS, a flood of predictive data are being added to the usual data, although this new type of data has never been directly input or calculated by determinist or classically deductive methods [6], [19]. Thus Predictive-Data – to be distinguished henceforward from other types of data – are considered an additional raw material to be exploited, even if their reliability must be closely scrutinized.

B. What about the current positions and statements made by main actors?

IT professional and industrial companies, but also start-ups working in that domain, are putting forward on promising business and marketing opportunities, furnishing and putting tool and technical solutions on market places[1].

But in parallel observers and researchers, often coming from social sciences fields, forewarn potential users from a new calculus order that could mean their end of autonomy. To mention only a few of them:

1) The end of Theory

Following George Box saying: *"All models are wrong, and increasingly you can succeed without them",* Chris Anderson [1], after he has been the editor-in-chief of WIRED magazine, puts forward that: "The Data Deluge Makes the Scientific Method Obsolete".

He claimed that: *"Petabytes allow us to say: Correlation is enough. We can stop looking for models. We can analyse*

[1]See for example http://www.bigdataparis.com/guide/BD14-15_Guide_BD_14136_2.pdf

the data without hypotheses about what it might show. We can throw the numbers into the biggest computing clusters the world has ever seen and let statistical algorithms find patterns where science cannot".

2) "Technological Solutionism"

In one of his recent bestseller books Evgeny Morozov [6], the Byelorussian expert analyst Evgeny Morozov explained: *"There is something, to me, that is very worrying about the idea of replacing causality with correlation, because if you do want to engage in reform, you do need to understand the causal factors that you will be reforming. If you just focus on correlations, all you'll be doing is basically adjusting the behaviour of the system without understanding the root causes that are driving it".*

And also: *"So the proliferation of big data and the ability to track things that we do is good only if we can actually understand why we engage in those behaviours. The ability to understand why I think is fundamental to understanding what it is that needs to be changed".*

3) "Algorithmic Governementality"

The European researcher Antoinette Rouvroy says [17]: "La rationalité post-moderne [engendrée par le Big Data] est fondée sur la découverte de corrélations entre des données recueillies avec des intentions et dans des contextes extrêmement divers, hétérogènes les uns aux autres, et qui ne sont reliés entre eux par aucun lien de causalité. C'est le renoncement au savoir causal, la dévaluation de l'expérience sensible elle-même au profit du calcul".

And also [16]: *"Operations of collection, processing and structuration of data for purposes of data mining and profiling, helping individuals and organizations to cope with circumstances of uncertainty or relieving them from the burden of interpreting events and taking decision in routine, trivial situations have become crucial to public and private sectors' activities in domains as various as crime prevention, health management, marketing or even entertainment. The availability of new ICT interfaces running on algorithmically produced and refined profiles, indiscriminately allowing for both personalization (and the useful, safe and comfortable immersion of users in the digital world) and preemption (rather than regulation) of individual and collective behaviours and trajectories appears providential to cope with the complexities of a world of massive flows of persons, objects and information, and to compensate for the difficulties of governing by the law in a complex, globalized world. The implicit belief accompanying the growth of 'big data' is that, provided one has access to massive amounts of raw data (and the world is actually submersed by astronomical amounts of digital data), one might become able to anticipate most phenomena (including human behaviours) of the physical and the digital worlds, thanks to relatively simple algorithms allowing, on a purely inductive statistic basis, to build models of behaviours or patterns, without having to consider either causes or intentions. I will call 'data behaviourism' this new way of producing knowledge about future preferences attitudes, behaviours or events without considering the subject's psychological motivations, speeches or narratives, but rather relying on data. The 'real time operationality' of devices functioning on such algorithmic logic spares human actors the burden and responsibility to transcribe, interpret and evaluate the events of world. It spares them the meaning-making processes of transcription or representation, institutionalization, convention and symbolization".*

C. Our point: Towards a scientific study of opportunity

Many of our modern computerized activities, may they be personal, professional or even artistic, involve searching, classifying and browsing large numbers of digital objects.

Until recently, the usual tools we had at hand, however, were poorly adapted as they were often too formal, because the current models for information search often assume that the function and variables defining the categorization are known in advance.

In practice, however, when searching for information, experimentation plays a good part in the activity, not due to technological limits, but because the searcher does not know all the parameters of the class he wants to create. He has hints, but these evolve as he sees the results of his search. The procedure is dynamic, but not totally random.

The collector's experimentation is always carried out by placing objects in temporary and metastable space/time. Here, the intension of the future category has an extensive figure in space/time. And this system of extension gives as many ideas as it does constraints. What is remarkable is that when we collect something, we always have the choice between two systems of constraints, irreducible one to the other. This artificial 'undifferentiation' for similarity/contiguity is the only possible kind of freedom allowing us to categorize by experimentation.

Nowadays, our software design could strongly become backed up by both artistic and psychological knowledge concerning the ancient human activity of collecting, which can be described as a metaphor for categorization in which two irreducible cognitive modes are at play: aspectual similarity and spatiotemporal proximity.

Inductive Data-Driven Algorithms and Big Data could help, definitely, to allow the creation of a new operational space, in between formal classes/categories and radical singularity.

III. The need for computer-aided collections management tools

A. An illustrative example

Let us illustrate this situation. First, let us suggest that looking for new material and classifying are two important processes involved in collecting. Indeed, when someone decides to start building a collection he usually already possesses a few items. Then, to extend this collection, new items must be added. In order to do so, the collector goes into the world and looks for these new items. Then as the collection builds up, the need to arrange the items into categories will become clearer, as the collection cannot simply remain a messy stack of unordered items [14].

Let us describe a particular example: the music collector. This collector will surely possess some initial items; these

may be some CDs or vinyl records. His first action involved in extending his collection could be a visit to the record shop for example. Here, the music is classified conformingly to the record companies' desires, which can sometimes be confusing for our collector, who is a fan of Jimi Hendrix, and just does not know where to look for his albums: in the blues section? Rock section? Is there a 'sixties' section? Anyway, despite finding them rather practical at first sight, our collector didn't create these labels, and finds it difficult adapting to them. However, as he browses through the shop, he also notices some nicely illustrated records, and discovers new artists he is interested in because their records are sitting next to Jimi's. Finally, when he has bought enough music records, and come back home, he will be able to start arranging his collection in a very personal and satisfying manner, which will be pleasing to the eyes, and also allow him to retrieve items quickly.

If he had decided to collect digital music, and go online to find new items for his collection, the process would have been rather similar. Commercial music download sites allow the user to browse through predefined music categories, thus implementing a kind of virtual record shop with the same problems mentioned earlier. The search tool however can come in handy, and allow the user to search for the name of an artist, a song, an album or even musical genre. All these are still editorial information, which aren't necessarily the most useful to the collector. Then, when the music is downloaded, the album consists of a group of compressed audio files, containing preset meta-tags, again storing editorial information. When browsing these files in his audio player, the songs are defined and classified automatically, not always according to the collector's desires. His final attempt is then to create a set of folders on his disk, and arrange his items in these folders. But how does he name these folders? What if he wants to arrange and browse the items in multiple ways? What if a particular item doesn't fit in any folder, or could be placed in two or three different categories?

As we see from this example, the tools that the everyday user has at hand are too formal, and are poorly adapted to the growing activity of collecting multimedia contents. Indeed, what we have said for music can also be said for the other kinds of media, and can also be said for information research, file sharing, etc.

Attempts have been made at putting the human user back in control of the collecting process, rather than relying purely on predefined categories and automated research algorithms.

However, it has become obvious that the other extreme of handing complete control over to the user isn't optimal either. Let us take a look at online content sharing sites, such as the famous FlickR™. There is no categorization here, but there are three main strategies when looking for photos: date, location, and tags. The first two are self-explanatory, but the tags are more interesting here. When someone uploads a photo to the website, they can link a certain number of keywords, called tags, to this photo. Then, we can either browse through the most popular tags, or type a tag into a textbox for a more precise search. The users then have complete freedom on the way they choose to define their photos. But the problem is that many photos aren't tagged, and the photos that are, often have poorly named tags, making them difficult to retrieve. Therefore, we believe that an optimal solution to the problem of digital collections could lie somewhere between these two polarities: predefined categories and total user creativity.

B. Artists and philosophers' fascination for collection regimes

As a matter of fact, artists and philosophers have always been fascinated by the rebellious nature of collections and have demonstrated this in their own way ([2], [20]).

Here, for example, is the analysis of [20] on the status of excess in a collection: *"Excess in a collection does not mean disordered accumulation; it is a fundamental principle: for a collection to exist as such-in the collector's eyes the number of objects must exceed the physical possibilities of exposing and storing the entire collection at home. Therefore, someone who lives in a studio can have a collection: it is only necessary for him to have at least one work he cannot hang in his studio. That is why the reserve is an integral part of collections. Excess also applies to the capacity of memorization: for the collection to exist, it is necessary for the collector not to be able to remember all the works he owns. In fact, the number of objects he owns must be so important that it becomes too important, so that the collector can forget one of them or leave a part of his collection outside of his home. To sayd it ifferently, for a collection to exist, the collector must not have full control over his collection anymore".*

And also: *"The scene of a collector is not his own apartment, it's the world. The main part of his collection is not at his place — his collection is to be, still scattered across the world, and every gallery and every fair is a way for him to go and find his future collection"*

As far as [2] was concerned, he had a very personal view on the subject: *"The art of collecting is a form of practical recollection, and, of all the profane manifestations of proximity, it is the most convincing. Everything that is present to memory, to thought, to consciousness, becomes a base, a frame, a pedestal, a casket for the object possessed".*

And also: *"What is decisive, in the art of collecting, is to free each object from its primitive functions, in order to establish a relationship as close as possible with similar objects. This relationship is diametrically opposed to usefulness, and belongs to the remarkable category of completeness. What is that completeness? An imposing attempt to go beyond the absolutely irrational nature of the simple presence of the object in the world, by integrating it in a new historical system, especially created, that is the collection".*

Thus, collections strongly differ from class, series, set, group, heaps, cluster, juxtaposition, accumulation, but also from organic whole/family.

C. Collections, between order and disorder

Until recently, a trend was mobilizing computers for the organization of our collections, considered like a group of

objects waiting to be organized in ad hoc classes that must be created simultaneously ([4], [9], [10]).

Because our collections seem to be nearer to order than disorder, attempting to assimilate them in classes is not so surprising. At least, collections look like they are waiting for their completion within a classification order, with the aim of turning into canonic achieved structures made of objects and classes. But something is also resisting this assimilation, as artists and philosophers have always noticed.

Undoubtedly impressed by artists and philosophers who considered the strange status of collections, computer program designers understood that computer modeling of object collections would necessarily involve the creation of hybrid structures including private characteristics – by which the collected objects are usually referred to – but also including characteristics that come from the activities in which these objects collectively engage.

Often, the approach implicitly chosen to characterize a collection is parsimonious and consists of over-determining the private referencing of the collected objects through a minimal description of the collective activity's context, even if it means predicting that the collection shall become a class or set of classes.

This practice presents the unquestionable advantage of not fundamentally opposing the traditional modeling of objects. However, it does not always live up to the collectors' high standards.

Here it is important to distinguish between figural and non-figural collections. This subtle distinction, introduced in the 1970s by Piaget and his research teams of child psychologists [12], brings more light to the situation. If it is certain that (non-figural) collections that adapt well to the aforementioned parsimonious approach exist, it is because they are completely independent of their spatial configuration. In that, they are already close to classification, of which they can only envy the formal completeness. On the other hand, there are collections we can label as *figural* because both their arrangement in space and the private properties of the collected objects determine their meaning.

D. Collections versus classes

In their book *La genèse des structures logiques élémentaires,* Jean Piaget and Bärbel Inhelder ([11], page 25) provide a precise distinction between figural and non-figural collections, which are still called classes or categorical collections. For the authors, a class requires only two categories or relationships, both necessary and sufficient, for its actual definition as a class:

1. The qualities common to its members and to those of the classes it belongs to, as well as the specific differences that distinguish its own members from the members of other classes (comprehension);
2. The relationship of a part to the whole (membership and inclusion) determined by the quantifiers "all", "some" (including "one") and "none" applied to the members of the class in question and to other members of the classes it belongs to, defined as extensions of that class.

For example, cats share in common several qualities owned by all cats, some being specific and some others belonging also to other animals. But no spatial considerations ever enter into such a definition: cats may be grouped or not in the space without any change concerning their class definition and properties (1) and (2).

Piaget then introduces figural collections, in which meaning defined by properties (1) and (2) is linked to the spatial arrangement of its elements. He claimed that: *"A figural collection composes a figure, through the spatial relationships between its elements, whereas non-figural collections and classes are free of any figure"*.

E. Figural versus non-figural collections

It is precisely these figural collections that computing is promising more and more an effective modeling of, pushed by an ever-growing social demand for on-line digital media browsing and information research amongst multiple sources.

But as we now understand, figural collections adapt poorly to their assimilation into non-figural collections or classes. Although according to Piaget, collections are destined to become classes, in the same way as subjects will grow psychologically so as to improve their cognitive capacity to classify. Still referring to [11]: *"It is a radical lack of differentiation that nudges figural collections out of the classical modeling field"*.

So classical IT approaches were unable to address and tackle the target of figural collections modeling.

On that particular point, we do not agree with Piaget, considering that, with the support of Inductive Data-Driven Algorithms and Big Data, even non-figural collections are about to be computerized.

IV. How Could IDDA & Big Data Help Supporting Our Figural Collections Deployment?

A. ReCollection: an experimental software for the creation of multimedia collections

ReCollection is a computer program for searching, arranging and browsing digital content, developed by Francis Rousseaux, Alain Bonardi and Benjamin Roadley [15].

As our collecting activities vary from one context to another, it is too ambitious to seek a general solution to the problem. Rather, particular application areas must be defined and isolated, in order for a specific answer to be given, however always relying on a set of basic principles. Here, we shall discuss the software prototype we have created for the digital opera/open form opera Alma Sola (designed by Alain Bonardi [3], and first performed at *Le Cube,* Issy les Moulineaux, October 2005).

1) The reserve

The ReCollection software has two main modes: reserve and gallery. The reserve allows us to store our objects that aren't exposed in the gallery. There are many objects in the reserve, and these are not always labeled; also they are rarely arranged in an orderly and tidy manner. So when we visit the reserve, we have no choice but to wander around,

picking up objects, inspecting and identifying them one at a time.

Fig 1. The reserve

The reserve can also be compared to the attic, in which our family possessions are stored similarly. As we explore our attic, we can happen to pick up an old photo album, which we had completely forgotten about. This item will surely bring back memories and emotions. We can then choose to keep this album under our arm, as we continue to explore the attic, or we can leave straight away, and put it on our fireplace, for example, making it visible to visitors. It is all these pleasant and familiar experiences, which we believe can be recreated thanks to the modeling of the reserve in our computer program.

The user can create any number of reserves. However, he must create at least one, and store at least one object in this reserve. When he is in reserve mode, he can only view one object at a time. When he decides to view another object, it is chosen randomly from the remaining items in reserve. During a visit, each object is viewed only once. If the user wants to view an item he has already visited, he may go through the history of items on the left side of the screen, as shown in figure 2. When he finds an object of interest, he can move it to the gallery. It will then be removed from the reserve, and saved in memory, with a group of objects waiting to be imported in the gallery. Then, in gallery mode, the user will see this heap of objects, and will be able to import it in the desired gallery, at the desired location.

2) The objects

The items in the Alma Sola collection are made up of three components:

1. A photo of the performance;
2. A sound recording of a few seconds of the singing;
3. A text: the line that is sang in the corresponding sound file.

These are all regular files stored on disk (bitmap, wave and .txt formats). Each item also has a name. In a more general context, the objects can be made up of any one of these types of media, a video (though not implemented in this version), or any combination of these.

Also, each object has a set of descriptors attached. There is a specific set of descriptors for each type of media, which describe the contents of the object, for example the average volume of the sound, the brightness of the photo, the number of words, etc. Depending on the application, we could also include editorial information, such as date, author, etc.

These descriptors may be assimilated to the private properties of traditional computer objects. But in the context of collecting objects, we also need to account for other properties that come from the activities in which these objects collectively engage.

3) The gallery

A collective activity involving a number of objects at a time is their relative arrangement in the gallery space. To the location of objects in this space, we have added their color; these two properties make up an extra conceptual layer, which is the framework for the creation and management of our collections.

In ReCollection, there is always at least one gallery, and the user can create as many as he wishes. There is always at least one item in a gallery, some basic content that the user can interact with, a starting point for his collection.

The objects can be placed and arranged manually in the gallery space, using click and move, just as in common user interfaces. The user can also rely on two algorithms to automatically dispose the objects. The first one, inspired by cataRT software [18], calculates the objects' positions and colors according to descriptors chosen by the user. The second calculates the positions depending on a sample of objects selected by the user. A Principal Components Analysis (PCA) finds out which descriptors vary most amongst the objects of the sample, the system can then rearrange the whole gallery according to these descriptors, as in the first method.

The arrangements resulting from the algorithmic calculations can always be modified manually in order to correct them (in the eventuality of rather subjective descriptors), to build up a global figure, or to bring items together. This way, through creative human-computer feedback loops, meaningful global figures can emerge through the arrangement in space of collected items, as well as local figures, soft pseudo-categories which are heaps of objects brought together by the system and/or the human user. These pseudo-categories are the building blocks for the classes the collection is implicitly aiming for. They are easily and constantly updated; items are added and removed instantly by being moved in space. They are loosely defined and never completely closed off from others, allowing some objects to be lost somewhere in between several heaps, when they cannot be placed in any one category. In a nutshell, this system allows for the creation of collections in which classes are in constant evolution, and are built by exploiting not only the objects' degree of similarity, but also their relative location in space and time.

Furthermore, the user may wish to search for objects in the gallery or in the reserve, in order to build on these categories, look for new kinds, or even fill in gaps in the gallery space. For this, the ReCollection system has two search tools

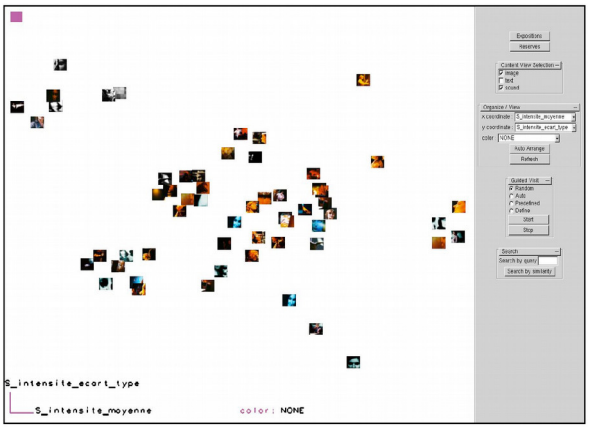


Fig 2. Gallery mode

he can use. The first is a simple 'keyword query', which searches for a keyword within the text or names of the objects. The second is a 'search by similarity'. The user selects an object, or group of objects, and the system searches for items that are similar (according to the descriptors). In both cases, the search is carried out in both the gallery and reserve, and a list of results is displayed in the gallery, ordered by similarity.

B. ReCollection at the time were IDDAs & Big Data technologies are now available

Once all the items of interest have been imported from the reserve, through browsing or searching, and once they have been arranged in the gallery space, the user has a first disposition he can play with. When he will browse the gallery space, his experience will be influenced by the fact that certain objects are close in space, and in time of visitation. Although this is interesting in itself, the system can help the user go further, by defining a set of guided visits, which are simply an order of visitation of selected objects in the gallery.

The type of interface we have chosen to implement these functionalities is a 2D zoomable user interface, inspired by Ken Perlin's Pad [5]. All objects are in the same 2D space, which has no borders. The point of view can be moved vertically and horizontally, and the user can zoom in and out. If he zooms in on an item, until it fills the screen, the sound is played back. This kind of interface has been experimented: Its intuitive approach is seducing to us, particularly in our goal of intuitively collecting digital media. Finally, the spatial metaphor takes advantage of the users' spatial memory and cognitive abilities [7].

ReCollection has typically been designed for supporting figural collections deployment, as the time where IDDAs & Big Data technologies were not yet available. That is why our 'search by similarity' tools were mainly based on PCA.

Nowadays, we are developing some IDDAs to allow the collectors to search by new kind of similarities through remove virtual digital reserves, distributed along the Web.

V. Conclusion

Inductive Data-Driven Algorithms really allow us to explore differently digital 'Big Data', if we waive the deductive requirement and permit the inductive heuristics to apply.

But technically, IDDA is not easy to develop, and ethically, it could be dangerous for our decision autonomy because of the lack of traceability that characterizes those IDDA approaches, readily producing 'results' without any linked explanation.

However, we have discovered a 'Big Data' field where IDDAs are fully legitimate and powerful, namely: 'figural' collections constitution, content curation and deployment. This task is much more important for human beings that it looks like, and the lack of efficient tools to support that was clearly identified.

Incidentally, that discovery has disclosed some important Abstract Truth of Big Data: if we accept, against Piaget, that collecting is more native than classifying or categorizing but not less powerful and intelligent, Big Data is the reserve of our future personal digital collections.

References

[1] C. Anderson: "Makers: The New Industrial Revolution", New York: Crown Business, 2012.
[2] W. Benjamin: "Paris, capitale du XIXe siècle — le livre des passages", Le Cerf, 1989.
[3] A. Bonardi: "New Approaches of Theatre and Opera Directly Inspired by Interactive Datamining", Proc. Sound & Music Computing Conference, pp. 1-4, Paris, 2004.
[4] P. Brézillon: "Context in Human-Machine Problem Solving: a Survey", Knowledge Engineering Review, Vol. 14, pp. 1-34, 1999.
[5] D. Fox, and K. Perlin: "Pad: An Alternative Approach to the Computer Interface", Proc. ACM SIGRAPH, 1993.
[6] A. Gkoulalas-Divanis, Y. Saygin, and V. Verykios: "Special Issue on Privacy and Security Issues in Data Mining and Machine Learning", Transactions on Data Privacy, Vol. 4, Issue 3, pp. 127-187, December 2011.
[7] L. Hasher, and R. T. Zacks: "Automatic and Effortful Processes in Memory", Journal of Experimental Psychology, 1979.
[8] E. Morozov: "To Save Everything, Click Here: Technology, Solutionism, and the Urge to Fix Problems that Don't Exist", FYP Editions, 2014.
[9] F. Pachet, J-J. Aucouturier, A. La Burthe, A. Zils, and A. and Beurive: "Multimedia Tools and Applications," Special Issue on the CBMI 03 Conference, 2006.
[10] F. Pachet: "Content Management for Electronic Music Distribution: The Real Issues", Communications of the ACM, 2003.
[11] J. Piaget, and I. Bärbel: "La genèse des structures logiques élémentaires", Delachaux & Niestlé, 1959.
[12] J. Piaget: "La psychologie de l'enfant", PUF, 2012.
[13] C. S. Pierce: "Pragmatism – The Logic of Abduction", Collected Papers, Vol. 5, pp. 196, 1903.
[14] K. Pomian: "Collectionneurs, amateurs et curieux", Gallimard, 1987.
[15] F. Rousseaux, A. Bonardi, and B. Roadley: "ReCollection: a Disposal/Formal Requirement-Based Tool to Support Sustainable Collection Making," Proc. ICCS Supplement, pp. 131-138, 2008.
[16] A. Rouvroy: "The end(s) of Critique: Data-Behaviourism vs. Due-Process", Privacy, Due Process and the Computational Turn — Philosophers of Law Meet Philosophers of Technology, Chapter 5, Mireille Hildebrandt & Ekatarina De Vries (eds.), Routledge, 2012.
[17] A. Rouvroy: "Le régime de vérité numérique — de la gouvernementalité algorithmique de fait au nouvel état de droit qu'il lui faut", séminaire Digital Studies de l'Association Ars Industrialis, 7 octobre 2014.
[18] D. Schwarz, G. Beller, B. Verbrugghe, and S., Britton: "Real-Time Corpus-Based Concatenative Synthesis with CATART", Proc. DAFx, 2006.
[19] B. Slavkovic, and A. Smith: "Special Issue on Statistical and Learning-Theoretic Challenges in Data Privacy", Journal of Privacy and Confidentiality, Vol. 4, Issue 1, pp. 1-243, 2012.
[20] G. Wajcman, "Collection", Nous, 1999.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1349–1354
DOI: 10.15439/2016F245
ACSIS, Vol. 8. ISSN 2300-5963

Churn Detection and Prediction in Automotive Supply Industry

Hasan Can Karapinar
Istanbul Technical University,
Faculty of Management, 34357,
Macka, Istanbul, Turkey
Email:
hasancankarapinar@gmail.com

Ayca Altay
Istanbul Technical University,
Faculty of Management, A304,
34357, Macka, Istanbul, Turkey
Email: altaya@itu.edu.tr

Gulgun Kayakutlu
Istanbul Technical University,
Faculty of Management, B309,
34357, Macka, Istanbul, Turkey
Email: kayakutlu@itu.edu.tr

Abstract—**Companies have both large certified enterprises and small unauthorized service providers as their competitors in the automotive supply industry. As technology related industries undergo more intensive competition, churn detection and prediction become essential to be precautious about leaving customers. The literature for churn detection offers numerous statistical and intelligent methods. In this study, Artificial Neural Networks and Decision Trees are applied to detect the churn in and analyze the validity of these methods for the automotive supply industry. The problem involves both categorical and continuous numerical decision inputs which cannot simultaneously fed into Decision Trees. In this case, continuous inputs should be divided into binary categorical ones by splitting into various intervals which are called buckets. Particle Swarm Optimization algorithm is implemented for finding optimal buckets for the churn problem data. Results indicate that while both algorithms are promising, the bucket tuning for Decision Trees complicate the churn detection process.**

I. Introduction

Retaining existing customers has become the main objective of Customer Relationship Management (CRM) departments in markets with fierce competition [1]. Automobile service industry stands for a real example when fierce competition is mentioned in Turkey. Maintenance and hard repairs are the most profitable cases in automobile service market, and every one of the authorized and non-authorized services are trying to take a big pie of this market. Developed subsidiary industry of automobile spare parts and low labor costs of non-authorized services are challenging automobile companies to retain their customers, making churn prediction an important part of this challenge. Retaining a customer is more profitable than acquiring a new one so that companies are willing to know which customer will churn like in other markets. Preventive actions, new marketing strategies and long term service contracts can be made in case of foreknow the churners. Churn prediction must be as accurate as possible to ease decision making. It is known that the cost of earning new customers is a lot more than sustaining existing ones. Hence, it becomes essential to select the best churn detection methods given company data [2].

In literature, there are various methods for data mining in churn prediction; yet, the common point of these methods are that they yield a classification and/or clustering decision where the customers are grouped as "churned" and "not churned" [3]. The methods applied for grouping customers for churn can mainly be subsumed under statistical, intelligent and hybrid methods. Statistical methods involve approaches that intend to minimize incorrect classification using statistical data processing tools such as Decision Trees or Regression [4]. Learning algorithms and intelligent methods make up the recent trend with the contribution of computer technologies in churn mining where methods such as Neural Networks and metaheuristics are implemented [5]. Lastly, hybrid methods that combine statistical and learning methods are stated to provide more accurate results [5]. These methods will be elaborated in the next section.

The literature of churn prediction is intensified in the service industry; especially in telecommunication and banking [3, 6-8]. Emerging from these industries, churn prediction is now known to contribute to companies in other industries, as well. In this study, we have applied two different data mining techniques on the problem of churning customers in the automotive supply industry. Artificial Neural Network and a hybrid method (Particle Swarm Optimization (PSO) based ID3 Decision Trees)s are used for determining the churn and their accuracies were compared in order to provide a better detection performance.

The structure of this paper is as follows: The next section provides a literature review on churn determination and prediction together with the motivation of this study. The third section introduces the methodology and an application is provided in the following section. Finally, the conclusions are presented in the last section.

II. Literature Review

Churn analysis and prediction aims to detect churn customers and/or predicting how likely the customers are to be churned [3]. The literature of churn analysis is intensified in industries where high rate of competition exists; such as

This work was not supported by any organization

the telecommunication industry where the churn rate reaches its peak with an annual average of 27% [9].

The implementation of data mining techniques in churn analysis is inevitable, since the extraction of patterns in churned customers is essential for determining and predicting churn [10]. An analytic view of the methods in churn analysis can be analyzed three-fold: statistical, intelligent and hybrid methods. The most exploited method in statistical churn analysis is Decision Trees and Random Forests [4]. Ahn et al. [11] use Factor Analysis in telecommunication industry and state that customer status is the most important criteria. Bin et al. [12] implement Decision Trees in phone sales. Wei and Chiu [3] implement decision trees that extract rules with C4.5 algorithm for churn prediction. Lariviére and Van del Peol [13] implement logistic and linear regression models and extend the Decision Tree approach and execute random forests in telecommunication industry. The implementation of decision trees is also diversified. For instance, Lemmens and Croux [14] use Bagged & Boosted Decision Trees in Telecommunications Industry.

The execution of intelligent and learning methods have been the focus of literature of the past two decades. In particular, Artificial Neural Networks (ANNs) methods, both supervised and unsupervised, represent the majority of studies in this branch. Hadden et al. [5] implement basic Back Propagation algorithm on Feed-Forward Neural Networks. Buckinx and Van den Poel [15] compare the performance of Random Forests and ANNs and conclude that the prediction by ANNs is more robust in case of FMCG industry example. Other variations of ANNs involve Support Vector Machines (SVMs) and Self Organizing Maps (SOMs). Coussement and Van den Poel [16] construct and SVM model to predict churn customers in telecommunications industry, whereas Tsai and Lu conduct a similar study where they implement an unsupervised approach by SOMs [17]. Even though, ANNs are known to be stronger churn predictors than Decision Trees, they are known to have disadvantages such as early convergence or being stuck at local optima [5]. Hence, these models are enhanced with other metaheuristics or Artificial Intelligence methods. For example, Karahoca and Karahoca [18] implement an Adaptive Neuro Fuzzy Inference System model integrated with a fuzzy clustering method which highly outperforms ANNs.

Hybrid methods are known to provide better comprehension of the churn structure. Xia and Jin [19] first find out the criteria that affect churn using Factor Analysis and using these criteria, they build an SVM model in order to predict churn. Idris et al. [9] find the optimum decision tree using Particle Swarm Optimization and conclude that metaheuristic-based decision trees outperform in terms of classification accuracy.

In this study, one main statistical method - Decision Trees - and one main intelligent method - ANNs - are tested on automotive supply chain determination. The very basic ID3 algorithm is used for Decision Trees. However, the ID3 (Interactive Dichotomizer 3 (ID3) algorithm for Decision Trees only operate when all variables are categorical and the data on hand involves continuous numerical variables. Hence a separation scheme is mandatory. The optimum separation points are found by Particle Swarm Optimization algorithm, yielding a hybrid method. A comparison of a hybrid and an intelligent method is constructed in terms of automotive supply industry perspective.

The factors used for classification in literature are observed to depend industry-wise. Telecommunication industry related studies - GSM churn, to be concise - are mainly focused on minutes of usage, monthly number of calls or monthly bills, which are specific to this industry [3]. Hence, it is essential to find out the criteria that are specific to the industry. However, the literature on automotive supply industry does not offer many criteria; which is why, mainly interviews with industry experts have determined the criteria for this study.

III. METHODOLOGY

Decision Trees and ANNs will be compared for this study, since they are the basic decision making.

A. Decision Trees

The basic ID3 Algorithm is used for Decision Trees. Proposed by Quinlan [20], the algorithm uses information gain in order to decide the splits or branches of the tree, which is calculated as given below:

$$IG(S) = \sum_{k=1}^{N} -p(k)\log_2 p(k) \quad (1)$$

where S is an element of the input attribute set I, N is the number of unique attribute values, $p(k)$ is the probability of this attribute value k [20-21].

For each node, the attribute that provides the most information gain is split until final branches (called leaves) are reached. The algorithm builds a decision tree from the data which are discrete in nature. The data in the case of automotive supply industry churn involves continuous variables such as mileage. In this case, the data of continuous attributes should be discretized. One way to handle such continuity problems is to assign separation or split points for the data and convert them into binary variables. This process is called “splitting into buckets” where each bucket is a class that specifies a range of values and a data point is subsumed under one of these ranges or buckets [22]. For example, if the range of a continuous variable is [0,10] and the separation point is 5, then, the buckets are [0,5] and [5,10]. A value, say 3, is classified in a binary fashion as 1 and 0, for two buckets, respectively. Moreover, the selection of separation points affects the structure of the input data, and hence, largely impacts the accuracy of the decision tree. Therefore, these separation points should be tuned. This tuning is achieved by the PSO algorithm.

B. PSO Algorithm

The PSO algorithm offered by Kennedy and Eberhart [23]. The steps of the conventional PSO algorithm are given below:

Let p_i be the i^{th} particle in the swarm which consists of N particles and let each particle have n variables.

Step 1. The particle velocities and positions are initiated as priorities

$$x_{i,j} = x_{min} + r(x_{max}\text{-}x_{min}), \quad i = 1, \dots, N, j = 1, \dots, n \quad (2)$$

$$v_{i,j} = \alpha \frac{x_{min} + r(x_{max}\text{-}x_{min})}{\Delta t} \quad i = 1, \dots, N, \ j = 1, \dots, \quad (3)$$

where x denotes the position, v denotes the velocity and is constant in the range [0,1].

Step 2. The objective function values of particles are calculated as $f(x_i)$.

Step 3. The best position for each particle and the global best position for the swarm are updated.

$$\text{If } f(x_i) < f\left(x_i^{pb}\right) \text{ then } x_i^{pb} \leftarrow x_i \quad (4)$$

$$\text{If } f(x_i) < f\left(x_i^{sb}\right) \text{ then } x_i^{sb} \leftarrow x_i \quad (5)$$

where pb denotes the particle best and sb denotes the swarm best.

Step 4. Particle velocity and particle position are updated, that is, the new velocities and positions are calculated for each particle.

$$v_{i,j} \leftarrow wv_{i,j} + c_1 r_1 \left(\frac{x_{i,j}^{pb}\text{-}x_{i,j}}{\Delta t}\right) + c_2 r_2 \left(\frac{x_{i,j}^{sb}\text{-}x_{i,j}}{\Delta t}\right), i = 1, .. \quad (6)$$

where $wv_{i,j}$is the separation term, $c_1 r_1 \left(\frac{x_{i,j}^{pb}\text{-}x_{i,j}}{\Delta t}\right)$is the alignment term and $c_2 r_2 \left(\frac{x_{i,j}^{sb}\text{-}x_{i,j}}{\Delta t}\right)$ is the cohesion term.

Step 5. Step 2 is returned until a termination criterion is satisfied.

C. ANNs

ANN models are basically derived from nervous systems which are able to perform functional input-output mapping such as machine learning and pattern recognition [24-25]. In this study, feed forward neural networks with back-propagation learning algorithm are utilized.

ANNs are constructed by layers of cells which are called neurons as can be seen in Figure 2. Neurons connect to other neurons in consecutive layers by a certain weight value. Inputs of each neuron are the weighted sum of the weights of incoming neural connections and a bias. This sum is subjected to a transfer or an activation function within the neuron [23]. The input of a neuron is calculated as

$$n = \sum_i w_i \cdot x_i + bias \quad (7)$$

where n is the input of a neuron, w_i is the weight of the i^{th} neuron in the previous layer and x_i is i^{th} input value.

The Back-propagation algorithm is a learning structure that minimizes the forecasted output of the network and the actual output by adjusting weights of the network. In order to achieve this error minimization, the data are first fed into the network and the error term obtained is propagated back into the network to adjust weights with formula given below

$$\Delta w_{ji}^k(n+1) = \text{-}\eta \frac{\partial E}{\partial w_{ji}^k} + \Delta w_{ji}^k(n) \quad (8)$$

where n is the index for iterations, η is the learning rate and E is the error term. Note that the derivative of the error term with respect to the weights yields the weight adjustment.

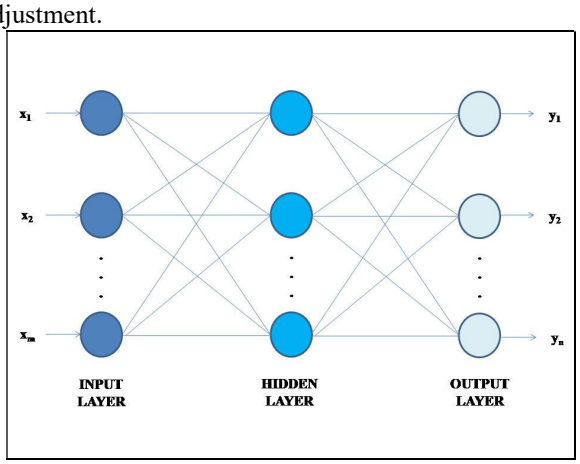


Fig. 2 The structure of an ANN

IV. APPLICATION AND RESULTS

In the last decade, there has been an increase in fleet customers in automotive market. Instead of buying a new car, companies started to rent fleet of cars. So that, detecting a fleet car has become critical because detecting one of the fleet car's churning means losing all of the cars and profit from that fleet.

In this study, the churn determination of customers of a leading car seller of Turkey is achieved using a PSO-based ID3 Decision Tree and ANNs. The criteria for churn evaluation are extracted from the interviews with industry experts and company's academic consultants.

The criteria for churn evaluation are listed below:

1. Annual penetration of 2013: It represents the total amount of cash inflow that customer did throughout the year 2013. It is a continuous attribute; hence, it should be bucketed.
2. Annual penetration of 2014: It represents the total amount of cash inflow that customer did throughout the year 2014. It is a continuous attribute and it should be bucketed as in the previous year's case.

3. Annual penetration of 2015: It represents the total amount of cash inflow that customer did throughout the year 2015. Just like the previous years, it is bucketed.
4. Annual frequency of maintenance of 2013: It represents the number of times that the car was taken to maintenance services throughout the year 2013. It is a discrete variable.
5. Annual frequency of maintenance of 2014: It represents the number of times that the car was taken to maintenance services throughout the year 2014. It is a discrete variable.
6. Annual frequency of maintenance of 2015: It represents the number of times that the car was taken to maintenance services throughout the year 2015. It is a discrete variable.
7. Year of purchase: It represents the year that the car was purchased by its owner. It is a discrete variable.
8. Insurance: It represents if the car is insured. It is a binary discrete variable.

Discretization of the continuous variables is achieved by splitting into buckets for the first three criteria; hence, PSO algorithm should be applied in order to find the optimum separation points. However, the optimum number of buckets is not known a priori; and thus, number of buckets are found by trial and error. The splits for bucket construction of 2, 3, 4 and 5 buckets are tuned by the PSO algorithm.

The performance measure (objective function) of the algorithm is taken as the rate of misclassifications. There are a total of 500 data points and their churn conditions are known. Two sample data points are shown in Fig. 3. The first customer is a churn and the second customer has not churned. 70% of the data are used for training and the rest is used for testing. PSO parameters are adjusted as 20 for the swarm size, 1 for the inertia coefficient (w), 2 for the cognitive coefficient (c_1) and 2 for the social coefficient (c_2). The learning rate (η) of the ANN is 0.7. Table 1 shows the average and the standard deviation of 30 runs of training of the PSO-hybrid and ANN results, respectively. Table 2 shows the same results for the testing data.

PSO-based Decision Trees yield the best result when the number of buckets is 2 for the training data and the testing data. Additionally, the PSO-hybrid with 2 buckets has one run without any error for the training data; whereas the ANN has one run without any error for testing data. In order to analyze if the results of these 30 runs are statistically significant, t tests among the algorithm results is applied. Tables 3 and 4 display the p values of the t test results of methods in a pairwise comparison fashion. The lower parts of the tables are not filled because the tables are symmetrical. For example, the value that is in the intersection cell of PSODT2 and PSODT3 also applies to the value of the intersection cell of PSODT3 and PSODT2.

TABLE I.
ALGORITHM RESULTS FOR TRAINING DATA

Method	Average misclassification Percentage (μ)	Standard deviation of the misclassification percentage (δ)
PSO – based Decision Tree with 2 buckets (PSODT2)	0.086	0.12
PSO – based Decision Tree with 3 buckets (PSODT3)	0.118	0.05
PSO – based Decision Tree with 4 buckets (PSODT4)	0.150	0.07
PSO – based Decision Tree with 5 buckets (PSODT5)	0.114	0.04
ANNs	0.109	0.02

TABLE IIII.
ALGORITHM RESULTS FOR TESTING DATA

Method	Average misclassification Percentage (μ)	Standard deviation of the misclassification percentage (δ)
PSODT2	0.154	0.05
PSODT3	0.170	0.03
PSODT4	0.167	0.05
PSODT5	0.198	0.04
ANNs	0.160	0.03

TABLE II.
SAMPLE DATA POINTS

	Annual Penetration			Annual Maintenance Frequency					
Rack Number	2013	2014	2015	2013	2014	2015	Year of purchase	Insurance	Churn
XXXXXXXXXX	0	478,546	647,431	0	1	1	2013	1	1
YYYYYYYYY	0	12,37	0	0	1	0	2012	0	0

TABLE IV.
P VALUES OF T TESTS FOR TRAINING DATA

	PSODT2	PSODT3	PSODT4	PSODT5	ANNS
PSODT2	-	0.2113	0.0640	0.7335	0.3047
PSODT3		-	0.0340	0.8648	0.4793
PSODT4			-	0.0175	0.0031
PSODT5				-	0.5427
ANNs					-

TABLE III.
P VALUES OF T TESTS FOR TESTING DATA

	PSODT2	PSODT3	PSODT4	PSODT5	ANNS
PSODT2	-	0.1383	0.3181	0.0004	0.5752
PSODT3		-	0.7791	0.0330	0.2018
PSODT4			-	0.0103	0.5134
PSODT5				-	0.0001
ANNs					-

The less the p value is, the more statistically different the results of the two methods compared. For example, for the training data, the p value between PSODT2 and PSODT3 is 0.0640. If a significance level of 10% is assigned, the given p value is less than 0.1, that is, the results of PSODT2 and PSODT3 are statistically significantly different. Moreover, from Table 1, it can be seen that the misclassification error rate of PSODT3 is larger, which means that PSODT2 provides better results that PSODT3. According to Table 3, for the training data, apart from PSODT4, other algorithm results are not statistically significantly different from each other. In terms of the testing data (see Table 4), PSODT5 is the worst result providing algorithms and other algorithm results are similar. Using these results, the non-dominated algorithms are PSODT2, PSODT3 and ANNs. ANNs provide better results for both training and testing data. Likely, the results of PSODT2 and PSODT3 are also as good; yet, in terms of Decision Trees, the bucket numbers have to be optimized. Moreover, considering the accuracy of the models, it is concluded that the criteria determined by industry experts are valid for churn detection.

V. CONCLUSION

Churn detection is difficult in industries where the competition is fierce. The churn is intensive in especially technology related industries. Even though not being as technologically intensive as online markets or telecommunication industries, automotive supply industry is also one of the industries that suffers from frequent churn. Therefore, churn detection and prediction becomes essential.

In this study, ANNs are constructed for churn prediction for automotive supply industry. Traditional churn approaches also offer the implementation of Decision Trees to be a powerful means of churn analysis. However, in the presence of non-categorical (numeric – continuous) data, the implementation of Decision Trees with raw data may be misleading. For this reason, it is obligatory that the continuous data are converted to binary and categorical data. This process is achieved by using 'buckets'. PSO algorithm is used for optimizing the number of buckets and the split values of the data. Once the data is processed into buckets, it can be fed into the decision tree. The results indicate that while both methods can be equally strong, the number of buckets is an important factor for determining the optimum decision tree.

Future studies that aim misclassification error reduction may involve improved ANN techniques such as dynamic networks, etc. Hybridization of decision trees with other metaheuristics and changing the decision tree structure are also worth analyzing. Additionally, different churn data would be helpful for the validation of the results. This study does not involve temporal predictors of churn; these predictors should be integrated into the churn model.

REFERENCES

[1] M. Xu and J. Walton, "Gaining customer knowledge through analytical CRM", *Industrial Management & Data Systems*, vol. 105, no. 7, pp.955 - 971, 2005.

[2] A. Niedziółka, "Management of Agritourism in the Sustainable Development of Rural Areas with the example of the Maiopolskie Voivodeship", *The Journal of Management and Sustainable Development*, vol. 17, pp. 70-75, 2007.

[3] C.P. Wei and I.T. Chiu, I.T., "Turning Telecommunications Call Details to Churn Prediction: A Data Mining Approach", *Expert Systems with Applications*, vol. 23, no. 2, pp. 103-112, 2002.

[4] W. Verbeke, K. Dejaeger, D. Martens, J. Hur and B. Baesens, "New Insights into Churn Prediction in the Telecommunication Sector: A Profit Driven Data Mining Approach", *European Journal of Operational Research*, vol. 218, pp. 211–229., 2012.

[5] J. Hadden, A. Tiwari, R. Roy and D. Ruta, "Churn Prediction: Does Technology Matter?", *International Journal of Intelligent Technology*, vol. 1, no.2, p. 104, 2006.

[6] W.H. Au, K.C.C. Chan and X. Yao, "A Novel Evolutionary Data Mining Algorithm With Applications to Churn Prediction", *IEEE Transactions On Evolutionary Computation*, vol. 7, no. 6, pp. 532-545, 2003.

[7] I. Khan, I. Usman, T. Usman, G. Ur Rehman and A. Ur Rehman, "Intelligent Churn prediction for Telecommunication Industry", *International Journal of Innovation and Applied Studies*, vol. 4, no. 1, pp. 165-170, 2013.

[8] Y. Xie, X. Li, E.W.T. Nigai and W. Ying, " Customer churn prediction using improved balanced random forests", *Expert Systems with Applications*, vol.36, no. 3, pp. 5445-5449, 2009.

[9] A. Idris, M. Rizwan and A. Khan, "Churn Prediction in Telecom Using Random Forest and PSO Based Data Balancing in Combination with Various Feature Selection Strategies", *Computers & Electrical Engineering*, vol. 38, no. 6, pp. 1808-1819, 2012.

[10] T. Vafeiadis, K.I., Diamantaras, G. Sarigiannis and K.Ch. Chatzisavvas, " A comparison of machine learning techniques for customer churn prediction", *Simulation Modelling Practice and Theory*, vol. 55, pp.1-9, 2014.

[11] J.H. Ahn, S.P. Han and Y.S. Lee, " Customer churn analysis: Churn determinants and mediation effects of partial defection in the Korean mobile telecommunications service industry", *Telecommunications Policy*, vol. 30, pp. 552-568, 2006.

[12] L. Bin, S. Peiji and L. Juan, "Customer Churn Prediction Based on the Decision Tree in Personal Handyphone System Service", *International Conference on Service Systems and Service Management*, Chengdu, China, June 9-11, pp. 1-5, 2007.

[13] B. Lariviére and D. Van den Poel, " Predicting customer retention and profitability by using random forests and regression forests

techniques", Expert Systems with Applications, vol. 29, no. 2, pp. 472-484, 2005.

[14] A. Lemmens and C. Croux, "Bagging and boosting classification trees to predict churn", Journal of Marketing Research, vol. 43, no. 2, pp. 276-286, 2007.

[15] W. Buckinx and D. Van den Poel, " Customer base analysis: partial defection of behaviorally-loyal clients in a non-contractual FMCG retail setting", European Journal of Operational Research, vol. 164, pp. 252-268, 2005.

[16] K. Coussement and D Van den Poel, " Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques", Expert Systems with Applications, vol. 34, no. 1, pp 313-327, 2008.

[17] C. F. Tsai and Y. H. Liu, " Customer churn prediction by hybrid neural networks, Expert Systems with Applications, vol.36, no. 10, pp. 12547-12553, 2009.

[18] A. Karahoca and D. Karahoca, "GSM Churn Management by Using Fuzzy C-Means Clustering and Adaptive Neuro Fuzzy Inference Skkkkystem", Expert Systems with Applications, vol. 38, no. 3, pp. 1814–1822, 2011.

[19] G. E. Xia and W. D. Jin, "Model of Customer Churn Prediction on Support Vector", Machine. Systems Engineering – Theory and Practice, vol. 28, no.1, pp. 71-77, 2008.

[20] J. R. Quinlan, Induction of decision trees", Machine Learning, vol. 1, no. 1, pp. 81-106, 1986.

[21] K. Gajowniczek, T. Zabkowski and A. Orlowski, "Comparison of Decision Trees with Renyi and Tsallis Entropy Applied for Imbalanced Churn Dataset", Proceedings of the Conference on

Federated Computer Science and Information Systems, 13-16 Sept., Lodz, pp.39-44, 2015.

[22] J. Du, R. He and Z. Zhechev, “Forecasting Bike Rental Demand”, CS 229 Machine Learning Project, Stanford University, 2014.

[23] J. Kennedy and R. Eberhart, “Particle Swarm Optimization”, ., IEEE International Conference on Neural Networks, vol. 4, pp. 1942–1948, Perth, WA, 27 November-01 December, 1995.

[24] S. Haykin, Neural Networks and Learning Machines, 3rd Edition, Prentice Hall, New Jersey, USA, 2008.

[25] K. Pytel, T. Nawarycz, W. Drygas, “Anthropometric Predictors of Artificial Neural Networks in the Diagnosis of Hypertension“, Proceedings of the Conference on Federated Computer Science and Information Systems (FEDCSIS), 13-16 Sept., Lodz, pp. 287–290, 2015.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1355–1358
DOI: 10.15439/2016F376
ACSIS, Vol. 8. ISSN 2300-5963

Spreadsheet-Based Business Process Modeling

Krzysztof Kluza and Piotr Wiśniewski
AGH University of Science and Technology
al. A. Mickiewicza 30, 30-059 Krakow, Poland
E-mail: {kluza,wpiotr}@agh.edu.pl

***Abstract*—Business Process models help to visualize processes of an organization. In enterprises, these processes are often specified in internal regulations, resolutions or other law acts of a company. Such descriptions, like task lists, have mostly form of enumerated lists or spreadsheets. We present a method how to generate a BPMN process model from a spreadsheet-based representation. In contrast to the existing approaches, our method does not require explicit specification of gateways in the spreadsheet, but it takes advantage of nested list form.**

***Index Terms*—Business Process Model and Notation (BPMN), process modeling, spreadsheets, spreadsheet-based modeling**

I. Introduction

PROCESS models constitute a useful knowledge representation. They are commonly used by organizations to depict the workflow of the company, especially to specify alternative flows of tasks and events. Such aspects are often specified using textual description in internal regulations, resolutions or other companies law acts. These descriptions consist of the specification of the steps taken to achieve the specific goal. Such steps can be easily specified using a spreadsheet or an enumerated list (an ordered list of steps can be almost directly transformed into a spreadsheet format).

In this paper, we present a method of generating BPMN process models from spreadsheet-based representation (see Fig. 1). A process can be described using one of the spreadsheet applications like MS Excel, Google Docs or OpenOffice Calc. Based on the CSV file of the model exported from the application, a graphical process model in BPMN can be generated according to the specified transformation rules.

According to the studies [1], up to 60% of the time spent on process management projects can be consumed by the acquisition of process models, which mostly is done manually by process designers or business analytics. Thus, generating or transforming the existing representation to models can shorten this time.

In the field of transforming some kind of process description into a process model, there are various research directions.

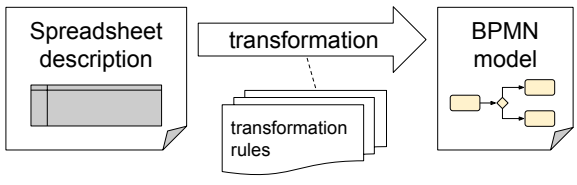


Figure 1. Overview of the spreadsheet-based approach

II. Related works

One of the existing methods is generating models from text description. A text can be provided in natural language [1] or in structured language. A part of the SBVR standard is SBVR Structured English [2]. There are methods of tranforming SBVR business rules into UML activity diagrams [3] (which are similar to BPMN models) or BPMN [4], [5]. Some papers consider the extended versions of SBVR, like SBPVR (Semantics of Business Process Vocabulary and Process Rules) [6] or sSBVRMM (simplified SBVR metamodel) [7]. There are also methods for analysis and validation of processes which use natural language generation [8] or spreadsheets [9].

Another method which provides good quality models is translation from other representation. An example of transformation approach is transformation of the Unified Modeling Language (UML) use case diagrams [10], [11]. The broad family of process model generation methods is process discovery, which is one of the process mining techniques. There are many existing algorithms, mostly implemented using the ProM process mining toolkit. Among them, there are algorithms to mine BPMN models [12].

The most similar approach to ours was presented by Krumnow and Decker in [13], [14]. They proposed three different approaches for business process modeling using spreadsheets:

1) Simple approach – this solution concerns simple processes modeled only using sequences of activities.
2) Branching approach – concerns not only sequences, but also more complex flow structures. It uses such elements like successors in order to represent complex control flow, as well as gateways and events. For the condition, the "Description" column is used. In the case of several successors, this can be realized by using a comma-separated string of row numbers as property value.
3) More Properties approach – extends the branching approach with additional specified explicitly properties like the assigned roles, input documents, etc. It allows for adding customized properties as new columns to the spreadsheet, which are not presented in the model.

The second and third approaches require from the user some kind of familiarity with business process modeling notation. According to these approaches, the user has to specified such elements like gateways or successors.

In our research, we propose a new approach, which solves the problem of familiarity with business process modeling notation. The solution is described in the following section.

Table I
SIMPLE EXAMPLE OF A XOR-GATEWAY IMPLEMENTATION

Order	Activity	Condition	(..)
4a	Send Ticket	Payment registered	
4b	goto 6	else	
5	...	...	
6	...	...	

III. TRANSFORMING SPREADSHEET INTO BPMN MODEL

The translation method presented in this section transforms a spreadsheet-based representation into a BPMN 2.0 model. In the following subsections, we describe the requirements and assumptions that we took into account during method development, the supported representation, as well as we provide the detailed description of transformations along with transformation procedure.

In order to make the solution widely applicable and simple to use, some assuptions were made. The following requirements need to be considered while providing a spreadsheet-based representation of a business model:

- The user should be able to create a business process model using his favourite popular spreadsheet application (e.g. MS Excel, Google Docs, OpenOffice Calc).
- A graphical model should be created using a CSV file (an exported spreadsheet).
- Only one pool is considered and the term "Who" is used instead of swimlane to distinguish different task executors.
- Logical gates are eliminated from the model. A process should be described as a set of phases. If two tasks are executed in the same phase, it means that there are parallel or connected by an inclusive or exclusive alternative (OR/XOR). The relationship between tasks is determined by a condition stored in a separate column (no condition – AND, two different conditions – OR, condition and „else" statement - XOR). Table I shows how logical gates can be represented.
- Loops are represented by "goto" tasks, which link one phase with another.

In the proposed solution, a business process is represented by a spreadsheet table, where rows correspond to tasks (or phases). Columns contain properties of the selected phase. Task and condition names should start with a capital letter, while all instructions are written in lowercase. Following column types are used:

- Order – number of the corresponding phase (starting from one). If two or more tasks are performed in parallel or alternatively, we use letters to distinguish different branches.
- Activity – name of the performed action or instruction. For example, if there is a need for a loop or skipping the phase, this field should be filled with a „goto" statement and number of the desired phase, e.g. „goto 7".
- Condition – a condition needed to perform the selected action. If the task executes every time this field should be left empty. In order to implement a XOR-gateway this field should be filled with an appropriate condition and the word „else" for the other task performed in the same phase. In order to implement an OR-gateway the fields mentioned before should be filled with two separate conditions.
- Who – a department or person that executes the task. This field corresponds to a swimlane in BPMN.
- Subprocess – information if the selected task contains a subprocess. If so, this field should be filled with „yes" and a sheet with a name of this subprocess should be created. Otherwise the field should remain empty.
- Terminated – information if the selected task terminates the process. If so, this field should be filled with „yes". Otherwise the field should remain empty.

Supported process elements and their spreadsheet model

1) Start events: In the proposed model an assumption is made, that a business process starts in one specified moment, which is called the „0 phase". First 3 rows of Table II present examples of none, message and timer start events.

2) End events: End events are represented by the statement „yes" in the column „Terminated". The field "Activity" can either contain the name of the last activity or remain blank. In 4th and 5th row of Table II the examples of none and message end events are presented.

3) Tasks: The name of a task is stored in the „Activity" column and should start with a capital letter. To skip a phase or go back to a previous phase it is possible to use the „goto" statement in the activity field.

4) Collapsed subprocesses: If a task is a collapsed subprocess, the „Subprocess" field should be filled with „yes" and the subprocess should be then modelled in a separate spreadsheet.

5) Parallel-, Exclusive- and Inclusive-gateways: Gateways are represented by alternative branches in the sequence (phase) flow. A phase preceded by a logical gateway is named as follows: NxM, where N is the (natural) number of the phase in the main process branch, x is a letter (a-z) corresponding to the alternative branch and M is the number of the phase in the selected branch. If the branch contains only one task M can be omitted. We assume that the branching is always ended by the same type of gateway that started it. It may be perceived as a limitation, but in fact it prevents model unconsistency.

The representation of a simple AND-gateway was presented in the 6th row of Table II. In order to implement an XOR-gateway, the field „Condition" should be filled in with the appropriate condition and with the word „else" for the other task performed in the same phase, but in another branch. An example of a simple XOR-gateway was presented in Table I. In order to implement an OR-gateway the fields mentioned before should be filled with two separate conditions. An example of an OR-gateway within multiple gateway spreadsheet representation was presented in the last row of Table II.

6) Pool/lane: Only one pool is considered. Different swimlanes are represented by the field „Who" containing the appropriate department or name.

Our spreadsheet-based approach is dedicated to people who are unfamiliar with business process modelling and that is

Table II
TRANSFORMATION RULES FROM SPREADSHEET-BASED REPRESENTATION TO BPMN PROCESS MODEL STRUCTURE

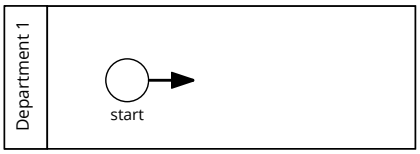
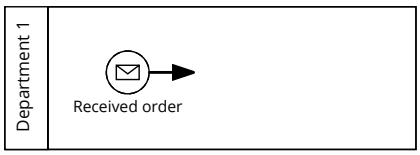
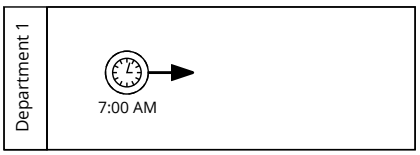
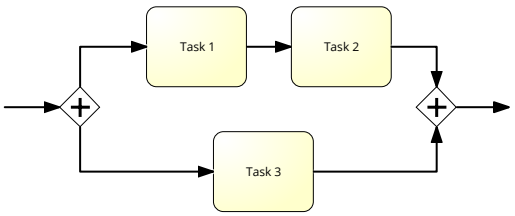
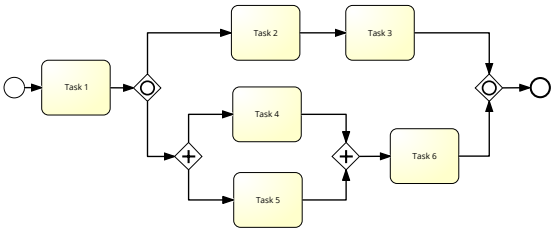
Spreadsheet representation	BPMN Element
<table><tr><td>Order</td><td>Activity</td><td>Condition</td><td>Who</td><td>(...)</td></tr><tr><td>0</td><td>start</td><td></td><td>Department 1</td><td></td></tr></table>	
<table><tr><td>Order</td><td>Activity</td><td>Condition</td><td>Who</td><td>(...)</td></tr><tr><td>0</td><td>message Receive Order</td><td></td><td>Department 1</td><td></td></tr></table>	
<table><tr><td>Order</td><td>Activity</td><td>Condition</td><td>Who</td><td>(...)</td></tr><tr><td>0</td><td>timer 7:00 AM</td><td></td><td>Department 1</td><td></td></tr></table>	
<table><tr><td>Order</td><td>Activity</td><td>Condition</td><td>(...)</td><td>Terminated</td></tr><tr><td>99</td><td>Request completed</td><td></td><td></td><td>yes</td></tr></table>	
<table><tr><td>Order</td><td>Activity</td><td>Condition</td><td>(...)</td><td>Terminated</td></tr><tr><td>99</td><td>message Send Report</td><td></td><td></td><td>yes</td></tr></table>	
<table><tr><td>Order</td><td>Activity</td><td>Condition</td><td>(...)</td></tr><tr><td>2a1</td><td>Task 1</td><td></td><td></td></tr><tr><td>2a2</td><td>Task 2</td><td></td><td></td></tr><tr><td>2b</td><td>Task 3</td><td></td><td></td></tr></table>	
<table><tr><td>Order</td><td>Activity</td><td>Condition</td><td>(...)</td></tr><tr><td>0</td><td></td><td></td><td></td></tr><tr><td>1</td><td>Task 1</td><td></td><td></td></tr><tr><td>2a1</td><td>Task 2</td><td>Condition 1</td><td></td></tr><tr><td>2a2</td><td>Task 3</td><td></td><td></td></tr><tr><td>2b1a</td><td>Task 4</td><td>Condition 2</td><td></td></tr><tr><td>2b1b</td><td>Task 5</td><td></td><td></td></tr><tr><td>2b2</td><td>Task 6</td><td></td><td></td></tr><tr><td>3</td><td>(end)</td><td></td><td></td></tr></table>	

Table III
COMPARISON TO THE EXISTING APPROACHES

Element type	Simple approach	Branching approach	More Properties approach	Our approach
Task	●	●	●	●
Events	❍	●	●	●
Collapsed Subprocess	❍	●	●	●
AND, OR and XOR Gateways	❍	●	●	●
Pool, Lane	❍	◑	◑	◑
Data Object	❍	◑	●	❍
Sequence Flow	◑	●	●	●
Message Flow	❍	❍	❍	❍

why it is important to transform the created spreadsheet into a BPMN diagram in a simple way. Such a diagram is a correct BPMN model, however, it can still contain some behaviour anomalies if the spreadsheet contains some loops [15].

The proposed set of transformation rules can use a CSV file that can be created from any spreadsheet software. The table containing information about the business process is represented as an array of structures, where each row is a unique part of structure with some element fields in the row.

Table III presents the overall comparison of our approach to the existing approaches in terms of supported elements by these solutions. One can noticed that our approach supports fewer elements than such complex approaches as Branching or More Properties approaches. However, the other approaches support the elements in a straightforward way, requiring the deeper knowledge of business process elements, even in the case of such simple structures like gateways.

Moreover, there are several points, in which our approach has the advantage over the existing approaches:

- We do not use explicit notion of XOR/OR/AND gateways, thus the user does not have to think about the kind of flow branching. Instead, in our model it is modeled in the implicit way.
- For describing condition, we use the ”condition” column, what is more clear than using the ”description” field.
- We do not use the notion of ”Successor” as it does not always show clearly the flow in the spreadsheet representation. Thus, in our opinion, a well-known jump statement ”goto” (one-way control flow transfer), existing in many computer programming languages, in this particular usage performs better, as it is clearly visible in the Activity field.

IV. CONCLUSIONS

In the paper, we present a new method of generating a BPMN process model based on its spreadsheet-based representation. In contrast to the existing approaches, our method does not require explicit specification of gateways in the spreadsheet, but it takes advantage of the spreadsheet with numbered rows in the form of a nested list. Thanks to this, our method does not require the knowledge of the BPMN notation or any business process notation.

As future works, we plan to extend the method in order to support multiple pools, message flows, as well as more type of events, including boundary events.

REFERENCES

[1] F. Friedrich, J. Mendling, and F. Puhlmann, “Process model generation from natural language text,” in *Advanced Information Systems Engineering*, ser. Lecture Notes in Computer Science, H. Mouratidis and C. Rolland, Eds. Springer Berlin Heidelberg, 2011, vol. 6741, pp. 482–496.

[2] F. Levy and A. Nazarenko, “Formalization of natural language regulations through sbvr structured english,” in *Theory, Practice, and Applications of Rules on the Web*, ser. Lecture Notes in Computer Science, L. Morgenstern, P. Stefaneas, F. Levy, A. Wyner, and A. Paschke, Eds. Springer Berlin Heidelberg, 2013, vol. 8035, pp. 19–33. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-39617-5_5

[3] A. Raj, T. V. Prabhakar, and S. Hendryx, “Transformation of SBVR Business Design to UML Models,” in *Proceedings of the 1st India Software Engineering Conference*, ser. ISEC ’08. New York, NY, USA: ACM, 2008, pp. 29–38.

[4] O. C. Tantan and J. Akoka, “Automated transformation of Business Rules into Business Processes,” in *Proceedings of the Twenty-Sixth International Conference on Software Engineering and Knowledge Engineering*, 2014, pp. 684–687.

[5] K. Kluza and K. Honkisz, “From SBVR to BPMN and DMN models. proposal of translation from rules to process and decision models,” in *Artificial Intelligence and Soft Computing*, ser. Lecture Notes in Computer Science, L. Rutkowski, M. Korytkowski, R. Scherer, R. Tadeusiewicz, L. A. Zadeh, and J. M. Zurada, Eds. Springer International Publishing, 2016, vol. 9693.

[6] A. Raj, A. Agrawal, and T. V. Prabhakar, “Transformation of Business Processes into UML Models: An SBVR Approach,” *International Journal of Scientific and Engineering Research*, July 2013.

[7] B. Steen, L. Pires, and M.-E. Iacob, “Automatic Generation of Optimal Business Processes from Business Rules,” in *Enterprise Distributed Object Computing Conference Workshops (EDOCW), 2010 14th IEEE International*, Oct 2010, pp. 117–126.

[8] L. Henrik, J. Mendling, and A. Polyvyanyy, “Supporting process model validation through natural language generation,” *IEEE Transactions on Software Engineering*, vol. 40, no. 8, pp. 818–840, 2014.

[9] J. Saldivar, C. Vairetti, C. Rodríguez, D. Florian, F. Casati, and R. Alarcón, “Analysis and improvement of business process models using spreadsheets,” *Information Systems*, vol. 57, pp. 1–19, 2016.

[10] J. R. Nawrocki, T. Nedza, M. Ochodek, and L. Olek, “Describing business processes with use cases,” in *BIS*, 2006, pp. 13–27.

[11] D. Lubke, K. Schneider, and M. Weidlich, “Visualizing use case sets as bpmn processes,” in *Requirements Engineering Visualization, 2008. REV ’08.*, 2008, pp. 21–25.

[12] A. A. Kalenkova, M. de Leoni, and W. M. van der Aalst, “Discovering, analyzing and enhancing bpmn models using prom?” in *Business Process Management-12th International Conference, BPM*, 2014, pp. 7–11.

[13] S. Krumnow and G. Decker, *Business Process Modeling Notation: Second International Workshop, BPMN 2010, Potsdam, Germany, October 13-14, 2010. Proceedings*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, ch. A Concept for Spreadsheet-Based Process Modeling, pp. 63–77.

[14] S. Krumnow, “Spreadsheet-based process modeling,” *Business Processes in the Real World*, pp. 55–70, 2010.

[15] A. Mroczek and A. Ligęza, “A note on bpmn analysis. towards a taxonomy of selected potential anomalies,” in *Computer Science and Information Systems (FedCSIS), 2014 Federated Conference on*. IEEE, 2014, pp. 1097–1102.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1359–1364
DOI: 10.15439/2016F324
ACSIS, Vol. 8. ISSN 2300-5963

Towards Rule-based Pattern Perspective for BPMN 2.0 Business Process Models

Krzysztof Kluza, Grzegorz J. Nalepa
AGH University of Science and Technology
al. A. Mickiewicza 30, 30-059 Krakow, Poland
E-mail: {kluza,gjn}@agh.edu.pl

***Abstract*—Rules and processes constitute powerful representation forms. Although the same notions can be expressed in both of these representations, there is a significant difference in abstraction levels between processes and rules. In practice, rules are mostly used for the specification of rule task logic in processes. In this paper, we present various options where and how rules can be perceived in business processes. We introduce rule-based pattern perspective for process models, focusing on BPMN 2.0 Business Process Models.**

***Index Terms*—BPMN, Business Processes, Business Process Hierarchization, Business Process Configuration**

I. Introduction

BUSINESS Process (BP) [1] models constitute visual representations of processes of an organization. BPMN is a dominant visual modeling language used for describing processes. A BPMN model should be easy to understand for non-business users. However, sometimes such a model becomes illegible or unclear due to its complexity. One of the ways of reducing complexity is to introduce rules in order to specify low level business logic of the process. Such Business Rules (BR) [2] can describe business policies, goals, strategies or guidelines, in a form of declarative statements, constraints or predicated actions. Another issue of improving comprehensibility of the model is to use the notation in a proper way.

As BPMN specifies only a notation, thus there can be several ways of using it. There are style directions how to model BPs [3], or guidelines for analysts based on BPs understandability (e.g. [4]). However, a proper business process modeling is still a challenging task, especially for inexperienced users.

Design patterns in software engineering propose reusable solutions independent from the implementation technology. Similarly, workflow patterns address business requirements independently from specific workflow languages, and describe the problem, conditions that should hold for the pattern in order to be aplicable, examples of business situation, as well as its realization or implementation [5]. As the existing workflow patterns perspectives does not take into account rules, we propose rule-based pattern perspective for process models.

The paper is organized as follows. In Section II we present the motivation for our research. Section III provides a short overview of related approaches. In Section IV we present our proposal of Rule-based Pattern Perspective for BPMN Process Models. Section V summarizes the paper.

II. Motivation

Although it is possible to express the same notions (concepts, objects, system behaviour description etc.) using processes as well as rules. These two approaches are not suitable for this purpose, as they use different languages, depict different aspects of a system, and thus there is a significant difference in abstraction levels between processes and rules. However, this is neither standardized nor obvious which method is better for expressing a particular semantics.

During modeling a system using processes and rules, several questions can arise, from the simple ones like which representation is better for a specific purpose, through the integration of proceses and rules, to detail aspects of representation pragmatics. Thus, in order to organize the knowledge about process models with rules, rule-based pattern perspective for processes is introduced. Based on workflow patterns, we selected several attributes which are applied to rule-based patterns in process models. Presenting the patterns, we focused on their BPMN representation. Such patterns can help in designing models with rules and increase the comprehension level of the model.

Similarly to other languages, in the BPMN modeling language, three aspects can be distinguished: syntax, semantics and pragmatics. In the case of syntax, the current BPMN 2.0 specification [6] describes it in detail, and our research is consistent with the specification describing the syntax of rule related elements (event, tasks). BPMN semantics have also been considered in several papers [7], however, not with an emphasis on rules. As in practice, the pragmatics of language use is essential, a short overview of the BPMN pragmatics is presented in the following section.

III. Related works

In the case of the BPMN pragmatics the following particular issues can be considered: modeling techniques and styles, workflow patterns, as well as domain patterns. These issues are elaborated in the next subsections.

A. Workflow Patterns and Domain Process Patterns

As design patterns provide reusable solutions [9] independent from the implementation technology, workflow patterns address business requirements independently from specific workflow languages. Such patterns describe conditions that should hold for them in order to be applicable, examples of business situation, description of the problem, as well as realization or implementation in current languages [5].

978-83-60810-90-3/$25.00

Table I
FRAGMENT OF IMPLICIT TERMINATION PATTERN SPECIFICATION[1]

Implicit Termination Pattern	
Description:	*A given process (or sub-process) instance should terminate when there are no remaining work items that are able to be done either now or at any time in the future and the process instance is not in deadlock. There is an objective means of determining that the process instance has successully completed.*
Motivation:	*The rationale for this pattern is that it represents the most realistic approach to determining when a process instance can be designated as complete. This is when there is no remaining work to be completed as part of it and it is not possible that work items will arise at some future time.*
Solutions:	*For simple process models, it may be possible to indirectly achieve the same effect by replacing all of the end nodes for a process with links to an OR-join which then links to a single final node. However, it is less clear for more complex process models involving multiple instance tasks whether they are always able to be converted to a model with a single terminating node. Potential solutions to this are discussed at length by Kiepuszewski et al. [8].*

Originally, van der Aalst et al. identified a set of 26 patterns that describe the control-flow perspective of business processes [5], [10]. Table I presents an example of a typical structural control-flow pattern for implicit termination of a process[1]. In the case of the BPMN notation, this pattern is directly supported from the very beginning (version BPMN 1.0) by ending every thread of a process with an end event. When the last token in the process generated by the start event is consumed, the process instance terminates.

Over the years, the existing workflow patterns have been evaluated [11], [12], [13], revised [14], [15], and extended to cover new perspectives like the data and resource perspectives [16], [17], or time perspective [18].

Until now, much more workflow patterns have been identified. The Workflow Patterns Initiative[2] distinguishes:

- 42 control-flow patterns,
- 43 workflow resource patterns,
- 40 workflow data patterns,
- 12 abstract and 8 concrete syntax modification patterns,
- over 100 exceptions patterns of various exception types.

Such patterns serves not only as good practices but also for a pattern-based evaluation of tools or standards [19], [20], [21].

What is more, one can also distinguish patterns concerning the process of modeling itself [22], [23], like fixation patterns during process model creation [24], or abstract and concrete syntax modifications patterns [25], [26].

Another kind of patterns – Domain Process Patterns (DPP) [27] – represent functions of process model fragments that are applicable to some modeling domain. They were introduced as a result of the investigation of business processes from the order management and the manufacturing production domains. Such patterns describes some domain related business operations representing a small fragment of the process. Although it is possible to consider involving rules in DPP, e.g. in *Inventory Pattern* one can consider rules involved in inventory management, it is not a purpose of DPP.

[1]An example can be accessed at http://www.workflowpatterns.com/patterns/control/structural/wcp11.php.

[2]See: http://www.workflowpatterns.com.

Neither workflow patterns nor domain process patterns do not consider the integration of processes with rules.

B. Business Rule Patterns

A business rule taxonomy that serves as a rich source of business rules can be found in [28], [29]. The authors specified more than 60 business rule structures that for a specific process instance: restrict the number of allowed instances of a specific process elements, restrict the coexistence of process elements of different types, specify the influence of specific data elements on the occurrence of process elements, a time restriction on process elements, or a property for a process element at a predefined process state. These rules were used for a comprehensive rule-based compliance checking approach. However, they were not analyzed from the business process representation perspective.

In the following section, we present a short overview of 10 selected rule-based patterns, analyze the rules from the business process representation perspective, and show how they can be observed in the BPMN process models.

IV. RULE-BASED PATTERNS IN BPMN PROCESS MODELS

In order to specify the rule patterns in BPMN process models, we use the following attributes:

1) **Pattern Name**: Descriptive name of the pattern.
2) **Description**: Description of the pattern.
3) **Motivation**: Description of the pattern purpose or a problem that is addressed by the pattern.
4) **BPMN Elements**: The list of the BPMN elements used in the pattern.
5) **Process Place**: Where in the process model the pattern can be applied: Start – at the beginning of the process, Intermediate – during the course of the process, or End – at the end of the process.
6) **An example**: An illustrative example presenting the pattern.

In the following subsections we present 10 selected rule patterns that can be observed in BPMN process models: Conditional Flow, Conditional Triger, Conditional Task/Subprocess Interruption, Conditional Process Interruption with Initiation, Rule-based Task, Simple Conditional Choice, Rule-based

Choice, Deferred Conditional Choice, Conditional Task Multiplicity and Task Performer Assignment.

A. Conditional Flow

Description: Conditional Flow provides the ability to control the flow of a token based on the evaluation of the *condition* expression in the process instance.
Motivation: Controlling the flow using conditional expressions serves as an additional building block for process model that allows for detailed controling of the flow of token through the branch of sequence flow. This pattern provides a condition for a sequence flow and is a variant of Data-based routing pattern[3] in the Data-based perspective.
BPMN Elements: Conditional Sequence Flow
Process Place: Intermediate
An example: Conditional flow with the condition: *An amount has to be higher than 100* is presented in Fig. 1.

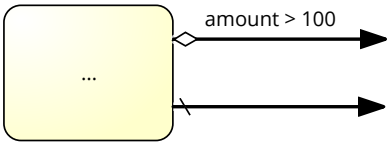


Figure 1. Conditional flow example

B. Conditional Triger

Description: Conditional Triger provides the ability to triger the flow of a token based on the evaluation of the *condition* expression in the process instance.
Motivation: This pattern provides a means of triggering the initiation or resumption of the token flow when a *condition* in the process instance is satisfied. The pattern provides a condition for an event and is a variant of Data-Based Task Trigger pattern[4] in the Data-based perspective.
BPMN Elements: Conditional Event
Process Place: Start / Intermediate
An example: Conditional Triger which trigers when *a customer credit rating will be higher than 4* is presented in Fig. 2.

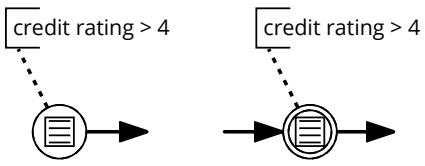


Figure 2. Conditional triger examples

[3]See: http://www.workflowpatterns.com/patterns/data/routing/wdp40.php
[4]See: http://www.workflowpatterns.com/patterns/data/routing/wdp39.php

C. Conditional Task/Subprocess Interruption

Description: When a specific condition is satisfied, a task/subprocess is interrupted and its further execution is abandoned.
Motivation: Conditional Task Interruption pattern provides the ability to abandon the execution of a task/subprocess based on fulfilling a *condition* of the Conditional Triger. The pattern is related to Cancel Task pattern[5] in the Control Flow perspective.
BPMN Elements: Interruptive Boundary Conditional Event attached to a Task or a Subprocess
Process Place: Intermediate
An example: A task that will be interrupted when *Credit rating is below minimum* is presented in Fig. 3.

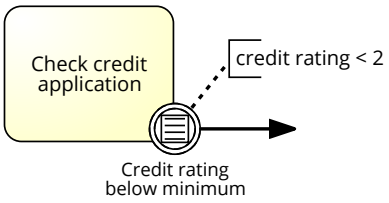


Figure 3. Conditional task interruption example

D. Conditional Process Interruption with Initiation

Description: When a specific condition is satisfied, the current process is interrupted and its further execution is abandoned, and a fragment of the process is started from the Conditional Triger. A started fragment is not part of the regular control flow.
Motivation: Conditional Process Interruption pattern provides the ability to abandon the execution of a process based on fulfilling a *condition* of the Conditional Triger and initiates a new subprocess not instantiated by normal control flow. The pattern is related to Cancel Case[6] and Cancel Region[7] patterns in the Control Flow perspective.
BPMN Elements: Event Subprocess with Conditional (Interruptive) Start Event
Process Place: Intermediate
An example: A Process will be interrupted when *a customer credit rating is invalid*. Then a procedure of handling the invalid credit rating will be initiated (see: Fig. 4).

E. Rule-based Task

Description: Rule-based Task allows for specification of the task logic using rules and delegating work to a Business Rules Engine in order to receive calculated or inferred data.

[5]See: http://www.workflowpatterns.com/patterns/control/cancellation/wcp19.php
[6]See: http://www.workflowpatterns.com/patterns/control/cancellation/wcp20.php.
[7]See: http://www.workflowpatterns.com/patterns/control/cancellation/wcp25.php.

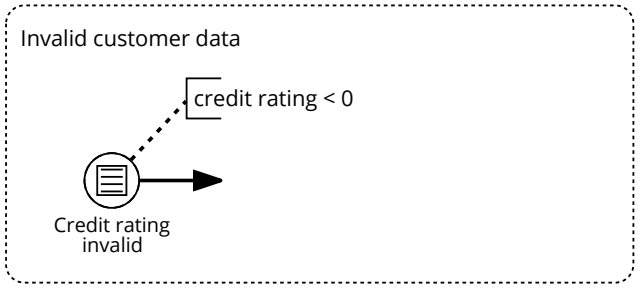


Figure 4. Conditional Process Interruption with Initiation example

Motivation: This pattern provides a means of using Business Rules Engine for execution of rules related to the task.
BPMN Elements: Business Rule Task, Business Rule Task (Call Activity)
Process Place: Start, Intermediate
An example: An example of Rule-based Task which determines a credit card type according to the defined rules is presented in Fig. 5.

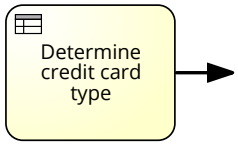


Figure 5. Rule-based Task example

F. Simple Conditional (Exclusive/Multi-Choice/Complex) Choice

Description: Simple Conditional Choice diverges a branch into two or more branches, and according to specified type of choice (Exclusive/Multi-Choice/Complex) the token from the incoming branch is passed to one or more outgoing branches based on the evaluation of the $condition$ expressions of the branches in the process instance with support of a mechanism that can limit the number of the outgoing branches.
Motivation: This pattern provides a means of detailed controling of the flow of token through the branches of sequence flow. It is a variant of Exclusive Choice[8] and Multi-Choice[9] patterns in the Control Flow perspective.
BPMN Elements: Exclusive Gateway, Inclusive Gateway, Complex Gateway
Process Place: Intermediate
An example: In Fig. 6 an example of Conditional Exclusive Choice is presented. In this example the choice is made according to the account limit (*either higher or equal 1000, or lower than 1000*).

[8]See: http://www.workflowpatterns.com/patterns/control/basic/wcp4.php.
[9]See: http://www.workflowpatterns.com/patterns/control/advanced_branching/wcp6.php.

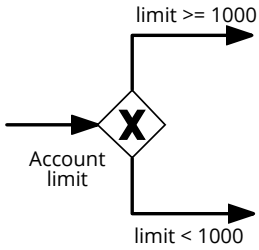


Figure 6. Simple Conditional Exclusive Choice example

G. Rule-based Choice

Description: This pattern is essentially an extension (combination) of the Rule-based Task and the Simple Conditional Choice pattern described above. It provides the Simple Conditional Choice behaviour, but ensures that a decision is made based on the output of the Rule-based Task (the result of inference).
Motivation: This pattern provides a means of detailed controling of the flow of token through the branches of sequence flow based on the on the output from the Business Rules Engine.
BPMN Elements: Gateway preceded by a Business Rule task
Process Place: Start, Intermediate
An example: In the example in Fig. 7, the choice is made according to the verification result (positive/negative) which is a value obtained from customer verification. The customer verification is performed automatically by a Business Rule Engine using the predefined rules.

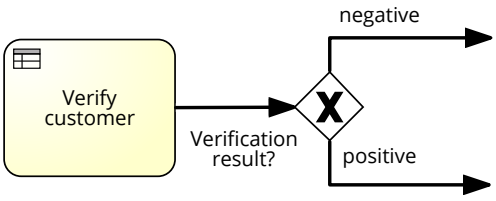


Figure 7. Rule-based Choice example

H. Deferred Conditional Choice

Description: Deferred Conditional Choice determines a point in a process where one or more branches are chosen according to the Conditional Trigers (see Sect. IV-B). The difference from the Simple Conditional Choice pattern is that in this pattern the decision is deferred and depends on the Conditional Trigers.
Motivation: This pattern provides a means of defering the moment of choice in a process to the last possible time and is based on the conditional trigers. The pattern is a variant of the Deferred Choice pattern[10] in the Control

[10]See: http://www.workflowpatterns.com/patterns/control/state/wcp16.php.

Flow perspective.
BPMN Elements: Event-based Gateway followed by Conditional Events.
Process Place: Start, Intermediate
An example: The choice in Fig. 8 is deferred to a point in which either a customer credit rating will be below minimum or a customer has more than 2 unpaid loan installments.

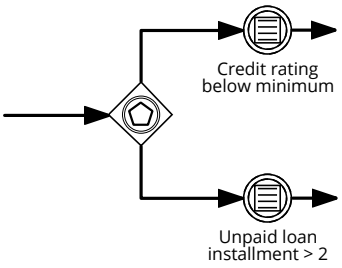


Figure 8. Deferred Conditional Choice example

I. Conditional Task Multiplicity

Description: Within a given process instance, multiple instances of a task can be created according to a condition specified in the rule.
Motivation: This pattern provides a means of rule-based specification of the task multiplicity. The pattern is related to the Multiple Instance Patterns[11] in the Control Flow perspective.
BPMN Elements: Multiinstance task with the specified attributes
Process Place: Start, Intermediate, End
An example: *Credit card application has to be approved by 2 bank employees* (see Fig. 9).

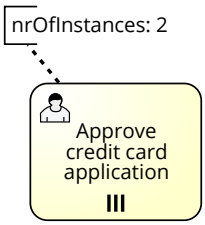


Figure 9. Conditional Task Multiplicity example

J. Task Performer Assignment

Description: This pattern specifies who performs which tasks in a process.
Motivation: This pattern provides a means of a kind of deontic rule that specify the performer role assigned to particular tasks. The pattern is related to the Role-Based Distribution pattern[12] in the Resource perspective.
BPMN Elements: Lanes
Process Place: Start/Intermediate/End
An example: See Fig. 10: *An approval task has to be performed by a supervisor.*

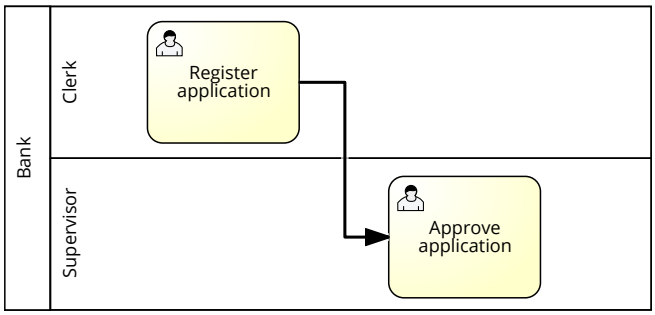


Figure 10. Task Performer Assignment example

V. Concluding Remarks

In the paper, various types of workflow patterns and domain process patterns were presented. However, these patterns does not take into account rule-based aspects. Thus, we considered various options where and how rules can be perceived in business processes.

The original contribution of this paper is presentation of rule pattern perspective for process models. We described 10 selected rule-based patterns and showed the examples of such patterns in BPMN 2.0 Business Process Models. We focused on the BPMN 2.0 notation, as it is standardized process representation use both by researchers and enterprises.

As future work, we plan to supplement the patterns with other decision aspects in BP models and extend them by investigating other process languages as well as extensions of the BPMN notation [30]. Furthermore, we plan to evaluate these patterns empirically using real-life process models as well as to check the possible anomalies that these patterns can induce [31].

We also want to include design issues using the proposed patterns [32]. as well integrate the pattern library with a process editor [33]. especially for model recommendations based on patterns during the design [34].

Acknowledgment

The paper is supported by the AGH UST Grant.

References

[1] M. Dumas, M. La Rosa, J. Mendling, and H. A. Reijers, *Fundamentals of Business Process Management*. Springer Berlin Heidelberg, 2013.
[2] D. Hay, A. Kolber, and K. A. Healy, "Defining Business Rules – what they really are. final report." Business Rules Group, Tech. Rep., July 2000.
[3] B. Silver, *BPMN Method and Style*. Cody-Cassidy Press, 2009.

[11] See: http://www.workflowpatterns.com/patterns/control/.

[12] See: http://www.workflowpatterns.com/patterns/resource/creation/wrp2.php.

[4] J. Mendling, H. A. Reijers, and W. M. P. van der Aalst, "Seven process modeling guidelines (7pmg)," *Information & Software Technology*, vol. 52, no. 2, pp. 127–136, Feb 2010.

[5] W. Aalst, A. Barros, A. Hofstede, and B. Kiepuszewski, "Advanced workflow patterns," in *Cooperative Information Systems*, ser. Lecture Notes in Computer Science, P. Scheuermann and O. Etzion, Eds. Springer Berlin Heidelberg, 2000, vol. 1901, pp. 18–29.

[6] OMG, "Business Process Model and Notation (BPMN): Version 2.0 specification," Object Management Group, Tech. Rep. formal/2011-01-03, January 2011.

[7] R. M. Dijkman and P. V. Gorp, "Bpmn 2.0 execution semantics formalized as graph rewrite rules," in *Proceedings from the Business Process Modeling Notation – Second International Workshop, BPMN 2010, Potsdam, Germany, October 13-14, 2010*, ser. Lecture Notes in Business Information Processing, J. Mendling, M. Weidlich, and M. Weske, Eds., vol. 67. Springer, 2011, pp. 16–30.

[8] B. Kiepuszewski, A. ter Hofstede, and W. van der Aalst, "Fundamentals of control flow in workflows," *Acta Informatica*, vol. 39, no. 3, pp. 143–209, 2003.

[9] A. Nolte, E. Bernhard, J. Recker, F. Pittke, and J. Mendling, "Repeated use of process models: The impact of artifact, technological and individual factors," *Decision Support Systems*, 2016.

[10] W. M. P. van der Aalst and A. H. M. ter Hofstede, "Workflow patterns: On the expressive power of (petri-net-based) workflow languages," in *Proceedings of the Fourth International Workshop on Practical Use of Coloured Petri Nets and the CPN Tools, Aarhus, Denmark, August 28-30, 2002*, K. Jensen, Ed., University of Aarhus. DAIMI PB-560, Aug 2002, pp. 1–20.

[11] W. van der Aalst, A. ter Hofstede, B. Kiepuszewski, and A. Barros, "Workflow patterns," *Distributed and Parallel Databases*, vol. 14, no. 1, pp. 5–51, 2003.

[12] W. Aalst and A. ter Hofstede, "Workflow patterns put into context," *Software & Systems Modeling*, vol. 11, no. 3, pp. 319–323, 2012.

[13] S. A. White, "Process modeling notations and workflow patterns," in *Workflow Handbook 2004*, L. Fischer, Ed. Future Strategies Inc., 2004, pp. 265–294.

[14] N. Russell, A. ter Hofstede, W. van der Aalst, and N. Mulyar, "Workflow control-flow patterns: a revised view," BPM Center Report, Tech. Rep. BPM-06-22, 2006, bPMcenter.org.

[15] M. Zapletal, W. M. van der Aalst, N. Russell, P. Liegl, and H. Werthner, "An analysis of windows workflow's control-flow expressiveness," in *Seventh IEEE European Conference on Web Services, 2009. ECOWS'09*. IEEE, 2009, pp. 200–209.

[16] N. Russell, W. M. Aalst, A. H. Hofstede, and D. Edmond, "Workflow resource patterns: Identification, representation and tool support," in *Advanced Information Systems Engineering*, ser. Lecture Notes in Computer Science, O. Pastor and J. Falcao e Cunha, Eds. Springer Berlin Heidelberg, 2005, vol. 3520, pp. 216–232.

[17] N. Russell, A. H. Hofstede, D. Edmond, and W. M. der Aalst, "Workflow data patterns: Identification, representation and tool support," in *Conceptual Modeling – ER 2005*, ser. Lecture Notes in Computer Science, L. Delcambre, C. Kop, H. Mayr, J. Mylopoulos, and O. Pastor, Eds. Springer Berlin Heidelberg, 2005, vol. 3716, pp. 353–368.

[18] A. Lanz, B. Weber, and M. Reichert, "Time patterns for process-aware information systems," *Requirements Engineering*, vol. 19, no. 2, pp. 113–141, 2014.

[19] M. Skouradaki, V. Ferme, C. Pautasso, F. Leymann, and A. van Hoorn, "Micro-benchmarking bpmn 2.0 workflow management systems with workflow patterns," in *International Conference on Advanced Information Systems Engineering*. Springer, 2016, pp. 67–82.

[20] K. Kaiser and M. Marcos, "Leveraging workflow control patterns in the domain of clinical practice guidelines," *BMC medical informatics and decision making*, vol. 16, no. 1, p. 1, 2016.

[21] A. Delgado, D. Calegari, P. Milanese, R. Falcon, and E. García, "A systematic approach for evaluating bpm systems: case studies on open source and proprietary tools," in *IFIP International Conference on Open Source Systems*. Springer, 2015, pp. 81–90.

[22] J. Claes, I. Vanderfeesten, F. Gailly, P. Grefen, and G. Poels, "The structured process modeling theory (spmt) a cognitive view on why and how modelers benefit from structuring the process of process modeling," *Information Systems Frontiers*, vol. 17, no. 6, pp. 1401–1425, 2015.

[23] J. Claes, I. Vanderfeesten, J. Pinggera, H. A. Reijers, B. Weber, and G. Poels, "A visual analysis of the process of process modeling," *Information Systems and e-Business Management*, vol. 13, no. 1, pp. 147–190, 2015.

[24] B. Weber, J. Pinggera, M. Neurauter, S. Zugal, M. Martini, M. Furtner, P. Sachse, and D. Schnitzer, "Fixation patterns during process model creation: Initial steps toward neuro-adaptive process modeling environments," in *2016 49th Hawaii International Conference on System Sciences (HICSS)*, Jan 2016, pp. 600–609.

[25] M. La Rosa, P. Wohed, J. Mendling, A. ter Hofstede, H. Reijers, and W. M. P. Van der Aalst, "Managing process model complexity via abstract syntax modifications," *Industrial Informatics, IEEE Transactions on*, vol. 7, no. 4, pp. 614–629, 2011.

[26] M. L. Rosa, A. H. M. ter Hofstede, P. Wohed, H. A. Reijers, J. Mendling, and W. M. P. van der Aalst, "Managing process model complexity via concrete syntax modifications," *IEEE Transactions on Industrial Informatics*, vol. 7, no. 2, pp. 255–265, 2011.

[27] A. Koschmider and H. A. Reijers, "Improving the process of process modelling by the use of domain process patterns," *Enterprise Information Systems*, 2013.

[28] F. Caron, J. Vanthienen, and B. Baesens, "Comprehensive rule-based compliance checking and risk management with process mining," *Decision Support Systems*, vol. 54, no. 3, pp. 1357–1369, 2013.

[29] F. Caron, J. Vanthienen, and B. Baesens, "Business rule patterns and their application to process analytics," in *17th IEEE International Enterprise Distributed Object Computing Conference Workshops (EDOCW 2013)*, Sept 2013, pp. 13–20.

[30] A. Ligęza and T. Potempa, "Artificial intelligence for knowledge management with bpmn and rules," in *IFIP International Workshop on Artificial Intelligence for Knowledge Management*. Springer, 2012, pp. 19–37.

[31] A. Mroczek and A. Ligęza, "A note on bpmn analysis. towards a taxonomy of selected potential anomalies," in *Computer Science and Information Systems (FedCSIS), 2014 Federated Conference on*. IEEE, 2014, pp. 1097–1102.

[32] G. J. Nalepa and A. Ligęza, *Software engineering: evolution and emerging technologies*, ser. Frontiers in Artificial Intelligence and Applications. Amsterdam: IOS Press, 2005, vol. 130, ch. Conceptual modelling and automated implementation of rule-based systems, pp. 330–340.

[33] K. Kluza, K. Kaczor, and G. J. Nalepa, "Enriching business processes with rules using the Oryx BPMN editor," in *Artificial Intelligence and Soft Computing: 11th International Conference, ICAISC 2012: Zakopane, Poland, April 29–May 3, 2012*, ser. Lecture Notes in Artificial Intelligence, L. Rutkowski and [et al.], Eds., vol. 7268. Springer, 2012, pp. 573–581. [Online]. Available: http://www.springerlink.com/content/u654r0m56882np77/

[34] S. Bobek, G. J. Nalepa, and O. Grodzki, "Integration of activity modeller with bayesian network based recommender for business processes," *Knowledge Engineering and Software Engineering (KESE10)*, p. 42, 2014.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1365–1368
DOI: 10.15439/2016F234
ACSIS, Vol. 8. ISSN 2300-5963

Concept of the urban knowledge. A case of Poland

Katarzyna Marciniak
Wroclaw University of Economics
ul. Komandorska 118/120, 53-345
Wroclaw, Poland
Email:
katarzyna.marciniak@ue.wroc.pl

***Abstract*— This document describes the concept of urban knowledge. Main aim of this paper is to define term, describe assumptions and possibilities of use. The paper consists of introduction, five chapters and the summary. In introduction author presents background of conducted research. First chapter is the explanation of the urban knowledge phenomenon. Second presents further characteristic with brief classification. Later, author presents how urban knowledge can be managed. In the last chapter, author is trying to point possible methods of urban knowledge acquisition based on available ICT solution. In the summary author will describe chances of application for presented model.**

I. Introduction

As the priority issues, which are defined by the National Spatial Development Concept 2030[18]., are considered: the low competitiveness of major urban centers and Polish regions against European, poor territorial cohesion of the country, the low level of infrastructure (especially transport, social, information, including information technology) areas urban and rural areas, the lack of a coherent system of environmental protection, insufficient resistance to the spatial structure of internal and external threats and spatial disorder.

According to the author's knowledge, a key element for the pursuit of development in line with the reasons stemming from the macro-cities is an urban knowledge. Taking into account the results of enterprises market functioning and the results that achieve through the use of knowledge management processes, the author believes that urban knowledge may be an aspect that will allow raising the competitiveness of cities, seeking to ensure the integration of infrastructure (mainly in the area of information technology, social and governance) and achieving spatial cohesion. Thus, the author believes that the improvement of management processes in the cities, on the basis of knowledge can contribute to the real fulfillment of assumptions NSDC 2030 in Poland.

This work was not supported by any organization

II. Urban knowledge definition

To be able to explain the urban knowledge concept, it is necessary to localize it roots in the concept of organizational knowledge.

Knowledge is not only one of the most important assets of an organization, but also the basis, the starting point used to define a strategy, especially for the implementation of management information systems. Knowledge, for the purposes of computing is based on data and information, where data is understood as a set of facts, measurements, statistics and information as the structuring of the data. In this sense, knowledge is a set of information that can be used in practice. However, the organization cannot exist without human capital. [9]

In the theory of knowledge management, there is the explicit and discreet knowledge. [6] Explicit knowledge (formal), as general, considers formal documents created by the organization. Discrete knowledge (informal) is defined as skills, qualifications and experience of human capital. This knowledge is not written anywhere, it is in minds of employees, what really constitutes a barrier to obtaining confidential knowledge. However, there are no barriers that cannot be overcome. [4]

In order to ensure the proper transformation of tacit knowledge into explicit one, as the first condition, it has to be pointed and realized by the organizational culture in the enterprise. [2] Secondly human resources have to be ensured in adequate mechanisms of facilitation the knowledge sharing process in a fair and codified why. It is therefore necessary to create appropriate procedures and systems which should enable knowledge capturing, processing and sharing according to requirements. [7] The weak point of this attitude is still the will of the employee, if they wants to share knowledge or not.

What it is the urban knowledge? The spine of the concept is equal to definitions presented above but in authors opinion it is needed to extend the classification.

First, it is needed to distinguish tacit and explicit knowledge of the city, bearing in mind that the city consists of municipal decision-making units and its external and

internal environment. Therefore, the explicit knowledge of the city can be considered as all kinds of documents, law, statistics, forecasts, reports, documentation, information, demands, portals relevant to public services or stakeholders, available for the public by each of the entities functioning within the city. Whereas the tacit knowledge of the city considers the qualifications, experience, skills, knowledge of both: employees of the administrative and governing units and knowledge of stakeholders, as well as citizens.

To have a comprehensive description of the urban knowledge, presented basics should be extended by several features important from the point of view of tacit knowledge. Urban knowledge is created independently by entities operating both inside and outside the structure of the city. This kind of knowledge must be gathered and stored in a very specific way, which will provide its uniqueness for a long time.

Urban knowledge is created by the local authorities, politicians, social activists, educational institutions, healthcare entities, religious, environmental bodies, research and development, administrative units, business, science, citizens and other stakeholders. According to the model of intelligent city, author decided to classify urban knowledge on six key aspects described in table 1.

TABLE I.
TYPOLOGY OF URBAN KNOWLEDGE

Urban knowledge type	Components
Economical knowledge	Cooperation, Productivity, Entrepreneurship, Finance/budgeting , Public relations/marketing, Planning, Infrastructure
Social knowledge	Social difference and classes, Social problems, Social needs, Believes
Environmental knowledge	Natural resources, Access to resources, Eco services, entrepreneurships, Pollution, Environment protection, Policy of sustainable environment
Mobility knowledge	Services, condition and sources of public transport, External sources of public transport, Availability and stability of ICT infrastructure, Transport systems
Governance knowledge	Public and private sector, Citizens involvement, Suppliers and recipients in decision-making processes, Clarity of managerial processes, Rules of sustainable development, Perspective of cities development
Quality of life knowledge	The attractiveness of available objects and events in the city, region, Health condition, access to medical units, Safety of residents and businesses, Living conditions, Education level, Tourist attractiveness,Social cohesion

All the presented in the table 1 aspects of urban knowledge will be characterized briefly in the next chapter.

III. THE ESSENCE OF URBAN KNOWLEDGE

The essence of the urban knowledge is to provide the most comprehensive knowledge of its stakeholder, resources, processes, conditions or predispositions. Proving, that in the context of smart city, transformation of the city is geared strictly to the needs of citizens, can be found in just presented the characteristics of the knowledge city.

A. Economical knowledge

Category of economical knowledge of the city revolves around the existing cooperation between the city and specific subjects. It focuses on information about potential cooperation, development opportunities, possible investments. Collaborations, which in the past have produced adverse effects are included as well. An important point of economic knowledge is also a problem of productivity of the city. It is not just about the number and quality of public services provided, but also about the knowledge associated with the ability of the city to meet the needs of its residents. The productivity of the city is an indicator of the level of wages, benefits, profits, and taxes provided for urban residents. Knowledge of entrepreneurship in the city is the basis for defining, classifying the level and direction of development of the city.

B. Environmental knowledge

Urban knowledge also focuses on the area related to the natural environment. "Talking about the natural environment of the city is so reasonable that urban development and functioning of the technical infrastructure largely transformed original, natural conditions and makes the individual elements of the natural environment in the city acquire specific characteristics." In this regard, the managers are obliged to collect and update the knowledge on natural resources and access to these resources. Ecological services available in the city needed to be implemented.

C. Mobility knowledge

Mobility of the city is also important for defining the urban knowledge. It is a very difficult area of decision-making for the city. Very often the local authorities must make decisions based on conflict of information or conflict of society needs or both of them in the same time. In order to facilitate a process of planning and then implementing solutions in due to meet the needs of society, it is important to acquire, collect and use of information and knowledge related to public transport services, its technical condition and possible alternatives in the form of external transport services. As well, access and the stability of ICT infrastructure supporting the management of transport and traffic in cities became crucial. Gathering such knowledge, and its subsequent use is mainly aimed at meeting the needs of ensuring a well-functioning, safe and optimal public transport in the city.

D. Managerial knowledge

Knowledge management within the city has to be concerned with the transparency of management processes, functioning on the basis of rational, prospects for the development of the city and make rational decisions based on their knowledge and experience. The source of such knowledge are certainly laws, their distribution task for all public administrations and local self-government, participation of citizens in the process of creating a vision

and strategy for the city, the characteristics of private and public sector, suppliers and recipients in decision-making and knowledge about the condition of the city, held resources and needs to meet. Possession of such knowledge by policymakers in the city can provide an excellent basis for constructing a holistic view on some issues, including defining the correct relationships of cause and effect.

E. Social and quality of life knowledge

Each of the six selected aspects of the urban knowledge concerns directly or indirectly on social layer. However, in the authors opinion, two aspects – social and quality of life knowledge, speak directly about the needs of the residents of the city.

IV. URBAN KNOWLEDGE MANAGEMENT

Knowing that each city of democratic country is obligated to develop and realize their own strategies, especially based on knowledge, it is possible to claim, that urban knowledge management involves the effective use of urban knowledge and transforming it into the lasting value for city's stakeholders and managerial staff. [2], [14] Urban knowledge management is clearly defined and systematic management of dynamic knowledge for the city and its associated processes of creating, gathering, organizing, diffusion, use and exploitation of knowledge, carried out in pursuit of the objectives of the city. [16]

Urban knowledge management can be also treated as a specially designed system that helps cities to acquire, analyse the use (re-use) of urban knowledge in order to make faster, smarter and better decisions, so that they can achieve a competitive advantage in case of covering cities stakeholders needs. [10] Urban knowledge management covers management of information, knowledge and expertise available within the city, i.e. mobility, environment, social, managerial, economical by the creation, collection, storage, sharing and use, to ensure the cities future development based on well prepared decision plans. Urban knowledge management deliberates cities strategy, which selects, distils, stores, organizes, packs and provides information relevant to the cities stakeholders in a way that improves efficiency and competitiveness of the city in the end. [5], [3]

A properly selected integrated management information system now enables the efficient management of the entire city, carrying out city's management functions with regard to both its internal environment and external, and thus, proper management of urban knowledge. "The selection of specific solutions and technologies in the field of urban knowledge management in the city should be carried out depending on the specifics of the city, its profile, individual economic situation, its strategy and approach to knowledge management. Each city or its institution should be regarded as an organization with its characteristic organizational culture, creative employees, principles and standards prevailing within it and look at it through the prism of ongoing business processes. In addition, each city must be aware that the overall urban knowledge management system cannot be based only on properly chosen technology" [17].

Resource management of urban knowledge, with respect to the characteristics presented in the article, the requirements of the assumptions Knowledge Based Economy or observable development solutions, can be supported by specialized tools of ICT dedicated for the specified area. As the main tasks carried out, supported by solutions that support urban knowledge management, include[1], [14]:

- the acquisition of knowledge from a variety of, often heterogeneous resources,
- creating and coding different elements of urban knowledge in order to incorporate them into basic information system in the city,
- supporting the exchange of urban knowledge and providing comprehensive operation of teamwork.

These examples of tasks supported by dedicated tools for urban knowledge management include: distribution of documents and their monitoring, defining and monitoring over the course of project implementation, communication between stakeholders and cities management, workflow control and monitoring of teamwork. Knowledge in these tools is treated very flexibly - starting with the classification of unstructured documents and finishing on the formal knowledge bases [11], [12].

V. ACQUISITION OF URBAN KNOWLEDGE

Ensuring constant contact with the public by means of ICT solutions, e.g. knowledge exchange platform, can contribute to enhance the competitiveness of the city, leveling of social inequalities, mitigate conflicts, alleviate the problems of infrastructure, economic, or investment. In addition, it will help to ensure consistency of information in the city.

As the purpose of the existence and operation of ICT infrastructure in the city is the integration of key information generated by its users, which provide a complete list of requirements, guidelines for maintenance and improvements in many aspects, including those associated with improving the quality of life by better knowledge management in the city. Because of this, it is a smart city project, the concept of a better life for the citizens, there is nothing surprising in fact, that talks about the need for citizens:

- a comprehensive contact, including the implementation of intelligent technology and information resources public, allowing the systematization and increase citizens' access to information and knowledge of using ICT infrastructures,
- full integration, assuming that the basic infrastructure scheme operating system of the target group heterogeneous information generated by specific entities are together entirely integrated into a single portal for example, and the assistance available to the presentation layer, respectively specific audiences,

- incentives for innovation (called encouragement for innovation), covering all the action of the city authorities for enterprises and public institutions, in order to propagate their use of new technologies as a means of enabling equality technological society,
- collaboration (called collaborative operation), based on intelligent infrastructure, critical systems and cooperating users (including public bodies, local authorities, professionals, non-governmental bodies, excluded people, etc.), which helps to improve effectiveness in the developing the public services.

Presented solution enables the integration of the information needs of citizens and all the entities interested.

For platform users, platform becomes a compendium of knowledge about the urban knowledge (economy, people, environment, mobility, governance, living).

As we can see, ensuring citizens by the city's authorities can contribute to the objectives of the NSDC 2013, which points combating low competitiveness of major urban centers and Polish regions against European, poor territorial cohesion of the country, low level of infrastructure (especially transport, social, information, including information technology) urban and rural areas, the lack of a coherent system of environmental protection, insufficient resistance to the spatial structure of internal and external threats and spatial disorder.

VI. Conclusion

Proposed in the paper approach with use of ICT solutions to gather urban knowledge, is unfortunately so idyllic that it does not take into account the problems which for centuries is faced by all cities of the world, i.e. excluded, homeless, not involved in the social life of the city, the mismatch solutions to the needs of users , lack of willingness to change thinking among municipal authorities, resistance among employees of local government, political, economic phenomena robbery, the gray zone, the fear of citizens, and cyber security.

Of course, each of these problems are trying to be solved by finding and implementing various solutions, but it not always provide the desired effect. On the other hand, not handling problems in case of improving quality of lives most of citizens by authorities would be a crime, which is governed in Poland by the Criminal Code (intentional act to the detriment of the customer, citizen, entity). As we can see, the decisions of investment and development in cities do not belong to the easiest. Their implementation is also not trivial.

However, author believes that the reestablishment of order and consistency in the city's functioning should be a top priority for the governments of each city in Poland. We have to remember also, that attempts of achieving it are determined by current investment decisions and made improvements. Therefore, it is essential that all urban transformation which are going to happen should include the use of information and communication technologies, systems operating in harmony and in favor of the environment and conducting researches on city's sustainable development based on knowledge.

Described in the paper concept of the urban knowledge provides the basis for further consideration on city's development determined by new management techniques. To produce a resource in the form of urban knowledge is not only to achieve positive economic impact on the city, region, or country but mainly aims to prepare themselves to meet the future needs of civilization, currently generated by the society, classified as information. In the information society the dominant role plays nothing else but knowledge. That why every city in Poland, according to the author assumptions, should seek to create, to clarify their urban knowledge. Especially, when it is in line with the objectives of knowledge management. It means, that city can contribute to make better decisions, accelerate information processes, reduce unnecessary costs, increase satisfaction or level citizens' lives quality.

References

[1] Albescu F., Pugna I., Paraschiv D., Business Intelligence & Knowledge Management – Technological Support for Strategic Management in the Knowledge Based Economy, Revista Informatica Economica, no 4(48)/2008

[2] Albescu F., Pugna I., Paraschiv D., Cross-cultural Knowledge Management, Informatica Economica vol.13, no 4/2009

[3] Baltzan P., Phillips A., Business Driven Information Systems, second edition, McGraw-Hill Irwin, New York 2009

[4] Bergeron B.m Essentials of Knowledge Management, John Wiley & Sons, New Jersey 2003

[5] Błaszczuk A., Brdulak J.J., Guzik M., Pawluczuk A., Zarządzanie wiedzą w polskich przedsiębiorstwach,Szkoła Główna Handlowa, Warszawa 2004

[6] Global Footprint Network, http://www.footprintnetwork.org/en/index.php/GFN/page/frequently_asked_questions/ [2015-03-15]

[7] Jakubczyc J., Mercier-Laurent E., Owoc M.L : What is Knowledge Management? Baborski A. (red.). Research Paprers of Wrocław Economic Academy no 815, Wrocław 1999

[8] Jakubowski T., Zarządzanie wiedzą w firmach konsultingowych, Gazeta IT nr 7, listopad 2002

[9] Karamalla-Gaiballa E., Matouk K., Zarządzanie wiedzą w przedsiębiorstwie a jego potencjał ludzki:

[10] Marciniak K., Owoc M., Knowledge Management in the Interactive Portal for Decision Makers on InKOM Example, INTERNATIONAL SCIENCE INDEX 9(1) 2015, eISSN: 1307-6892, p.705-712

[11] Nycz M., Business Intelligence in Enterprose 2.0 in Knowledge Acquisition and Management no 232 Research Papers of Wrocław University of Economics, Publishing House of Wrocław University of Economics, Wrocław 2010, ISSN 1899-3192

[12] Richard T. Herschel and Nory E. Jones, Knowledge management and business intelligence: the importance of integration, in JOURNAL OF KNOWLEDGE MANAGEMENT VOL. 9 NO. 4 2005, Emerald Group Publishing Limited, ISSN 1367-3270

[13] Skyrme D.J., Knowledge Networking. Creating the Collaborative Enterprise, Butter-worth-Heinemann, Oxford 1999

[14] Tiwana A., The knowledge management toolkit, PTR 1999

[15] Turban E., Leidner D., Mclean E., Wetherbe J., Information Technology for Management Transofrming Organizations in Digital Economy, 6th edition, John Wiley & Sons, 2008

[16] R. Żubera R, S. Sudak (red.), „Koncepcja Przestrzennego Zagospodarowania Kraju 2030", Ministerstwo Rozwoju Regionalnego, Warszawa 2012

[17] Cambridge, MA Rep. ARCRL-66-234 (II), 1994, vol. 2.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1369–1373
DOI: 10.15439/2016F555
ACSIS, Vol. 8. ISSN 2300-5963

Knowledge Management &Risk Management

Eunika Mercier-Laurent
University Jean Moulin Lyon 3,
6, Cours Albert Thomas, 69008 Lyon, France
Email: eunika.mercier-laurent@univ-lyon3.fr

***Abstract*—**Knowledge Management methods, techniques and experience can bring a considerable help in preventing all kind of risks. The current economic context amplified the existing risks and introduced new ones such as environmental, political and social. Among the most critical risks of this century is this of lost of knowledge and "memory" in particular in industry involved in long-term activities.

After listing the principal risks and recalling of main approaches and techniques of Knowledge Management and its application for risk management an example of nuclear industry is given. Discussion follows in final conclusion and some perspectives are presented.

I. Introduction

HUMANITY have always been exposed to all kinds of risks - natural, industrial and those related to human activity and human nature, including motivations and passions. While the management of "soft" industrial risks is well established, the serious, unpredicted and rare risks are managed after the disaster.

Knowledge management approaches powered by artificial intelligence techniques can bring considerable help in risks prevention and management, but they are not sufficiently used.

This paper presents a broad spectrum of today risks and main works connecting these two topics. A short recall of principal knowledge management methods and disciplines follows. Some techniques for risks prevention and fixing are mentioned. These purposes are illustrated by a case study in the field of nuclear energy management. This paper ends by conclusions and perspectives for future work.

II. Typology of Risks

Among the industrial risks we can distinguish those caused by a malfunction or by the "human factor".

Risks resulting from human activity such as air and water pollution, destruction of natural ecosystems and global warming are the consequence of the ignorance of natural ecosystems, that we are the part, of the context and motivations as selfishness and desire to become rich quickly.

The pride and power can also lead to irresponsible decisions at high political and management levels, as for example wrong choice of strategy.

Other risks arise from human nature and motivations, such as sabotage, crime and cybercrime, wars, and lack of communication or of compromise. Risks arising from ignorance, obscurantism or indoctrination can cause unwanted destruction, bombings and wars.

The perpetuation of the same mental patterns is a risk of not finding the right solution when solving problems, which is a strong barrier for innovation. For example, onboard car computers are programmed using traditional decision trees and the user wanting to go to a given place has to choose step by step instead of making a direct voice request. Some of them are still equipped with a mouse and using such a system when you drive may lead to accident.

The recent risks are related to the use of technology, digitization, virtualization, security of sensitive data on clouds, the unexpected impact of biotechnology, nanotechnology and others [1, 2]. The precautionary principle, very strong in France [3], is applied in some cases when the impact is not known or not considered. An alternative of applying systematically such a principle is the understanding and simulation of the potential impacts by a multidisciplinary and experienced team possessing the related knowledge.

Another risk could come from the lack of management of the intellectual capital. Consequently, knowledge and know-how are not explored because little known at the strategic level.

A simplified ontology of risks is presented in Figure 1.

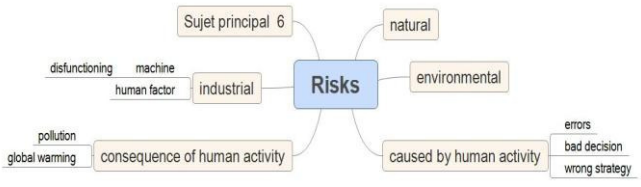


Fig. 1. Simplified ontology of risks.

The ISO 31000 norm considers as a high risk a theft, loss, obsolescence, no recovery and the lack of the use of knowledge. All have the negative effect on the achievement of objectives.

III. Related Work

The industrial interest in risk management is not new. We can find numerous work devoted to managing risks in energy [5, 6, 7, 8], in chemistry and pharmaceutical industry [9, 10, 11, 12] aeronautic and space [13, 14], automotive [15] and recently those related to the use of information technology [16, 17]. Other fields are also concerned, such as finance and banking [18], insurance and health [19].

In France there is a long tradition of managing industrial risks using statistic methods, recently extended by Bayesian and some data mining [20, 21]]. The Institut pour la Maîtrise des Risques (IMdR) groups research and industrial actors involved in this area and organize Lambda Mu, scientific conference and other events aiming in sharing experiences and solving problems collaboratively [22].

The approaches and techniques of Artificial Intelligence such as experts systems, case-based reasoning and multi-agent systems have been successfully used for managing risks as well for prevention as for fixing [23].

The development of methods for risks management and operating reliability, based on probabilistic networks from 1990 has been complementary to expert systems.

The bayesian networks, belief networks and Valuation Based Systems provide graphic representations of knowledge, more comprehensive by the user. These techniques are well suited to the uncertain context and in particular to epistemic uncertainty and are a part of toolbox currently used by engineers of risk control or operational safety, such as fault tree and Bayesian network [21]. These tools are currently used after collecting related data.

The professionals of risk management and operating reliability are used to apply known methods to existing data, opposite to the "knowledge thinking" approach based on problem understanding and collecting or mining the related knowledge vital for solving a given problem [23, 24]. The risk related to intellectual capital such as loss of knowledge and of competency, know-how is still weakly managed [4].

Globalization and hyper-competition has increased the risk in business. Political push to create start-ups should be associated with adapted policy, which is not the case today. The collected feedback from these experiences will be very helpful to improve the innovation policies.

The recent collogue of French Ministry of Defense pointed out the influence of climate changes on the raise of criminality [25]. Over twenty presentations with various points of views demonstrated the cause-to-effect dependence of delinquency and terrorism with climate change. But it is not only influence; the others factors such as unemployment and lack of the clear vision for the future and professional perspectives for young, low educated people, enhance the probability that they may choose Jihad.

To anticipate and manage such risks, it is necessary to gather and manage the related multidisciplinary knowledge and skills and use the right approaches and tools.

IV. Role of knowledge in risks management

Globalization, reinforced by the development of various modes of transportation, Internet and other information and communication technologies has changed the way we learn, work and do business and have introduced other risks. In search of cheaper workforce, companies relocate production and customer service taking a risk of knowledge loss and of customers' deception.

Production of spare parts in low cost countries may entail a risk of quality caused by the lack of conformity with the specifications and the use of alternative materials that can degrade the performance of given equipment.

Internal transfers, retirements, turnover and cross-competition generate impacts such as loss of know-how, loss of time to reinvent and remake what has been already done, additional costs, loss of competitiveness, responsiveness, security/safety flaws, which are the factors determining differentiation in an open economy. The knowledge transfer "at time", the use of decision support systems, knowledge base of products, projects and customer relations, managing the feedback from experience, e-learning and m-learning systems integrated to the overall knowledge flow connecting may help the involved actors in preventing and managing the risks.

The relocation may also lead to other risks, such as loss of employment and exclusion in developed countries.

Faced with the economic crisis in developed countries, governments consider innovation as the only remedy that can improve growth and boost the job creation. But entrepreneurship requires risk taking and know-how of managing a start-up for success cannot be acquired only in school, but mostly by doing or with an experienced mentor.

Innovation needs multidisciplinary knowledge to succeed and many products and services are knowledge intensive. Sometimes ignoring ergonomics and the real needs of users and customers may lead to the risk of non acceptation of given software or other products. Such products will not be used and may cause a company fail.

Computers and other technological devices and applications, currently used in industry are now present in all fields and their development accelerates. Computers, sensors and other electronic devices are more and more embedded into products. The products such as IoT, conceived according to the designers' vision only, autonomous devices, M2M communication and fully automatic client support systems raise the risk of elimination of human from the decision making.

If the machine thinks instead of us, it may influence the progressive lost of human cognitive capacity.

Can we trust the perfect operation of fully automated systems?

Internet is considered by many as a source of knowledge, but what is the reliability of information found on the web?

Decision taking based only on information found in various wikipedia, forums, and social networks maybe risky.

Can we find a reliable knowledge or information only in such results? Is it helpful and sure to avoid risks in the future? Is it efficient for knowledge discovery?

The prevention and management of all kinds of risks require an understanding of all facets, contextual knowledge and a method that fits to these components while the most of risks managers use the data and statistics.

A. *Prevent or fix?*

The industrial risks management methods include preventive maintenance, annual and five or ten-year controls of nuclear power plants, aircrafts and other complex equipments and feedback. The feedback is usually recorded in a large database conceived for all related professionals. The data processing after collection uses mainly statistics and the results are available to those concerned. Data is often missing because the persons that should record then do not understand what they have register because of weak adapted interface; sometimes we can find errors related to the data value such as missing comma, shifted, wrong or missing data [35].

In a few critical cases we manage the crisis after it happens (Tchernobyl, Fukushima, AZF Toulouse and others - http://www.aria.developpement-durable.gouv.fr/).

This involves repairing of the effects, as in Western medicine, but how to prevent these accidents using efficiently existing knowledge and expertise?

Practical experience demonstrates that sometimes a given problem, i.e. terrorist attack, may be detected considering weak signals often ignored. It is vital to collect all related knowledge including contextual and check consistency, especially when it comes from several sources.

While industrial maintenance is usually well managed, the related knowledge is often not organized and poorly managed. Nevertheless, knowledge plays a central role in our activities. But what is the necessary knowledge to collect and preserve in aim to avoid and control risks? What to preserve, how and why? It depends on the organizational/company strategy, that decides about the key knowledge and strategic competencies to maintain develop and preserve. Than the best adapted approach and tools will support the realization of this task.

V. KNOWLEDGE MANAGEMENT

In the early 1990s we saw the emergence of a new management method - Knowledge Management or business based on knowledge [26]. It has been a foundation of a new economy of knowledge [27].

The term "Knowledge Economy" is not new. Although it was proposed in the 1980s, such economy has always existed. It is about generating value from knowledge, know-how and experience. This is a radical change for companies that must now recognize and manage this intangible capital.

In this context, knowledge management can be defined as "*an integrated system of initiatives, methods and tools designed to create an optimal flow of knowledge within and throughout an extended enterprise to ensure stakeholders success"* [28, 29]. Stakeholders are partners, distributors and customers. Success includes among others leadership, industrial performance, efficiency, quality and safety.

The purpose of this approach is to organize and manage knowledge applicable or applied by the company or organization, for its activities related to current or future business.

Our research focuses on critical knowledge and know-how of experts, which in majority of cases is not formalized. While knowledge transfer has always been the scope of the trainers, we often ignore that knowledge can also be "transmitted" to the computer using something else than traditional databases [7]. Artificial intelligence has invented very early methods and techniques for knowledge acquisition, discovery and processing by computer. They have been tested successfully for over 50 years and are now encapsulated in tools.

Many decision support systems have been designed to assist users through expert systems, help in process control, in medical and industrial diagnostics, to provide design assistance, solve complex problems with constraints such as logistics, scheduling, planning and others. AI techniques are also useful for intelligent e-commerce, effective matching of offer and demand, very useful for recent platforms. All these applications should be organized as a Knowledge Flow. Knowledge modeling methods such as KADS for conceptual modeling, ontologies and other models [24] should be adequate to the nature of given knowledge. Graphical interface will facilitate their collection for preservation and sharing, but also reusing for a variety of applications.

The development of methods based on probabilistic networks in risk management and operational safety in the early 1990s also focuses the use of expertise. Bayesian networks, transferable belief models (TBF), the Valuation Based Systems use a graphical representation of knowledge which facilitates the understanding of data. These techniques seem well adapted to the context of uncertainty (including epistemic uncertainty) and generalize the methods commonly used by traditional engineer in risk management or operational safety, such as fault tree and Bayesian network. Current applications include risk analysis, reliability analysis, and analysis of the human factor [21].

Deep learning, born from the experience and the need to explore appropriately gigantic amounts of data connects several automatic machine learning techniques (machine learning) as data mining, natural language processing and image mining, among others.

At this stage a distinction between thinking "data" and "knowledge" must be considered. While the first focuses on the collection of data, the second focuses on the collection of

knowledge, essential for knowledge-based problem solving.
Knowledge transfer from human to computer via different knowledge models is now mastered and should be generalized. Many knowledge bases exist, but the knowledge capitalization process is not widespread. It is very often postponed, due to lack of time and underestimated by many companies. But it is a false excuse, because they continue losing time using still "data approach" which generates huge amount of data that should be "cleaned" and analyzed.
"Knowledge approach" may seem difficult for those who are used to work with data, because it requires changing a mental schema. However the trend of knowledge management and a necessity to preserve knowledge of retiring experts handed capitalization up to date.

For over two decades, companies and organizations have experimented Knowledge Management, beginning with their motivations and perspectives - as integrated in a company strategy, by managing of Intellectual Capital, Business Intelligence or by industrial applications [23].
Many companies have taken steps, but for the most part, without considering the feedback from Artificial Intelligence. Internal networks of experts (communities of practice) are among the "easy and quick" initiatives aimed at mitigating the risks related to the loss of knowledge and experience, allowing the sharing of "best practice", such networks requires a facilitator, who can progressively become Chief Knowledge Officer.

Two major factors hindering the implementation of KM as a management method are following:

- is the necessity of demonstrating the utility of this approach and its ROI,
- misunderstanding of the role the Chief Knowledge Officer must play; it is not a social position, but a cross communication and facilitation.

The initial objective of Knowledge Management remains improving of the innovation capacity of enterprises and organizations by better use of talents, knowledge, detection of opportunities, integration of feedback and use of the best fitting technologies.
As Knowledge Management involves stakeholders, the respect of ethics and trust are vital to achieve a common goal – to be successful together.
It is not easy to implement because it involves spending time to understand before doing, thinking "knowledge" and to change attitudes: listening instead of push, collaborating instead of competing, collecting and sharing experience and feedback, evaluating the benefits in terms of tangible and intangible values.

The knowledge to preserve includes not only this associated with long life technical equipments, such as nuclear power plants, but also the ability to solve problems. This is an art to gather individual and collective knowledge serving to solve a specific problem, but also the know-how in considering the various contexts that influence decision taking.

If we consider the risk of crime and cybercrime, solving this problem is not just to find the guilty and put him to jail, but to understand the real causes and motivations of involved persons and their environment. As mentioned before, the study of French Ministry of Defense has demonstrated the link between climate change and crime. For example, the drought caused by deforestation and intensified cultures lead to the depletion of the earth, causing famine and pushed more vulnerable to selling drugs or stealing.

The risk prevention requires awareness of their existence and their possible consequences, the ability to solve complex problems, holistic and system thinking and to select the necessary knowledge to prevent, minimize the causes, and impact. If this approach had been applied in the case of German Wings the company could have prevented the accident because the knowledge on the health of employees and their relations are part of the overall approach.

Finally the KM approach producing the best results is those that starts with an application addressing the real needs, developed with an incremental approach "do small and think big".

VI. EXAMPLE: NUCLEAR INDUSTRY

In France we have 18 nuclear plants, as shown in Fig 2.

Fig. 2. Map of French nuclear power plants [30].

Some of them have been built over 60 year ago. Nuclear plants represent three main risks: loss of initial knowledge, ageing of component materials and management of pollution at the end of life. Designers of these plants are retired. The initial knowledge and context of the whole lifecycle related to plants has been transmitted through documents and reports. There are still some elements lost because all details of lifecycle, replacement of elements, minutes from meetings, verbal experts' exchanges are not transmitted. Reports from the programmed control also exist, but it is a known fact that everything is not written [7]. Related database contains only data. Some expert systems have been

developed in specific critic fields, such as ageing of materials and their consequences on maintenance [36], mastering of ageing,

Last year Paris has been hosted the COP21 and the strong international lobbying claims to stop the oldest power plants. On the other hand, the management of nuclear waste is a long term process and requires an updated long term memory. ANDRA (Agence Nationale pour la Gestion des Déchets Radioactif) involved in managing nuclear waste is aware of the problem and has initiated a Knowledge Management activity [31]. As there are not only one organization involved in the whole nuclear plants cycle of life they initiated with ANR (Agence Nationale de Recherche) a collaborative program supposed to bring together all stakeholders in aim to organize and manage the related knowledge [33]. The task is not trivial, because the majority of involved actors retains knowledge and do not realize the consequences of such an attitude. We expect their awakening.

Despite a global awakening on planet and living protection the most of industrial people focus on quick business and do not assess the multiple impacts of their activity on our biosphere [34].

VII. Conclusion

Facing the systemic risks generated by globalization, quick business, vertiginous progress in technology and replacing human by machines requires different way of thinking and problem solving. Preventing and management of such risks require the "knowledge thinking" and considering the multidisciplinary knowledge and the related context. Deep problem understanding will guide the choice of Knowledge Management method and well suited tools. AI techniques support the managing the flow of knowledge by systematic collection and exploration of related knowledge and experience and knowledge. However a global awakening and change of attitudes is required. It can be done using i.e. serious games and various simulators. The future work related to risks management focus on the risks prevention by understanding the key factors through the games.

References

[1] Crichton M. "Prey", *HarperCollins*, 2002

[2] Fukuyama F. "Our Posthuman Future: Consequences of the Biotechnology Revolution, *Picador* 2002

[3] Le Déaut J-Y., Sido B. "Le principe d'innovation" - *Compte rendu de l'audition publique du 5 juin 2014 et de la présentation des conclusions les 4 et 26 novembre 2014, Rapport Senat* n° 133, 2014

[4] Edvinson L., Hofman-Bang P, Jacobsen K. "Intellectual capital in waiting – a strategic IC challenge", *Handbook of Business Strategy*, Vol. 6 Iss: 1, pp.133 – 140, 2005

[5] Bousquet N. "Analyse bayesienne de la durée de vie de composantes industrielle", PhD, University Paris XI, 2006

[6] Cohen B.L. "Risks of nuclear power", *The Health Physics Society*, Univ of Michigan, 2005

[7] Dourgnon A., Roche C., Mercier-Laurent E. "How to Value and Transmit Nuclear Industry Long Term Knowledge" *ICEIS* (2), 323-326, 2005

[8] Leveson N.G. " Risk Management in the Oil and Gas Industry", May 17, 2011, http://mitei.mit.edu/news/risk-management-oil-and-gas-industry

[9] Adis W. "A Risk Modeling Framework for the Pharmaceutical Industry", *Communications of the IIMA*, 2007 Volume 7 Issue 1

[10] Barbier P. " Urban Growth Analysis Within a High Technological Risk Area. Case of AZF Factory Explosion in Toulouse (France), CASITA Project, CASITA PROJECT, Ecole Nationale des Sciences Géographiques, 2003

[11] Meel A., L. O'Neill L., Levin J.H. and Seider W.D., "Operational Risk Assessment of Chemical Industries by Exploiting Accident Databases" , *Journal of Loss Prevention in the Process Industries*, 20(2) 113-127, 2007

[12] Grabowski H.G., Vernon J.M., "Returns to R&D on new drug introductions in the 1980s, *Journal of Health Economic*, Volume 13, Issue 4, 1994, p. 383-406

[13] Cook R. "Experts in uncertainty: Opinion and subjective probability in science", *Oxford University Press*, 1991

[14] Rosetta Mission http://www.dlr.de/dlr/en/desktopdefault.aspx/tabid-10394/, 2016

[15] Autonomous vehicles - Consideration for Personal and Lines Insurers, Munich RE, 2015

[16] Jurison J. "The role of risk and return in information technology outsourcing decisions" *Journal of Information Technology* (1995) **10,** 239–247

[17] Saripalli P. "QUIRC: A Quantitative Impact and Risk Assessment Framework for Cloud Security", *IEEE 3rd International Conference on Cloud Computing*, 2010

[18] Risk Management in Banking*, INSEAD Program*, 2016

[19] Dusansky R., Koç Ç. "Implication of the Interaction Between Insurance choice and Medical Care Demand", *Journal of Risk and Insurance*, Vol 77, p.129-144, March 20

[20] Lannoy A. Mercier-Laurent E., de Miramon B., Obama J.M., "Maîtriser la connaissance pour maîtriser les risques" *Documentaliste, Sciences de l'Information*, Vol 51, N° 3, 2014

[21] Sallak M. " Réseaux values (VBS) appliqués à la maîtrise des risques et à la sûreté de fonctionnement dans le domaine ferroviaire" , *Journée IMdR*, Paris, 2013

[22] Lambda Mu http://ipgr.fr/2015/04/20eme-congres-lambda-mu-11-12-et-13-octobre-2016-saint-malo/

[23] Mercier-Laurent E. " Innovation Ecosystems", Wiley, 2011, chapters 4et 5

[24] Mercier-Laurent E. "Rôle de l'ordinateur dans le processus global de l'innovation à partir des connaissances", *Habilitation pour diriger les recherches*, University Jean Moulin Lyon 3, 2007

[25] "Climat et Défense, quels enjeux ? The Implication of Climate Change for Defence", October 14, 2015, Paris, http://www.defense.gouv.fr

[26] Drucker P. "The New Society of Organizations", *Harvard Business Review*, September-October, 1992

[27] Amidon D., Formica P. and Mercier-Laurent E. "Knowledge Economics : Emerging Principles, Practices and Policies", Tartu University Press, 2005

[28] Amidon D. "The Innovation Strategy for The Knowledge Economy", *Butterworth Heinemann*, 1997

[29] Jakubczyc J. Mercier-Laurent E. and Owoc M. "What is Knowledge Management?", *KAM*, 1999, Wroclaw, Poland

[30] http://www.world-nuclear.org/information-library/country-profiles/countries-a-f/france.aspx.

[31] H. Bienvenu H. "La gestion des connaissances au sein de l'Agence Nationale pour la gestion des Déchets Radioactifs (ANDRA) ", *Qualitique*, N° 256

[32] http://www.agence-nationale-recherche.fr/financer-votre-projet/appels-ouverts/appel-detail0/appel-a-projets-andra-optimisation-de-la-gestion-des-dechets-radioactifs-de-demantelement-2015/

[33] Mercier-Laurent E. "The Innovation Biosphere – Planet and Brains in Digital Era", *Wiley* 2015

[34] Dourgnon A. L'art et la méthode de préservation des connaissances à longue durée de vie, EDF R&D, H-P1A-2012-02513-FR, 2013

[35] Mahé S. Développement et capitalisation des connaissances techniques pour la maîtrise du vieillissement, *Qualitique* N° 256 p. 41-44, 2014

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1375–1378
DOI: 10.15439/2016F129
ACSIS, Vol. 8. ISSN 2300-5963

QtBiVis: a software toolbox for visual analysis of biclustering experiment

Artur Pańszczyk
AGH University of Science and Technology,
Dep. of Automatics and Bioengineering,
al. Mickiewicza 30,
30-059 Kraków
Email: panszczyk.artur@gmail.com

Patryk Orzechowski
AGH University of Science and Technology,
Dep. of Automatics and Bioengineering,
al. Mickiewicza 30,
30-059 Kraków,
Email: patrick@agh.edu.pl

***Abstract*—In this article we introduce QtBiVis - a novel software intended for the comparative analysis of biclustering results. This modular tool has been efficiently implemented in C++ with Qt framework GUI. It may be successfully used for coverage analysis of the results of biclustering as well filtering or sorting biclusters by Gene Ontology (GO) identifiers or bicluster enrichment values. It may also be useful for parameter studies of biclustering algorithms. In future releases we plan to add different modules for visualizing and comparing different GO terms and biclusters.**

I. Introduction

MICROARRAY technology has become a subject of multiple biological experiments since its theoretical foundations in 1980's and first application in 1995 [1]. As labeled nucleic acids immobilized on a solid surface proved to be capable of monitoring the expression levels of nucleic acids molecules, the technology has been predominately used for measuring in parallel multiple gene expression patterns. It has also gained wide scope of application in disease diagnostics, drug discovery and comparative genomics.

The result of microarray experiment after background adjustment, normalization and summarization at the probe level is structured into a data matrix of real numbers, in which each value corresponds typically to a gene expression level under the specified condition. The whole microarray data may contain up to tens of thousands of gene expressions. Biclustering algorithms have been applied to identify groups of genes that show resemblance under particular subsection of conditions. Multiple biclustering methods have been developed so far [2], [3], [4]. Different metrics have been adapted to measure gene expression level [5], [6]. The collection of the most recognizable biclustering approaches applied to GDS datasets is included in Eren et. al. [7].

A. Methods of visualization of biclusters

The most popular way to visualize a single bicluster uses a heatmap, in which cells are colored based on the gene expression level. This perspective has been implemented in multiple software tools among which the most popular are BiVoc [8], BiVisu [9], BicAT [10] and its extension BicAT-Plus [11], BicOverlapper [12], [13], Furby [14], Bicluster Viewer [15] or BiGGEsTS [16]. A single bicluster is usually resorted and drawn in the upper left corner with its rows and columns rearranged.

The second popular visualization method, which is very useful for gene expression profile comparison, uses parallel coordinates, in which conditions are vizualized on horizontal axis. Gene profiles are represented by lines, which join corresponding values of corresponding conditions. This perspective is popularly used by multiple software tools (including BiVisu, BicAT, BicOverlapper or Bicluster Viewer).

A widespread visualization method for multiple biclusters uses heatmaps or heatmaps with dendograms. For example hierarchical biclustering is represented by a tree and a heatmap, in which rows are reordered to fit the recognized bicluster. This perspective is offered for example by BiGGesTS.

There are also more sophisticated visualization perspectives, which are provided by some of the available tools. For example BicOverlapper offers a transcription regulatory networks view, wordcloud or bubble map perspective. Furby presents graphically the attracting force between each pair of the biclusters. This resembles a class diagram, which is popular in relational databases. The number of rows and columns shared between each pair of biclusters are reflected by the transparency of the interconnecting line.

B. Motivation

The major motivation for QtBiVis is to provide a fast and reliable software, which would be able to support the researchers in parameter study of the biclustering methods. We believe that the analysis of biclustering experiments based on the inspection of biclusters' coverage may provide a novel insight for assessment of biclustering methods capabilities. It may also become helpful in determining the optimal setting of parameters for each algorithm.

The second justification of QtBiVis emergence are limitations of the available software. As the majority of the tools has been implemented in Java, existing tools encounter different performance issues during analyzis of hundreds or thousands of biclusters. This limits analysis of complex experiments in which thousands of biclusters need to be visualized and compared simultaneously. QtBiVis has been implemented in C++ with Qt used for graphical user interface.

978-83-60810-90-3/$25.00

II. METHODS

In this article we present QtBiVis, an open-source toolbox, which implementation may be found at github.com/Archi0/QtBiVis. In this section we present an overview of QtBiVis and detailed information about its design.

A. Main features of QtBiVis

The QtBiVis tool, which is herein presented, has been implemented in C++ programming language with Qt 5.5 framework used for graphical interface design. A Dynamic_bitset from Boost C++ library has been used for performing bitwise operations. The main features of QtBiVis include:

- loading main data set which contains microarray data,
- loading files with biclusters with Gene Ontology ID (GO ID) and p-values,
- filtering biclusters by a specific GO ID,
- calculations, plotting and saving results of the degree of coverage of the bicluster environment,
- displaying information about a single bicluster with its values, labels, level of coverage, GO IDs and p-values,
- plotting statistics of the bicluster based on average of values in columns and standard deviations of the values in columns in examined cluster,
- calculating, plotting and saving information about the relation between number of occurrences of value in different biclusters and size of the bicluster,
- drawing a heatmap based on the number of occurrences of a given value in the loaded clusters.

B. Overview of application

The main workflow of QtBiVis includes loading a microarray dataset and definitions of series of biclusters from multiple biclustering experiments. Information about loaded biclusters is shown in the table on the right-hand side of the main window (see Fig. 1). Each row of the table contains a Gene Ontology ID, p-values (before and after multiple test correction) and the identifiers of rows and columns of the bicluster. Filtering is applied for biclusters after entering a value in "GO Filter" text area.

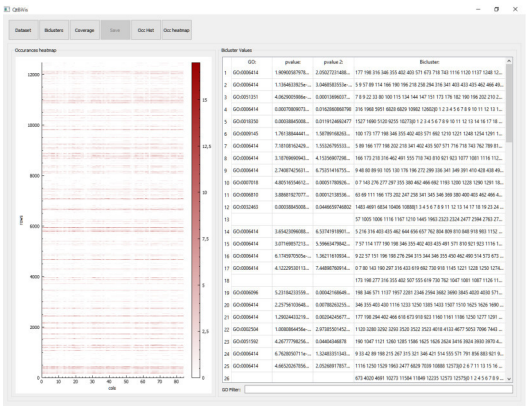


Fig. 1. The main window of QtBiVis.

The application provides access to a standard overview of a single bicluster (i.e. its expression values, labels of rows and columns and parallel coordinates plot), as well as tools for visual comparison of relations between multiple biclusters. This includes the statistics of coverage and the frequency of occurrences of expression values presented in form of heatmap and histogram.

C. Analysis of a single bicluster

A single bicluster analysis overview, which is demonstrated in Fig. 2, contains biclusters values with rows and columns names, gene ontology ID's with p-values before and after correction and neighborhood of the selected bicluster with percentage values.

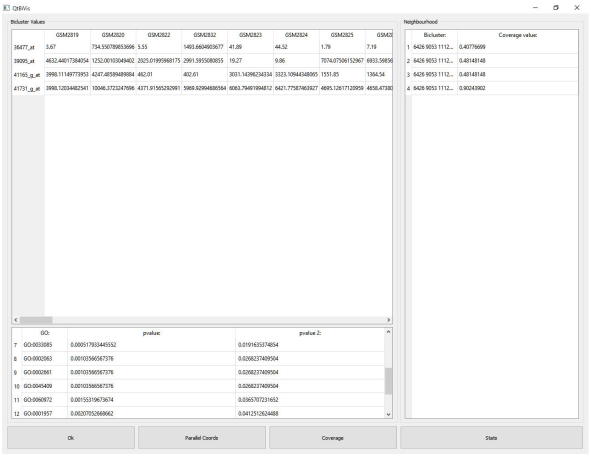


Fig. 2. A single bicluster overview.

This perspective offers access to three different analysis tools. The first one, called "Parallel Coords", provides access to profile analysis using parallel coordinates plot (see Fig. 3).

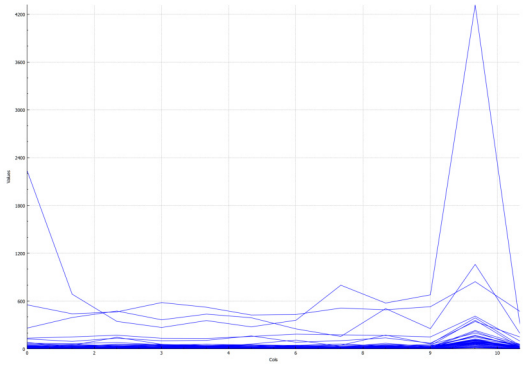


Fig. 3. Gene profile analysis of a single bicluster.

The second one, named "Coverage", displays a histogram, which presents the level of bicluster overlap with respect to other biclusters (see Fig. 4). A histogram presents on horizontal axis a degree of coverage (i.e. a percentage of shared area with the bicluster) and on vertical axis - a number

of occurrences (i.e. number of biclusters that share the same area).

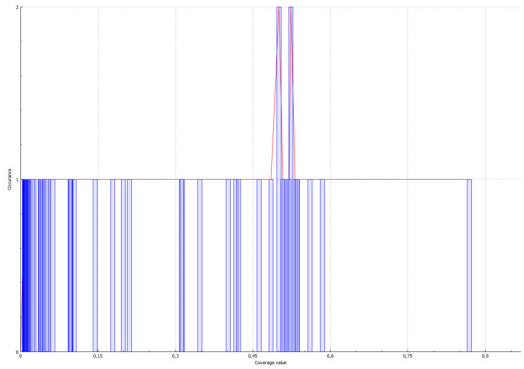


Fig. 4. Coverage analysis of a single bicluster.

The third one, "Stats", shows means and standard deviations of particular columns of a bicluster. In current version, for each column the plot displays mean and standard deviations as dots, whilst the average standard deviation of all columns is represented by a line (see Fig. 5).

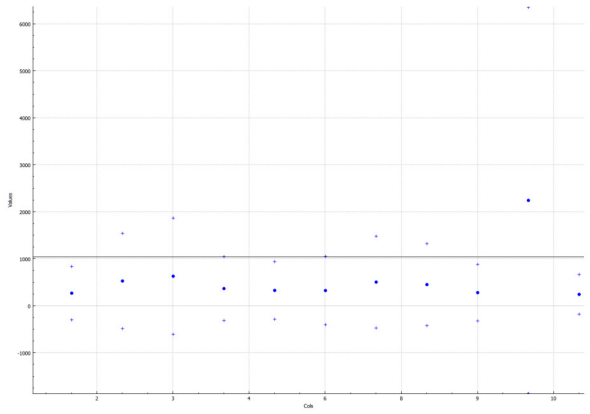


Fig. 5. Statistics of mean and standard deviation of the whole bicluster and its values in columns.

D. Coverage statistics

Another statistics, which is provided by QtBiVis, is analysis of the degree of bicluster intersection. Degree of overlap for two biclusters (A and B) is determined by Jaccard index (1), which takes into account the number of intersecting biclusters' elements with respect to the total area occupied by both biclusters.

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} \tag{1}$$

Clicking on the "Coverage" button on main window calculates the percentage coverage for every loaded bicluster, which is presented on a histogram (see Fig. 6). The button "Save" stores the results in a selected file.

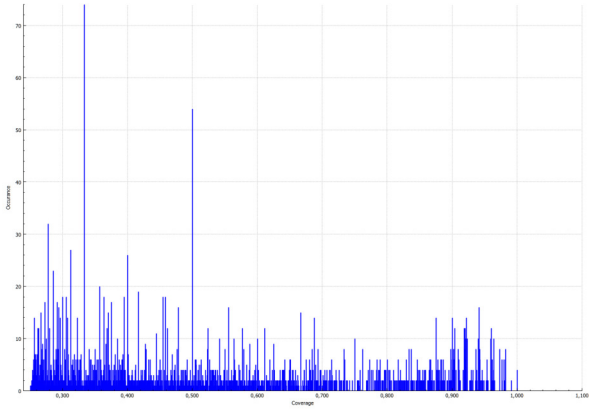


Fig. 6. A histogram presenting the degree of coverage of multiple biclusters with each other.

III. Implementation

QtBiVis has been designed as a modular application. Each module is responsible for providing a different perspective. Several coding optimizations have been used to reduce the computation time of certain calculations. For example row and column of a bicluster are represented in application by bits. If the row or column belongs to a bicluster it is set to '1', otherwise it remains '0'. Bitsets are used for computations of their intersections or unions.

A. Application design

The class diagram of the application is presented in Fig. 7. The main components of the application have been presented hereafter.

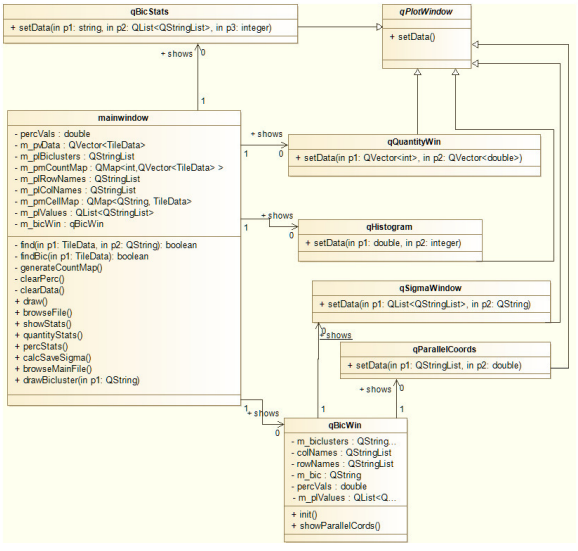


Fig. 7. Class Diagram of QtBiVis.

1) MainWindow: The MainWindow module is responsible for loading microarray data as well as loading, displaying and filtering the biclusters.

2) QBicWin: This module is used for displaying statistics about a single bicluster: its values and rows and columns labels. The module also presents the neighboring biclusters by showing the percentage of coverage. Different modules for single bicluster analysis may be triggered from here (i.e. qBicStats, qParallelCoords and qSigmaStats).

3) QBicStats: This module calculates of degree of coverage for the examined biclusters with the rest of biclusters.

4) QHistogram: The QHistogram module is responsible for calculation and presenting a histogram based on degree of coverage for every detected bicluster.

5) QSigmaStats: QSigmaStats is another module, which calculates and shows common statistics of values in bicluster, such as an average of values, a standard deviation and an average of standard deviation of values in each column of bicluster.

6) QParallelCoords: The QParallelCoords module is responsible for vizualization of a single bicluster with parallel coordinates perspective.

7) QQuantityWin: QQuantityWin is used for displaying the histogram based on the number of occurrences of values in microarray data.

IV. Results

For demonstration purposes eight GDS datasets have been taken, which have been previously used by Eren et al. [7]. As the original dataset haven't been provided by the authors in supplementary materials, we downloaded their copies from Gene Omnibus and tried to follow the authors' preprocessing procedure (i.e. missing value imputation, validation and Benjamini-Hochberg multiple value correction [17]). Unfortunately, we didn't manage to obtain similar results to the authors. The reason for this is that Eren et al. specify neither the method of missing values imputation, nor the method of gene universe creation (i.e. algorithm used for filtering features from an ExpressionSet, which exhibit a little variation or where GO or Entrez Gene identifiers are missing). Thus, we decided to parse the original file with their biclustering results, which have been publicly available. The file has been split into separate folders corresponding to each dataset and divided into the separate files for each biclustering method respectively.

The detailed analysis of the results of coverage of each biclustering method remains out of scope of this paper and has been mentioned only to demonstrate the usage of the QtBiVis.

V. Conclusions & future work

The QtBiVis tool performs extraordinarily fast for analysis of multiple biclusters and their coverage. Its capabilities include filtering by a specific GO term across all detected biclusters and sorting by p-values. Based on this, we hope to assess if the relation between the uniqueness of the detected biclusters and their biological significance exists.

We also plan to use the software for parameter study of biclustering algorithms. By analyzing the results obtained from a series of biclusters, each run with different input parameters, QtBiVis may allow us to support analysis of the most commonly detected biclusters by a specific algorithm. Thus, we hope to assess how the input parameters affect the stability of the results.

The current version of the algorithm doesn't take full advantage of the enrichment level of specific biclusters. In future releases of the software we plan to add different statistics, which would present the impact of the biclustering uniqueness on biological significance. This may be performed for example by visualizing only those biclusters or GO terms, which significance is higher than the adjustable threshold.

Acknowledgments

This research was funded by the Polish National Science Center (NCN), grant No. 2013/11/N/ST6/03204. This research was supported in part by PL-Grid Infrastructure.

References

[1] M. Schena, D. Shalon, R. W. Davis, and P. O. Brown, "Quantitative monitoring of gene expression patterns with a complementary dna microarray," *Science*, vol. 270, no. 5235, pp. 467–470, 1995.

[2] S. C. Madeira and A. L. Oliveira, "Biclustering algorithms for biological data analysis: a survey," *Computational Biology and Bioinformatics, IEEE/ACM Transactions on*, vol. 1, no. 1, pp. 24–45, 2004.

[3] A. Oghabian, S. Kilpinen, S. Hautaniemi, and E. Czeizler, "Biclustering methods: biological relevance and application in gene expression analysis," *PloS one*, vol. 9, no. 3, p. e90801, 2014.

[4] B. Pontes, R. Giráldez, and J. S. Aguilar-Ruiz, "Biclustering on expression data: A review," *Journal of biomedical informatics*, vol. 57, pp. 163–180, 2015.

[5] P. Orzechowski, "Proximity measures and results validation in biclustering—A survey," in *Artificial Intelligence and Soft Computing* (L. Rutkowski, M. Korytkowski, R. Scherer, R. Tadeusiewicz, L. A. Zadeh, and J. M. Zurada, eds.), vol. 7895 of *Lecture Notes in Computer Science*, pp. 206–217, Springer Berlin Heidelberg, 2013.

[6] B. Pontes, R. Girldez, and J. S. Aguilar-Ruiz, "Quality measures for gene expression biclusters," *PloS one*, vol. 10, no. 3, p. e0115497, 2015.

[7] K. Eren, M. Deveci, O. Küçüktunç, and Ü. Çatalyürek, "A comparative analysis of biclustering algorithms for gene expression data," *Briefings in Bioinformatics*, 2012.

[8] G. A. Grothaus, A. Mufti, and T. Murali, "Automatic layout and visualization of biclusters," *Algorithms for Molecular Biology*, vol. 1, no. 1, p. 15, 2006.

[9] K.-O. Cheng, N.-F. Law, W.-C. Siu, and T. Lau, "Bivisu: software tool for bicluster detection and visualization," *Bioinformatics*, vol. 23, no. 17, pp. 2342–2344, 2007.

[10] S. Barkow, S. Bleuler, A. Prelić, P. Zimmermann, and E. Zitzler, "Bicat: a biclustering analysis toolbox," *Bioinformatics*, vol. 22, no. 10, pp. 1282–1283, 2006.

[11] F. M. Al-Akwaa, M. H. Ali, and V. M. Kadah, "Bicat_plus: An automatic comparative tool for bi/clustering of gene expression data obtained using microarrays," in *Radio Science Conference, 2009. NRSC 2009. National*, pp. 1–8, IEEE, 2009.

[12] R. Santamaría, R. Therón, and L. Quintales, "Bicoverlapper: a tool for bicluster visualization," *Bioinformatics*, vol. 24, no. 9, pp. 1212–1213, 2008.

[13] R. Santamaría, R. Therón, and L. Quintales, "Bicoverlapper 2.0: visual analysis for gene expression," *Bioinformatics*, p. btu120, 2014.

[14] M. Streit, S. Gratzl, M. Gillhofer, A. Mayr, A. Mitterecker, and S. Hochreiter, "Furby: fuzzy force-directed bicluster visualization," *BMC bioinformatics*, vol. 15, no. Suppl 6, p. S4, 2014.

[15] J. Heinrich, R. Seifert, M. Burch, and D. Weiskopf, "Bicluster viewer: a visualization tool for analyzing gene expression data," in *Advances in Visual Computing*, pp. 641–652, Springer, 2011.

[16] J. P. Gonçalves, S. C. Madeira, and A. L. Oliveira, "Biggests: integrated environment for biclustering analysis of time series gene expression data," *BMC research notes*, vol. 2, no. 1, p. 124, 2009.

[17] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: a practical and powerful approach to multiple testing," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 289–300, 1995.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1379–1382
DOI: 10.15439/2016F20
ACSIS, Vol. 8. ISSN 2300-5963

Hard lessons learned: delivering usability in IT projects

Krzysztof Redlarski, Paweł Weichbroth
Gdansk University of Technology
Faculty of Management and Economics Gdansk, Poland
Department of Applied Informatics in Management
ul. Narutowicza 11/12, 80-233 Gdańsk, Poland;
krzysztof.redlarski, pawel.weichbroth@zie.pg.gda.pl

***Abstract*—Effective project management requires the development of a realistic plan which aims to ensure the success of the project and ultimately deliver a high quality product to customers. However, experience shows that the majority of software vendors managing projects suffer from numerous problems to provide usability in IT solutions and complete a project in a given time with success. In this paper we discuss, analyze and synthesize the outcomes of a study conducted among IT firms in Poland. As a result, we have identified eight stimulants and three non-stimulants that affect the usability of software products, which later were stratified into three levels. Finally, we outline some of the lessons learned, summarized and expressed as a set of eleven goal-oriented rules.**

I. Introduction

PROJECT management is now one of the fastest growing areas of science, and its development and continuous improvement is sure to be a long-term up-to-date research topic. Many firms, during the realization of new IT projects, come across problems with their completion in line with the starting assumptions (i.e. budget, scope and schedule). Strong competition on the market may force them to make difficult decisions to optimize costs. Thus, they aim to find more effective methods of project management which allow them to increase project effectiveness. On the other hand, they look for adequate methods to manage a given project, best suited to its specific needs. However, in practice they have to face many different kinds of problems and obstacles, associated with the usability of the developed product, which directly translates into the success of the entire project. Evidently, a low-end usability product, as a whole, will be perceived as low-quality by its customers.

In this paper, we present and discuss the results obtained from a survey conducted among IT firms located across Poland. Our research includes the main findings of the survey, conducted among project managers, and an analysis of project documentation. We have identified, and later described, the factors (stimulants and non-stimulants) that affect the quality of software products in selected IT projects. In the end, we outline a set of eleven goal-oriented rules as a guideline for software vendors, regarding pro-usability organization and cooperation with end-users.

II. IT Product Life Cycle

The dynamic pace of change and the development of the IT society bring about numerous problems related to the choice of effective methods of project management, as well as their use and interaction with IT products. Taking appropriate action in the early stages of the development of IT products and services plays a significant role in effective product management (Fig. 1).

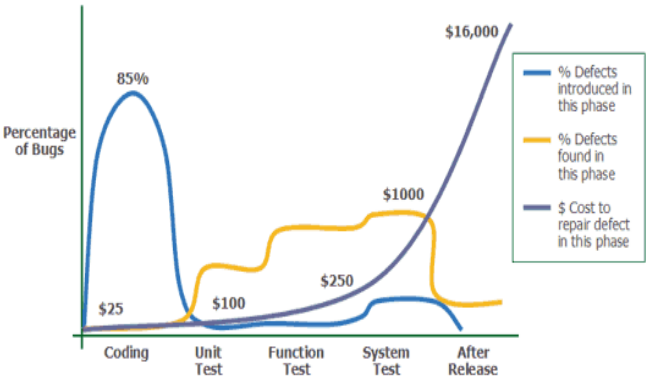


Fig. 1. Cost of the removal of defects implemented in stages of an IT project [1].

Defects implemented in the product in its initial stage of development are particularly costly from the perspective of the product life cycle. They may lead to the eventual failure of the entire project if their number exceeds permissible limits. Hence, it is particularly important to start the process of product quality provision at the earliest possible stage of any IT project.

III. The Development Of IT Project Management

The awareness of business decision-makers related to IT product development for end-users is constantly changing. Until the 1980s developers were only focused on software design and perspectives, entirely excluding end-users from the development process. Fortunately, from the early 80s till now, the end-user role has changed. The software vendors have realized that a user-oriented approach can advance product quality on a higher level. In such a way, end-users gradually began to be actively involved in projects, and their

importance over the years successively increased [2]. Among others, the so-called "human factor" has become an important and relevant element in ensuring high usability of the final product. To put it simple, the problem was to ensure efficiency, effectiveness and satisfaction with everyday use. On the other hand, as a result of a lack of acceptance on the part of the end-users, the sales value of new products can significantly fall, along with the profit of the software vendor.

A user-oriented approach is reflected in changes in both the technology of IT product development, as well as in the methodology of IT project management. Currently, one can particularly observe an intensive usage of soft approaches in project management. For example, the Agile manifesto states that "*our highest priority is to satisfy the customer through early and continuous delivery of valuable software*". In our opinion, the customer is all different kinds of users, with respect to such values as trust and transparency in communication facility. In turn, in literature of Human-Computer Interaction (HCI), one can distinguish an approach that addresses user-oriented design, defined as User Centered Design (UCD). This approach is characterized by active user participation in the whole design process (analogously to Agile), whilst putting the biggest emphasis on the fulfillment of the requirements and needs of the expected users. However, on the one hand, the goal is to close a project within a given time and budget, while, on the other hand, to deliver a product with high usability. In such a case, an impartial and tenable compromise between each project stakeholder group is a must.

The attempt to identify the factors affecting the provision of high product usability, undertaken in this analysis, is particularly important, because the correct identification of these factors allows us to determine not only the changes taking place in current developments in IT projects, but also to indicate the relationship occurring between selected factors and stakeholder groups.

IV. Usability Factors

We undertook both quantitative and qualitative research among five IT firms:

- *Infovide-Matrix JSC* (IVMX), a company which, according to the IDC report [3], is in the top 10 IT services companies in Poland, with revenues of USD 56.12 million, and employing up to 500 people; the company specializes in consulting, developing and deploying product solutions and technologies in the IT area;
- *Aiton Caldwell JSC*, a firm specializing in providing SaaS services (Software as a Service), based on remote software sharing via the Internet. The leader of the Polish market of hosted telecommunications services for SMEs and individual clients. The company employs about 50 people, and has about 40% of the Polish market share of VoIP (Voice over Internet Protocol);
- *Kolibro LLC*, a company specializing in consulting and deploying internal communication management systems, i.e. corporate intranet portals and dedicated applications for business process management. The company employs about 40 people, and has mainly used Microsoft SharePoint technology and Enterprise Project Management;
- *One2tribe LLC*, a firm specializing mainly in game development for mobile devices. It has, in its portfolio, a number of applications used in popular social networks for marketing purposes. The company employs about 40 people and has been present on the market for eight years;
- *Webstruments GP*, a firm specializing in the development and implementation of Internet and intranet applications. The company employs about 10 people and has been in business for 7 years, having a portfolio of very interesting projects for major firms, mainly from the public sector.

We purposely selected miscellaneous firms to examine, offering a diverse range of IT products, varying in the total number of employees, annual income, scale of operations and duration of business activity. In fact, the study included the largest IT companies in Poland, which have vast experience and resources involved in projects (taking into account workforce, materials and costs), but also micro-enterprises, which often equally effectively compete in the market with the biggest market participants. The major goal of the study was to identify the factors affecting the assurance of usability in IT projects, and to assess their impact on the final success of a given project. The study has been divided into two parts:

- The first part was to *conduct interviews* with project managers, who, whilst sharing their insights, would point to the factors which, in their opinion, influence product usability and the final success of a project.
- The second part was to *analyze the documentation* of a given project, which described how the project was realized, and indicated the methods and measures applied in order to assure product usability by the personnel responsible for the project.

Next, the identified factors were examined, using an electronic questionnaire in which respondents answered the following questions:

- In your opinion, *what was the product usability like*? The answer was represented by a 5-point quality scale (very poor, poor, average, good, and very good), where 1 stood for very poor quality, and 5 for very good quality.
- In your opinion, *did the project end in success*? The answer was represented by a 5-point quality scale (total success, partial success, hard to say, partial failure, and total failure), where 1 stood for total failure of the project, and 5 meant that the project was totally successful. The study was conducted on a group of 30 respondents (participants of IT projects).

The obtained results allowed factors from the first part of the evaluation and the relationship between them to be verified, and juxtaposed with their usability in IT projects.

Factors with a positive impact (stimulants) on product usability in IT projects are:

- s_1: *product specification,* is a set of documented requirements to be satisfied by the software product;
- s_2: *user participation,* refers to the assignments, activities and behaviors that a user (human) performs during the product development process; these can take a variety of forms: direct (individual action) or indirect (represented by others), local (physical attendance) or remote (physical non-attendance), formal (using legal agreements, documented and obliged) or informal (through informal relationships, discussions, and tasks), performed alone or in collaboration with others (in groups or teams), moderated (driven by other individuals) or liberated.
- s_3: *project analysis*, is a retrospective inspection of the finished project;
- s_4: *project team size,* is the total number of hired staff, assigned to perform a particular set of tasks;
- s_5: *project team experience,* is the total number of similar projects finished by the assigned group of people;
- s_6: *project team knowledge,* is the collective, tacit and incorporeal expertise in a specific domain, demonstrated by the assigned group of people;
- s_7: *project manager experience,* is the ability to manage and mitigate risks, and skills gathered in using risk management methods in previous projects;
- s_8: *project manager knowledge,* reflects the mastery of domain expertise in project management as applied to a particular field or multiple fields, which brings natural authority and solid strategic insight.

On the other hand, we distinguished three factors with a negative impact (non-stimulants) on product usability in IT projects, including:

- d_1: *project innovativeness,* is originality by virtue of introducing a new software product (newness to the developing firm);
- d_2: *project implementation method,* reflects the assumptions, principles and best practices that are used in the software product life cycle.
- d_3: *project outsourcing,* is the presence of external services hired in the project to perform particular tasks.

In addition, we distinguish three separate groups of factors, embodied in three dimensions named, respectively from top to bottom: project, team and project manager (table below).

TABLE I.
THREE DIMENSIONS OF USABILITY FACTORS

level	stimulants	non-stimulants
project	s_1, s_2, s_3	d_1, d_2, d_3
team	s_4, s_5, s_6	-
project manager	s_7, s_8	-

Each individual factor was assigned to a distinct level and each is subsequently represented by a goal-oriented rule in the next section. The project level is the most general and is the only one to have all three non-stimulants assigned as well as the first three stimulants.

The project team and the project manager experience (s_5 and s_7) are particularly important to ensure high project usability. It is also recommended to carry out a so-called post-project analysis after completion of a project. This allows usability errors to be significantly eliminated in subsequent projects. The analysis of the projects surveyed also indicated that a certain level of formalization of the project documentation favored ensuring high usability in a project.

Among the non-stimulants that could negatively affect product usability in IT projects, project innovativeness (d_1) was highlighted. This factor greatly conditioned the participation of end-users in a project. Firms that developed projects of a rather reproductive nature (and thus largely repetitive), would substantially, or even entirely, resign from user participation in a project, thereby aiming to reduce the cost of the entire project. In this case, the risks associated with the failure to provide usability were attributed to the project manager or the whole company responsible for the project. The analysis of the methods used in the surveyed IT projects indicated that there is no direct relationship between one method of project implementation (classic, agile, informal), and the usability of a product. A project which ended in complete success would primarily use informal methods of project management, which combined two main approaches used in project management: the classic and the agile. Moreover, project outsourcing (partial or full range), proved to be a factor negatively influencing product usability. Companies carrying out projects based on the services of external companies fared far worse compared to companies independently pursuing projects.

Therefore, analyzing the development of project management methodologies (see par. 2), it should be stated that there is a continuous tendency to increase the active participation of end-users in projects. However, despite the fact of the increased importance of their participation [4], practice among businesses can be seen to limit the participation of users in selected types of projects. This applies mainly to projects where the product being developed is repetitive and imitative. The active participation of the user in a project corresponds to the necessity to increase the budget of the project, since their presence increases the total costs of the project. Hence, companies willing to increase the chances of a project being realized, and increase the probability of acquiring new customers, deliberately restrict user participation in their projects. Reducing the costs associated with user involvement in a project is, in fact, a significant part of the project budget. The experience of those responsible for the project is, therefore, particularly important for ensuring product usability.

V. Lessons Learned

Rule no. 1 states that *the higher the degree of documentation formalization, then the higher the usability and quality of the product being developed*. In other words, if you and your client pay more attention to detail, use formal specifications and modeling languages, and share the same dictionary, then the product will demonstrate perceivable values to its users concerning accuracy, efficiency, and satisfaction.

In current practices of usability testing we observed four prototype methods: thinking-aloud, subjective ratings, history files and eye-tracking. The major goal of usability testing is to identify and eliminate obscure problems with the user interface. Rule no. 2 states that *the higher the user participation in the life-cycle of software design, then the higher the usability and quality of the product being developed*. An important step in addressing these issues is the recent effort to develop a better measure of user participation, constructed by combining different methods.

The post-project analysis documents the results of conducting a deep and broad project assessment from its kickoff to finalization. Rule no. 3 states that *the higher the degree of comprehensive post-project analysis, then the less expense and risk together with higher usability are expected in future solutions*. A unified analysis should capture successes and failures, challenges and threats, in such areas like: planning, resources, scheduling, development and design, testing, communication, team and organization, solutions and tools.

On the team level we identified three stimulants that concern its size, experience and knowledge. Rule no. 4 states that *a bigger project team size boosts a richer essence in knowledge flow between its members and engages more experience and skills, which together move toward higher usability of the product being developed*. We also have to keep in mind an adequate arrangement of members of a goal-oriented team.

Satisfaction (subjectively pleasing) is one of Nielsen's five usability attributes, especially important in home computing environments (e.g. video games) where entertainment value comes first, before compatibility, efficiency and reliability that, on the contrary, play a major role in work-related environments. Rule no. 5 states that *a more experienced team delivers a technology more preferred by a user, reflected by higher satisfaction of usage.*

The ISO 9241 presents a set of usability heuristics which applies to the interaction of people and information systems. In this standard this interaction was called a "dialogue" and the following seven "dialogue principles" were defined: conformity with user expectations, controllability, error tolerance, self-descriptiveness, suitability for individualization, suitability for learning and suitability for the task. They apply to broad and narrow groups of artifacts. The former includes two sets of recommendations: (1) a presentation of information, defined in three main areas such as: organization of information, graphical objects, and coding techniques and (2) user guidance that covers general advice: prompts, feedback, status information, error management and on-line help; the latter includes four sets of recommendations: menu dialogues (such as pop-up, pull-down and text-based menus), command dialogues (command line interface), direct manipulation dialogues and form-filling dialogues. Rule no. 6 states that *the higher the knowledge capital of the team developing a particular artifact, then the less ambiguous the dialog between the end-user and the product will be.*

One of the most desirable attributes of the project manager is action management, which means acting in such a way that leads to the achievement of expected results through the successful and timely completion of activities and the delivery of the product. Rule no. 7 states that *the higher the experience of the project manager, then the more proper usability practices are applied.*

Rule no. 8 states that *the higher the knowledge demonstrated by the project manager, the more relevant the allocation of time and resources to various means*. Regarding rules no. 7 and 8, a skilled project manager should wield knowledge, skills and experience commensurate with the complexity, risk and size of the project.

In a low level innovative project, its specification is often created *ad hoc*, based on the project manager's experience and knowledge. However, knowing your audience is key to any successful innovation and this is notably true for new products [5]. Rule no. 9 states that *higher project innovativeness requires a higher degree of documentation formalization*. On the other hand, the level of active user participation in testing usability should always be considered in conjunction with the project innovativeness.

The context specificity of explicit users' requirements does not allow usability to be compared across different IT systems, unless they share comparable functionalities and user interfaces [6]. Rule no. 10 states that *the context of use analysis sets up the borders of usability heritage between heterogeneous systems*.

Numerous undesirable consequences of IT outsourcing have been reported so far i.e. service debasement, absence of cost reductions, disagreements. Rule no. 11 states that *the higher the level of outsourced resources hired in the project then the lower the internal usability of captured know-how*.

References

[1] *Why Unit Testing? Develop software with confidence*. http://www.agitar.com/solutions/why_unit_testing.html

[2] K. Redlarski, *The impact of end-user participation in IT projects on product usability*, ACM, 2013.

[3] *International Data Corporation*. http://idcpoland.pl/eng.

[4] E. L. Wagner, and G. Piccoli, *Moving beyond user participation to achieve successful IS design*. Communications of the ACM, vol. 50 (12), ACM 2007, pp. 51–55.

[5] D. Jelonek, and I. Pawełoszek, *Technologie semantyczne w zarządzaniu platformą otwartych innowacji*. Informatyka Ekonomiczna, vol. 4(30), 2013, pp. 169–180.

[6] P. Weichbroth, and M. Sikorski, *User Interface Prototyping. Techniques, Methods and Tools*. Studia Ekonomiczne 2015, 184–198.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1383–1387
DOI: 10.15439/2016F123
ACSIS, Vol. 8. ISSN 2300-5963

Speculative Query Execution in Relational Databases with Graph Modelling.

Anna Sasak-Okoń
University of Maria Curie-Skłodowska
in Lublin
Pl. Marii Curie-Skłodowskiej 5, 20-031 Lublin, Poland
Email: anna.sasak@umcs.pl

***Abstract*—In computer architecture, speculative execution is the process of executing instructions ahead of their normal schedule[1]. Grama et al.[2] introduce the concept of speculative decomposition as a possibility to execute one or more of possible branches in parallel with computation which are expected to determine the branch choice. The following paper introduces the method of speculative query execution in relational databases. Query queue can be seen as a line of sequential instructions and thus changing their order can result in some errors. Author introduce a middleware called the Speculative Layer which, based on a specific graph representation, executes some additional Speculative Queries. Results of those Speculative Queries can be used while executing queries from the queue providing a befit which is a shorter response time. The paper describes the process of graph modelling for groups of queries in order to initiate speculative computations, metrics used to evaluate Speculative Queries and experimental results for a test database and a group of input queries.**

I. Introduction

ORIGINS of the speculative execution are the early works of branch prediction[3][4]. The sequential semantics imposes a certain order in which instructions should be loaded, decoded, executed and ended[5]. Code branches, usually dependent on some logical conditions, disturb the fluency of loading and executing, causing delays. As an attempt to prevent delays there were experiments to predict a branch direction and to execute an instruction or a group of instructions in advance.

In general, there are three types of Thread Level Speculations (TLS)[6]:

- Control Speculation – origins from branch prediction strategy. The assumptions could be made based on some static (e.g. op codes) or dynamic values [7][8].
- Data Dependence Speculation – If two instructions are fully independent, only then the parallel execution is possible. Before memory access instructions are executed, the addresses they refer to are often undetermined. To prevent data load from an address where store should be executed earlier, a certain secure mechanism should be introduced[8].
- Data Value Speculationis – is expected to prevent data dependency with the value prediction mechanisms which allow to propagate data values to succeeding instructions in advance.

II. Related Work

There is already much research done around the world in adopting speculative execution in database computations.

Polyzotis N. and Ioannidis Y[10]. introduce speculation as a parallel, intelligent technique of query processing assistance. Exploiting idle time[11] of the system the application processes some asynchronous database manipulations which in case of success would be beneficial for the final query.

Barish G. and Knoblock C.A.[12][13] in order to overcome the limits imposed by binding patterns between data sources propose mechanisms of applying speculative execution for Information Gathering Plans. The general process of speculative execution involves issuing operations ahead of their normal schedule based on data (hints) received earlier in the plan.

Hristidis V. and Papakonstantinou Y[14] analyse speculative computations for ranked queries. Authors create a speculative version of a ranking algorithm which in case of a slower data source assumes speculatively that there are no tuples satisfying the preference function and thus can return top-N results faster but with some inaccuracies.

Reddy P., Kitsuregawa M.[15], Ragunathan T., Krishna R.P.[16][17] deal with speculative execution for transaction protocols in database systems. They introduce the speculative protocol (SL). With SL the waiting transaction is able to access locked data as soon as blocking transaction produces its images.

III. The Speculative Layer

An inspiration for this experiment is an idea of speculative execution briefly described in the previous section. Authors proposes a speculative execution mechanism for relational databases executing SQL queries of accepted structure which are CQAC queries with additional IN and LIKE operators. What is important, an analysed database must show a specific use template. Databases which suit our interests usually have to execute similar queries from different users. What's more, data modifications are rare and usually concentrate around some fixed points.

A queue of queries awaiting for execution, called the input queries, presents an interesting analogy to the sequential order of instructions. The consecutive queries can, like sequential instructions, show some dependencies. On the other hand,

978-83-60810-90-3/$25.00

carefully identified similarities allow to use some of the results many times[18][19][20].

A model described above is implemented as an additional middleware between users and DBMS called the Speculative Layer, which dynamically supports execution of input queries. The Speculative Layer, based on precise Speculative Analysis, creates a subset of data in RAM called further a Speculative DB. The data from the Speculative DB used while executing an input query improves system throughput and shortens users waiting time.

All actions of the Speculative Layer are controlled by the main worker thread called the Manager. In each step N input queries are analysed. This group of input queries is called the Window of Speculations. Based on those analysis, supported by a specific graph representation described in Sections IV and V, Manager assigns tasks to K Worker Threads.

In particular Manager implements the following functions of the Speculative Layer:

1) System Start.
   All initial actions required for the first run of the Speculative Layer. In particular graph representations are created for N input queries from the Window of Speculation. Next, those representations are combined to create the Queries Multigraph ready for the Speculative Analysis.
2) Nonspeculative Query Execution.
   Process of executing the first input query from the Window of Speculation, called the Nonspeculative Query. If there are Executed Speculative Queries assigned to the Nonspeculative Query, then it must by modified so it would use those results. If there are more than one Executed Speculative Query assigned, then the choice which to use is based on the values of the defined metrics - Horizontal an Vertical Selectivity. Vertical Selectivity models the reduction of the number of columns while the Horizontal Selectivity approximates of number of records returned by the Speculative Query.
3) Speculative Analysis.
   Process of the Queries Multigraph analysis which identifies speculation points and generates Awaiting Speculative Queries. The other result of the Speculative Analyses is a Speculative Queries Multigraph i.e. Queries Multigraph with additional speculative edges representing points and types of speculations.
4) Window of Speculation Move.
   After the Nonspeculative Query is executed and its results are returned to the user the Window of Speculations moves. It means that the representation of executed Nonspeculative Query in the Queries Multigraph is replaced by the representation of the next input query from the queue.
5) Speculative Query Execution.
   If there are idle Worker Threads then, if it is possible, they should be assigned available Speculative Queries from the Awaiting Speculative Queries List, according to the values of aforementioned metrics. The highest execution priority should have those queries which provide the highest potential reduction of records or/and can be used by the most of Input Queries.
6) Executed Speculative Query Assignment.
   After an Awaiting Speculative Query is executed and becomes an Executed Speculative Query, it has to be assigned to the specific Input Query/Queries from the Window of Speculations, which marks the possibility to use its results.
7) Speculative DB Refreshment.
   When the Speculative DB reaches its maximum size, it has to be reduced. The reduction process consists in removing the results of chosen Executed Speculative Queries based on its characteristics. First to remove are always those queries with the highest Vertical and Horizontal Selectivity and those which are used the least often.

IV. Query Graph

A. CQAC queries

Each accepted CQAC query is represented by its Query Graph $G_Q(V_Q, E_Q)$. Graph creation rules follow the example of [22][23] works, and are as follows. Each Query Graph Vertex is one of three types:

- Relation Vertex (R_i) – one for each relation,
- Attribute Vertex (A_j^i) – one for each attribute,
- Value Vertex $(\Omega) = \{Val_j^i | A_j^i\}$ – one for each value or set of values.

Each Query Graph Edge is one of the following types:

- Membership Edge – $e_\mu : R_i \underline{^\mu} A_j^i$ – one between relation R_i and each of its attributes A_j^i from SELECT clause,
- Predicate Edge – $e_\theta : A_j^i \underline{^\theta} \{Val_j^i | A_k^m\}$ – one for each predicate of WHERE clause $A_j^i \theta \Omega$, where θ is one of accepted operators. Ω is a single value or a set of values (Val_j^i) or an attribute (A_k^m) for JOIN condition.
- Selection Edge – $e_\sigma : R_i \underline{^\sigma} A_j^i$ – one for each predicate of WHERE clause $A_j^i \theta \Omega$, where θ is one of accepted operators. Ω is a set of values (Val_j^i) or an attribute (A_k^m) for JOIN condition.

B. Embedded queries

Each embedded query q_m is represented by its own query graph joined with its parent query Q graph in the following way – for each predicate $A_j^i \theta A_k^m$ where $A_j^i \in (Q$ WHERE clause) and $A_k^m \in (q_m$ SELECT clause), there is a predicate edge between A_j^i and A_k^m.

C. Modifying queries

Next to SELECT queries there are also modifying queries which are accepted by the Speculative Layer and thus need to have a proper graph representation. For each type of modifying query there is another edge type representing a possible change in the database state:

- DELETE: $e_\delta : R_i \underline{^\delta} A_j^i$ and $e_\delta : A_j^i \underline{^{v\theta}} \Omega$ where θ is one of accepted operators and Ω is a set of values,

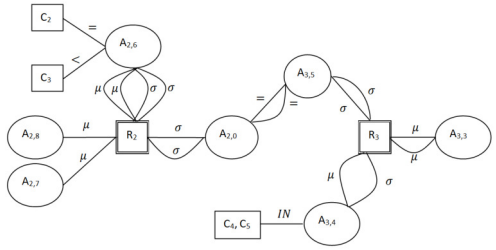


Fig. 1. Queries Multigraph

- INSERT: $e_\eta : R_i \underline{\eta} R_i$
- UPDATE: $e_\upsilon : R_i \underline{\delta} A_j^i$ and $e_\upsilon : A_j^i \underline{\upsilon\theta} \Omega$ where θ is one of accepted operators and Ω is a set of values.

V. Queries Multigraph

To represent a group of queries with one graph some additional rules have to be defined. Such graph $G_s(V_s, E_s)$ will be called the Queries Multigraph or QM. QM Vertices set is an union of Vertices of all component query graphs: $V_s = V_{q1} \cup V_{q2} \cup \ldots \cup V_{qn}$. QM Edges set is a multiset of all component query graphs edges: $E_s = E_{q1} + E_{q2} + \ldots + E_{qn}$. This way multiple edges of the same type are allowed. It is important for the Speculative Analysis process raising the importance of some edge connections. Fig.1 presents the Queries Multigraph representing three following component queries:

- SELECT $A_{3,3}, A_{2,6}, A_{2,7}$ FROM R_2, R_3 WHERE $A_{2,0} = A_{3,5}$ AND $A_{2,6} = C_2$
- SELECT $A_{2,6}, A_{2,8}$ FROM R_2, R_3 WHERE $A_{2,0} = A_{3,5}$ AND $R_2.A_{2,6} < C_3$
- SELECT $A_{3,3}$ FROM R_3 WHERE $R_3.A_3, 4$ IN (C_4, C_5)

VI. Types of Speculative Queries

The process of Speculative Analysis is expected to determine a set of Speculative Edges. Those edges represent different strategies of creating Speculative Queries. Based on the results usage, there are three types of Speculative Edges which represent three types of Speculative Queries:

- Speculative Parameter – those edges/queries relate to the presence of embedded queries. Due to separating an embedded query as a Speculative Query, it is possible to use its results as a parameter in a parent query. The Speculative Parameter Speculation is identified the moment the query with an embedded select enters the Window of Speculation. An embedded query as a whole is added to the head of the Awaiting Speculative Queries List (Q^i_{PS}). As a consequence those queries are always first to be executed by the Worker Thread.
- Speculative Data – the aim of those speculative queries is to obtain and save in the Speculative DB a specific subset of records or/and attributes of a relation. The main goal is to create this subset so as it could be used while executing as many input queries as possible. Speculative Data queries are the most frequent in the group of identified speculations.
- Speculative State – those edges/queries relate to the presence of modifying queries. If there is a modifying query in the Window of Speculation then both Executed and Awaiting Speculative Queries are in danger of processing invalid data. Speculative State edges are referring to the modifying queries represented by $e_\delta, e_\eta, e_\upsilon$ edges. If a modifying query has a number K in an input queries queue, then all succeeding queries (K+) are in danger of processing invalid data.

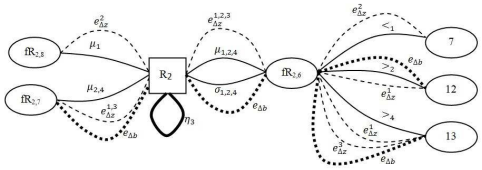


Fig. 2. Multigraph with Speculative Data and Speculative State edges

The Fig. 2 presents an example of Queries Multigraph with Speculative Data and Speculative State edges.

VII. Experimental Results

The Speculative Layer was implemented with C++ and Visual Studio 2013 with Pthread library and SQLite 3.8.11.1. Experimental results were obtained in Windows 8.1 64b with Intel Core i7-3930K and 8GB RAM.

A. Test Database and Test Input Queries

The Database used for experiments was generated with the TPC[21] data and structure generator. It consists of 8 relations storing 1GB data. It represents 150 000 orders from 150 000 customers which include chosen from 200 000 products delivered by 10 000 suppliers. Such database is a fine example of a medium sized Internet store. Eight SQL Query Templates were prepared and used to generate a set of Input Queries executed with the Speculative Layer.

B. Window of Speculation Size and the Number of Speculative Threads

At the beginning the series of experiments were conducted to determine the size of Window of Speculation and the number of Speculative Threads for which the experiments would continue. The size of the Window of Speculation stands for the number of input queries represented by the Query Multigraph and thus it determines the number of generated Awaiting Speculative Queries. Fig.3 presents how the Size of Window of Speculation affects the number of generated Awaiting Speculative Queries.

The number of executed Speculative Queries depends on the number of active Speculative Threads and not on the size of the Window of Speculation itself. In the Fig.4, two series of data are presented. First one, marked with black squares, presents the number of Executed Speculative Queries for the number of active Speculative Threads. The second one, marked with

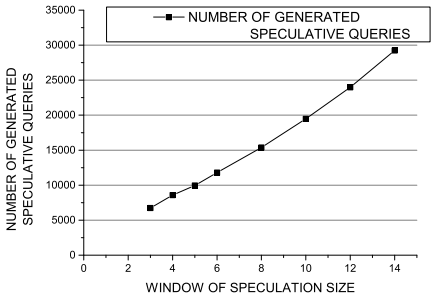


Fig. 3. The number of generated Awaiting Speculative Queries.

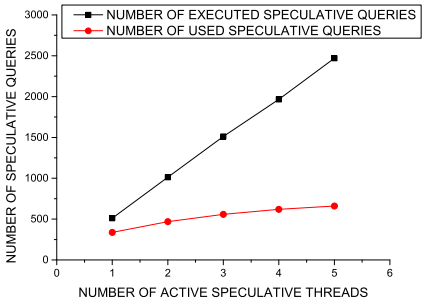


Fig. 4. The number of executed and used Speculative Queries.

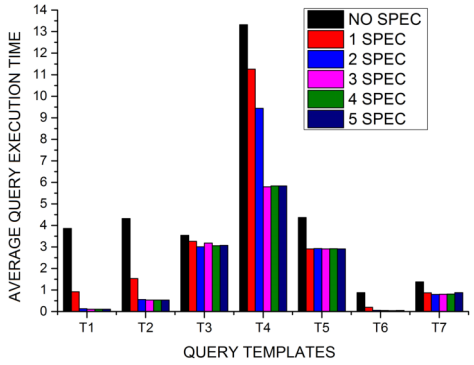


Fig. 5. The average Execution Time for Each Template for 0-5 Active Speculative Threads.

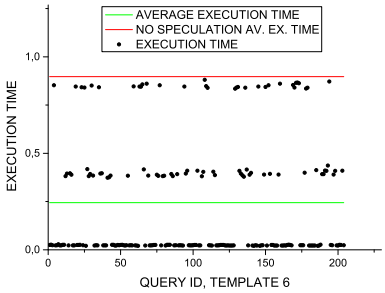


Fig. 6. Query Execution Times for Template 6.

red dots, presents the number of Used Speculative Queries for the number of active Speculative Thread. The experiment was carried for the Window of Speculation Size = 5.

Fig. 3 shows almost linear dependency between the Size of the Window of Speculation and the number of generated Awaiting Speculative Queries. Fig. 4 also presents almost linear dependency between the number of executed Speculative Queries and the number of active Speculative Threads. On the other hand the number of Used Speculative Queries hardly changes for the number of threads 3 and more and thus the percent of Used Speculative Queries is decreasing. Based on those observations it was decided that the further experiments would be carried for the Window of Speculation Size=5. The experimental results are presented for the set of 1000 input queries generated from Templates T1-T8. The size of Speculative DB is 700MB RAM.

C. Query Execution Times

Fig. 5 presents the reduction of average execution time for input queries of each Template. First column represents the sequential execution time when there were no active Speculative Threads. The following columns represent execution times obtained for each query Template for 1 to 5 active Speculative Threads, which execute Speculative Queries. It appears that the highest execution time reduction was obtained by initiating the first Speculative Thread (up to 55% execution time reduction for Template 1). Further improvement, up to 20% brings the second active Speculative Thread. Activating more than two Speculative Threads doesn't affect the Average Execution Time.

Fig. 6 presents execution times of input queries for Template 6, with the Window of Speculation Size=5 and 2 active Speculative Worker threads. In the picture there are two additional lines presented showing the average execution time with (green line) and without (red line) speculative execution. Template 6 was chosen for presentation as its queries show an interesting behaviour. There is a clear division for three groups of input queries. First group are the input queries executed without the opportunity to use results of Executed Speculative Query. They have the longest execution times and thus are located close to the red line. The remaining two groups are input queries which were able to use the results of Executed Speculative Queries, however, the obtained execution times vary significantly. It turned out the group with the lowest execution times had an opportunity to use results of the Speculative Query with the Horizontal Selectivity equal to 0,01. The rest of them had to use the results of Speculative Queries with the Horizontal Selectivity equal to 0,9 which are almost full copy of the original ORDERS relation.

Fig. 7 presents how many Input Queries of each Template were executed using results of the Executed Speculative Query. It is expressed as a proportional dependency where each column stands for 100% of Input Queries of each Template. As you can see Templates T3 and T4 are the least responsive

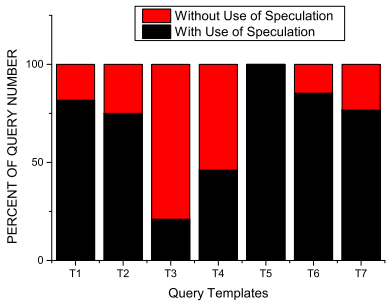


Fig. 7. The number of Input Queries which used the Executed Speculative Query Results as a proportional relation to the total number of Input Queries of each Template.

to Speculation and thus the average execution time reduction of those templates (Fig.5) was also the lowest. The reason is that those queries has low occurrence density (10%) in the Input Query Set. As a consequence its multigraph vertices has low Selection Degrees and thus they are rarely chosen to be executed by Speculative Threads. What's more, in both those Templates, T3 and T4, IN and LIKE operators are dominating. Those operators refer to the narrow subgroups of attribute values, making it especially difficult to generate Speculative Queries useful for more than one query from the Window of Speculation.

VIII. Conclusion

The following paper presents a model of Speculative Execution to support SQL query execution in relative databases. The Speculative Layer executing Speculative Queries is carefully described. In particular the details of an adopted query graph representation are presented. Experimental results obtained for the test database and a group of 1000 input queries are very promising. In case of Template 5, 100% of input queries were execute using the results of executed Speculative Query. Templates: 1, 2, 6 and 7 show around 75% and above execution with Speculative Query. All groups of queries show its execution time reduction: from 10%(Template 3) to 70%(Template 6). Further work should concentrate on the improvement of the number of input queries executed with the Speculative Query results (especially for Template 3 and 4). Worth consideration is also allowing more flexible structure of accepted queries and intensifying the number of modifying queries which would require more sophisticated speculation validation methods.

References

[1] D. Kaeli, P. Yew, "Speculative Execution in High Performance Computer Architectures," Chapman Hall/CRC, 2005, ISBN:978-1-584-88447-7.

[2] A. Grama, A. Gupta, G. Karypis, V. Kumar, "Introduction to Parallel Computing (Second Edition)," Addison Wesley, 2003, ISBN: 978-0-201-64865-2.

[3] E. A. Jr. Liles, B. Wilner, "Branch prediction mechanism.," IBM Technical Disclosure Bulletin, 1979, Vol.22(7), p. 3013-3016.

[4] J. E. Smith, "A study of branch prediction strategies.," ISCA 98 Conference Proceedings, 1998, New York, p.202-215, http://dx.doi.org/10.1145/285930.285980.

[5] D. Padua, "Encyclopedia of Parallel Computing A-D.," Springer, 2011, ISBN: 978-0-387-09765-7.

[6] A. Kejariwal, X. Tian, W. Li, M. Girkar, S. Kozhukhov, H. Saito, U. Banerjee, A. Nicolau, A.V. Veidenbaum, C.D. Polychronopoulos, "On the performance potential of different types of speculative thread-level parallelism.," International Conference on Supercomputing Proceedings., 2006, Cairns, p.24, http://dx.doi.org/10.1145/1183401.1183407.

[7] J. Šilc, T. Ungerer,B. Robič, "Dynamic branch prediction and control speculation.," Int. Journal of High Performance Systems Architecture, 2007, Vol.1(1), p.2-13, http://dx.doi.org/10.1504/IJHPSA.2007.013287.

[8] D. Kaeli, P. Yew, "Speculative Execution in High Performance Computer Architectures.," Chapman & Hall,CRC, 2005, ISBN:978-1-584-88447-7.

[9] S. T. Pan, K. So, J. T. Rahmeh, "Improving the accuracy of dynamic branch prediction using branch correlation.," International Conference on Architectural Support for Programming Languages and Operating Systems, 1992, Boston, p.76-84, http://dx.doi.org/10.1145/143371.143490.

[10] N. Polyzotis, Y. Ioannidis, "Speculative query processing," CIDR Conference Proceedings, Asilomar, 2003, p.1-12, http://dx.doi.org/10.1.1.11.8541.

[11] R. M. Karp, R. E. Miller, S. Winograd, "The Organization of Computations for Uniform Recurrence Equations," Journal of the ACM, 1967, Vol.14(3): p.563-590, http://dx.doi.org/10.1145/321406.321418.

[12] G. Barish, C.A. Knoblock, "Speculative Plan Execution for Information Gathering," Artificial Intelligence, 2008, Vol.172(4-5), p.413-453, http://dx.doi.org/10.1016/j.artint.2007.08.002.

[13] G. Barish, C.A. Knoblock, "Speculative Execution for Information Gathering Plans," AIPS Conference Proceedings, Toulouse, 2002, p.184-193, http://dx.doi.org/10.1.1.11.3505.

[14] V. Hristidis, Y. Papakonstantinou, "Algorithms and Applications for answering Ranked Queries using Ranked Views," The VLDB Journal, 2004, Vol.13(1), p.49-70, http://dx.doi.org/10.1007/s00778-003-0099-8.

[15] P.K. Reddy, M. Kitsuregawa, "Speculative locking Protocols to Improve Performance for Distributed Database Systems," IEEE Transactions on Knowledge and Data Engineering, 2004, Vol.16(2), p.154-169, http://dx.doi.org/10.1109/TKDE.2004.1269595.

[16] T. Ragunathan, R. P. Krishna, "Performance Enhancement of Read-only Transactions Using Speculative Locking Protocol," IRISS - Sixth Annual Inter Research Institute Student Seminar in Computer Science, Hyderabad, 2007.

[17] T. Ragunathan T, R. P. Krishna, "Improving the performance of Read-only Transactions through Speculation," Databases in Networked Information System, 2007, Vol.4777, p.467-474, http://dx.doi.org/10.1007/978-3-540-75512-8_15.

[18] A. Sasak-Okoń, M. Brzuszek, Speculative execution plan for multiple query execution systems, Annales UMCS Informatica, 2010, Vol 10(2), p.41-50,

[19] A. Sasak-Okoń, M.Brzuszek, The example of speculative execution for multiple query execution systems, Metody Informatyki Stosowanej, 2011, Vol.3(28), p.157-166, ISSN 1898-5297.

[20] A. Sasak-Okoń, M. Brzuszek, Graph modeling as a support technique for speculative computations in multiple query execution systems, Data Analysis Selected Problems, 2013, p.55-68, ISBN 978-83-7518-602-4.

[21] TPC benchmarks, http://www.tpc.org/tpch/default.asp, 2015.

[22] G. Koutrika, A. Simitsis, Y. Ioannidis, "Conversational Databases: Explaining Structured Queries to Users", 2009, Technical Report Stanford InfoLab.

[23] G. Koutrika, A. Simitsis, Y. Ioannidis, "Explaining Structured Queries in Natural Language.", ICDE Conference Proceedings, Long Beach, 2010, p. 333-344, http://dx.doi.org/10.1109/ICDE.2010.5447824.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1389–1396
DOI: 10.15439/2016F314
ACSIS, Vol. 8. ISSN 2300-5963

Searching for information and making purchase decisions in b2b online stores. The case of the technical articles wholesale

Łukasz Wiechetek
Maria Curie-Skłodowska University,
Plac Marii Curie-Skłodowskiej 5
20-031 Lublin, Poland.
Email: lukasz.wiechetek@umcs.lublin.pl

Mieczysław Pawłowski
Onninen Sp. z o.o. Email:
mpawlowski@onninen.com

***Abstract*—This paper aims to extend current research in the area of on-line business-to-business clients preferences. A quantitative research in a form of transaction log analysis (TLA) performed by the authors allow to conclude that search log and shop basket logs are sources of viable information about business customers. Transaction log analysis allows to identify different customer groups, such as searchers and buyers. It also allows for better customization of cloud of tags, so that the results are better tailored to customers' preferences. It turns out that a large list of products (result of search process) is not a deterrent to purchasing on-line. However, in order to facilitate the purchasing process b2b platform should offer additional filtering mechanisms.**

The main results of the research could be used by managers of professional e-commerce b2b platforms to better understand customers' needs and develop strategies to build long lasting partnerships.

Research limitations. The authors examined single b2b platform that may limit the findings generalizability, so the future research of other b2b platforms should be explored to validate the results.

Our future research should be performed to obtain the interplay between quantitative and qualitative methods of b2b clients' analysis and online shopping. It will allow for better understanding of customers' engagement and clients' unique organizational culture that may have an impact on purchasing decisions and building long time relationship.

I. Introduction

To avoid many pitfalls b2b companies should systematically analyze clients' decision making processes and improve supplier-client relationship. The Gallup research shows that in many cases b2b companies don't know the business clients opinions and needs, therefore they hardly need the feedback from the customers in order to sustain and develop business relationships [3]. The source of that feedback can be data stored in the Internet, social media platforms, that can be automatically summarized by IT systems [25, 26]. Data can be also collected by b2b platforms in a form of transaction logs (TL). Transaction logs are collected automatically and can be generated by applications, operating systems, network devices and other programmable hardware. Transaction log analysis (TLA) can be a good starting point for better understanding the purchasing preferences of business customers.

A. Collecting the data about b2b customers

E-commerce is rapidly developing phenomenon. Using information systems (IS) allows for gathering a lot of data about customers' behavior. Many companies believe that more data give better customer insights, however Gallup stated that it is not always true in b2b relations [1]. It is difficult to build lasting partnerships offering only good product or service price. This price related relationship will continue until the client finds a cheaper supplier.

In order to improve relations with customers Gallup advises to take the following steps [3]:

- Asses the company's knowledge about the business clients.
- Analyze the customer engagement drivers.
- Take the ownership of the relationship.

The above steps lead to better understanding the customer engagement and determine the strengths and weaknesses of the relationship. The weak points should be quickly identified and corrected and the strengths properly maintained. Taking the ownership of the relationship needs dialog with customer and continuous questioning about what was done well, in what areas the relation could be improved, what was missed, etc. The change should be sustainable.

As Gallup research shows fully engaged customers buy more and more often. The organic growth of the b2b business needs the understanding of business clients' needs. Some customer characteristics could be derived from the transition log analysis (TLA). Transaction logs are files that contain data collected automatically by b2b platforms. These files can consist of data about customers' search preferences: search phases, search time, session duration; buying preferences: price, discount, payment method; transportation preferences: duration, cost, volume, and many more. Therefore, the analysis of these files can be a good starting

point for exploring needs of business customers. To better understand these needs we should distinguish data stored in logs, generated by bots from the traffic generated by business users [27]. Well-identified needs are the foundation for further good relationships with customers.

The main drivers for purchasing decisions in b2b market are functionality, utility of the product and information about vendor. However in some cases just as important are some factors related to business customer personality [8]. Sometimes business clients act as common consumers and have human qualities. Therefore in order gain more profound knowledge about preferences of b2b clients pure quantitative research like transaction log analysis or knowledge extraction from the professional e-mails [28] should be complemented with psychological analysis. Chlupsa, Döhl, Lean and Hanoch claim that decision making processes are influenced by implicit motives [9]. In order to obtain a complete data describing customers some additional qualitative research like in-depth interviews, observations, and focus groups should be performed.

B. B2b customers purchasing decisions

When customer wants to buy a product he has to go through buying decision process. Psychologists have developed many models describing this process [11] most of them goes across the stages from the need to the purchase decisions. For example Engel's, Blackwell's, Kollat's model consists of five stages: need recognition, search for information, alternatives comparison, purchase decision and post-purchase behavior.

Business e-commerce customers need some tools preferably in the form of b2b platform that will allow them to convert needs into purchase. Therefore b2b platform should offer not only the mechanisms for searching the right product but also give the possibility for product comparison and finally offer a simple mechanism for filtering and purchasing. Furthermore, the additional mechanisms should collect information about the course of the purchasing process and also store the purchase history. This functionality results in higher hardware requirements but allows for better fitting the IT system to the needs of the business consumer and build customer loyalty.

IT solutions are not sufficient condition for the success. Khan, Naumann and Williams [12] examined factors that drive customer satisfaction and repurchase intentions. They explored Japanese business-to-business service customers. According to their findings personal interactions have great impact on repurchase intentions. The researchers observed that in Japanese context, product perception is less important than personal contact. They claim that personal business relationships are very important.

Belonax, Newell, Plank [13] investigated buyer perception of trust and expertise of the salesperson. They claim that this perception was higher in less important purchases than extremely important purchases. They also confirmed positive relationship between trust and expertise. Perception of trust and expertise is positively affected by the frequency of purchase contacts.

Cano, Boles and Bean [14] examined the communication media preferences in business-to-business transactions. They concluded that in most cases buyers and sellers prefer face-to-face or telephone communication rather than other types of communication tools. The communication ways vary throughout process purchase. For the efficiency of the sales process salesman must understand buyer's communication needs and adapt to this type of communication. Finally researchers conclude that communication process should be managed in order to be cost efficient in short time but also build stable long-run relationships.

There can be concluded that the purchase act in business to business environment is preceded by a number of preparatory steps. The buyer has to trust the seller and believe in his expertise. Good communication is crucial for building the trust. The communication preferred by b2b customers should be face-to-face but in many cases, through the development of technology it can be successfully supported by b2b platforms.

C. Building relationships with b2b clients with IT

Trusting and more satisfied clients are more likely to continue cooperation. Attracting and keeping customers can be achieved no only through functionality, utility of the product and information about vendor but also by offering additional services i.e. shipping cost and duration, payment form and conditions. Mingming and Parlar [10] used leader-follower game to check whether buyer is willing to increase the purchase value to get free shipping.

B2b platform can be also the tool for building customer loyalty. The structural equation modeling was used by Hsu, Wang and Chih [5] to explore how web platform characteristics influence customer loyalty and positive opinions. It was found that web site characteristics has a positive influence on relationships. Performed research showed that web site attributes can be good predictors for customer's loyalty, high-quality e-commerce platform results in more satisfied user.

Sila [6] collected surveys form 275 North American companies using b2b electronic commerce and found that the biggest contributor to b2b usage is scalability.

Taehee, Jonghoon, Junho, Sang-goo showed that the quality of b2b platform can be improved by using ontology-based product recommender system instead of systems based on the text-retrieval technique [7]. They addressed the results ranking problem by modeling the product ontology as a Bayesian belief network.

As we can see, there are many ways to build relationship with customer e.g. price, quality, delivery terms. B2b platforms can be an important part of maintaining

relationships with business customers. They play the role of an organization business card but in some sense also a kind of expert system. Moreover, the data recorded by the platform in the form of transaction logs can be a source of valuable information about customers' purchase preferences and can be used for increasing the efficiency of trade.

D. Shopping basket analysis

B2b platform can be used as a tool for collecting data about customers, tool for building relations with customer or image building tool. An interesting data for the analysis that can be collected by the b2b platform are the shopping baskets. Analysis of the shopping cart can provide a valuable information about the needs and behavior of on-line and off-line consumers. We can find many researches in the area of: what (and when) is being purchased [15][19], shopping basket size and value analysis [16], effect of sales promotions on shopping basket [17][18], the cart filling time [20], the influence of smart shopping card on shopping behavior [21], or determinants of consumers' online shopping cart abandonment [22][23][24].

Mallapragada, Chandukala and Qing [15] investigated the impact of what product and where being shopped. They examined 773,262 browsing sessions resulting in 9,664 transactions and noticed that website communication tools are positively correlated with basket value. They confirmed also that number of products available in the on-line store is positively associated with duration of the visit and the value of the basket.

Anesbury, Nenycz-Thiel, Dawes and Kennedy [20] examined in details behavior of 40 shoppers by recording (screen recording) online shopping trip of new, inexperienced grocery customers. They concluded that shopping process is quite fast. The 12 item shopping took less than 10 minutes. Clients mostly have chosen items from the first results page. Customers used rather default display options presented on the on-line store. Finally researches stated that in terms of time and efforts grocery on-line clients are quite similar to off-line shoppers.

List of the determinants of consumers' online shopping cart abandonment was presented by Kukar-Kinney and Close [22]. They confirmed that categories of cart abandonment determinants should include entertainment value (using web page only for information purpose), concern about costs, waiting for better price, privacy and security concerns. They concluded that customers leave the basket not only because of dissatisfaction with the product. They often use on-line platform just for tracking prices. The basket leaving does not necessarily mean abandoning the purchase. Some clients are going to decide to buy selected products in the near future, e.g. waiting for right time or maybe better price.

Mentioned researches were mainly related to retail customers. The above cases confirm that the shopping cart analysis provides valuable information that can be used to build clients loyalty and improve platform conversion rate.

In this article we want to present the shopping cart analysis of the customers of the technical articles wholesale. In order to describe searching process and better understand purchase decisions of b2b clients.

II. Methodology and Research Model

A. Research questions

The main research questions are:

- What are the main characteristics of the search phrase?
- How many items are returned in a single search procedure?
- What kind of users buy products using the business-to-business platform?
- Do b2b clients use during search procedure the advanced manual filtering mechanism to complete the purchase?
- What is the average size of the list returned by the query? This may describe both the ability to find the right product by the customer and product range offered by the platform.
- What is the optimal size of results list allowing to finalize the purchase? It seems that too small or big size of the results list is not conducive to adding items to the shopping cart.
- How many shopping baskets were filled by single user during the examined period?

B. Explored data

The Authors' explored log files generated by users of the online technical articles wholesale. The use case diagram of explored b2b platform is presented on Fig. 1

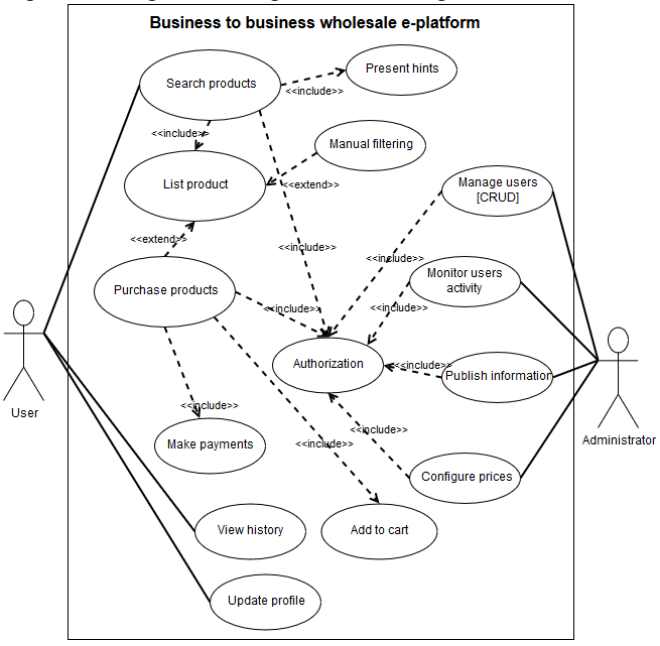


Fig. 1 The use case diagram of explored b2b platform

The explored b2b platform is technical articles wholesale. It has two groups of actors: business users and administrators. Registered business customers can use the platform to search for the products, view products characteristic, make a purchase, make a payment or view purchase history. Administrators can create and manage users' accounts, monitor users' activity, configure prices or promotions and finally add news.

The sample screens of searching mechanism of investigated b2b platform were shown on Fig. 2 and Fig. 3.

Fig. 2 Search inbox and prompt screen

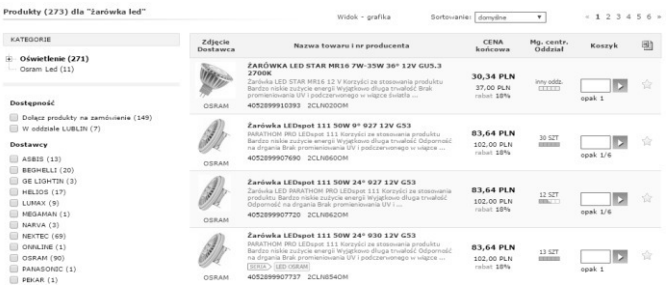


Fig. 3 Product list filters

We explored 596130 logs generated between September 1st 2015 and October 31st 2015. The logs were generated by 4824 b2b customers.

A log file collected by the platform contained information about search phrase, search date and time, customer ID, number of products found (Fig. 4).

```
Phrase;DateAndTime;CustomerID;NoOfRes
22x3/4 mufa ;2015-9-1 06:10:39;0093689001;3
kolano 22x3/4 ;2015-9-1 06:10:56;0093689001;12
mufa 22  ;2015-9-1 06:11:15;0093689001;28
filtr 3/4 płuk;2015-9-1 06:11:35;0093689001;3
forum   ;2015-9-1 06:12:13;0096764001;15
króciec 3/4 ;2015-9-1 06:12:31;0093689001;10
```

Fig. 4 Part of an analyzed log file

C. Research procedure

The research procedure was presented on Fig. 5. We divided the procedure into the following stages: data collection, data selection, data completion, basket linking, quantitative data analysis and conclusions.

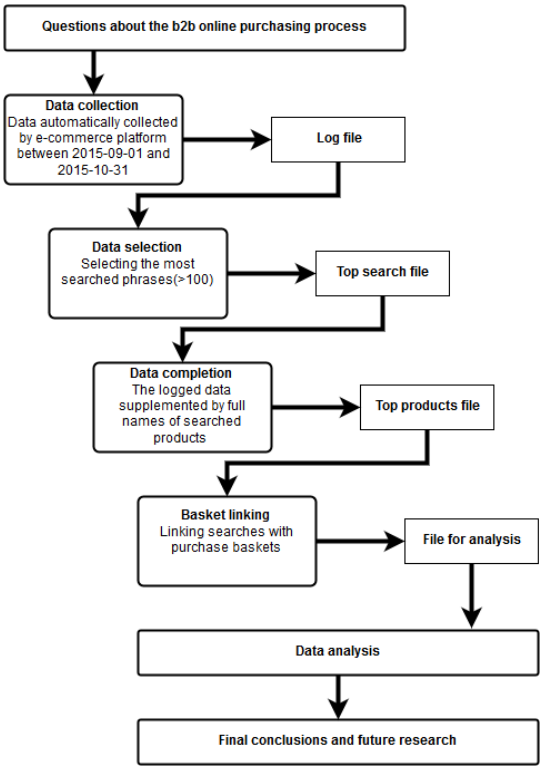


Fig. 5 The research procedure

- The mechanisms implemented on the b2b platform allowed for the collection of transaction logs without the need to turning on the debug mode.
- The automatically generated log file was limited to the most popular phrases. We selected the phrases that were searched more than 100 times in the period of two months. As the result we obtained the list of 367 the most popular phrases (Table 1).
- The list of the most popular search phrases has been supplemented with the list of full products names that were returned by searching mechanism. We used SQL command `[select product from product_list where product_name like "*searched_phrase*"]` This extension was performed only for the 367 most popular search phrases.
- Knowing the indexes of searched products we linked them with shopping basket logs stored in the system. The obtained file became the input for further quantitative analysis.

III. Data Analysis

The qualitative analysis of the incidence of search phrases showed that the most frequently occurring phrases are: pipe, pump, LED, boiler, valve, heater, sink. Mainly pure text phrases (3-11 characters) that included general name of searched object or the manufacturer name. The "word cloud" of the most popular search phrases used by b2b customers was presented on Fig. 6.

Fig. 6 The most frequent search phases (Polish spelling)

The number of products returned by search mechanism for less popular phrases (100-500 searches) is 10-100. For the more frequently used search phrases, the number of returned products is from 100 to 1000. We can say that these are the most popular group of products but also their recognition in the form of phrases, it is very inconvenient for the user. The search phrase that returns 1000 items creates about 20 result pages. Therefore to find the right product user have to use the additional filtering mechanism. In the explored case, clients used additional filtering mechanism which is reflected in a large number of baskets containing these products (Fig. 7).

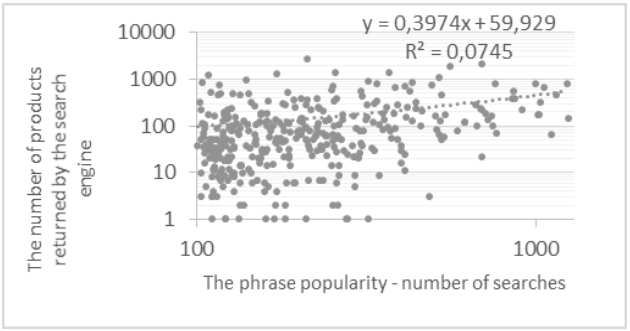


Fig. 7 Phrases popularity vs. the number of products returned by the search engine

In the second study we analyzed how the number of products in search results affected adding them into the shopping baskets. Some search processes returned a lot of products (>1000 items that is >20 pages). That should not foster adding products to the shopping carts. This seems to be problematic for the customer because there are too many items to analyze manually. However, it turned out that these common phrases were source of many purchase decisions. The most shopping carts consisted of items that were chosen from the lists of more or less 250 products (about 5 pages).

From the Fig. 8 we can conclude that the optimum number of products fostering the purchase act is from 30 to 150 items (1 to 3 pages). A larger number of results gives a minimal purchase increase till the point where we have again an increase in the number of items in a baskets. The second area of purchase increase is observed in the range from 600 to 800 returned items. However we assume that it is due to product attributes but not due to the user's convenience. We can say that this kind of inconvenience does not prevent the user from buying. Users seem to continue buying process using manual filters. The question for the future analysis is what kind of additional filtering has been applied to finalize the purchasing process? Currently, the platform does not register information about usage of manual filter.

TABLE I.
THE NUMBER OF PRODUCTS RETURNED FOR A PARTICULAR SEARCH PHRASES

Number of items	Phrase count
1-10	56
11-20	25
21-30	31
31-40	32
41-50	18
51-60	18
61-70	15
71-80	19
81-90	8
91-100	13
101-110	6
111-120	10
121-2720	7 - 1
Total	367 phrases

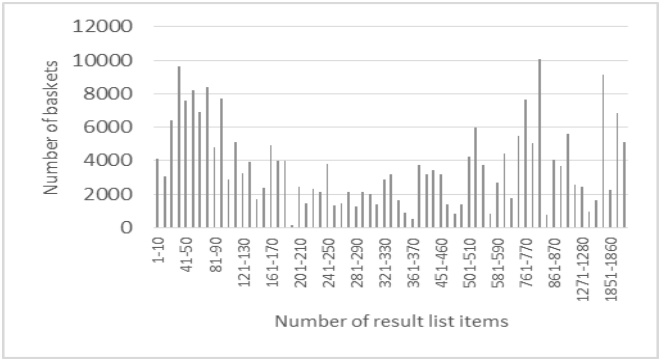


Fig. 8 The most frequent search phrases

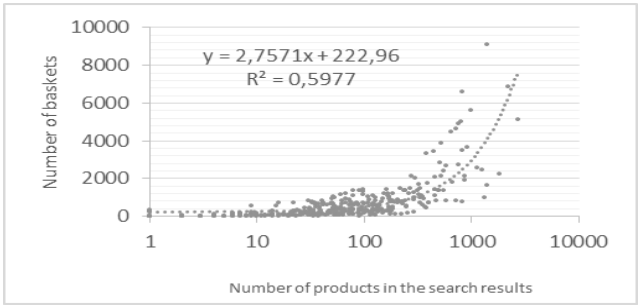


Fig. 9 Search results vs the number of baskets

In the next study we tried to find how the search results are distributed among the population of users. Fig. 10 shows that the majority of users receive less than 200 products as the search results (1-4 pages). That amount of results is possible to read and familiarize with the offer. Larger lists of products was obtained by much smaller group of users.

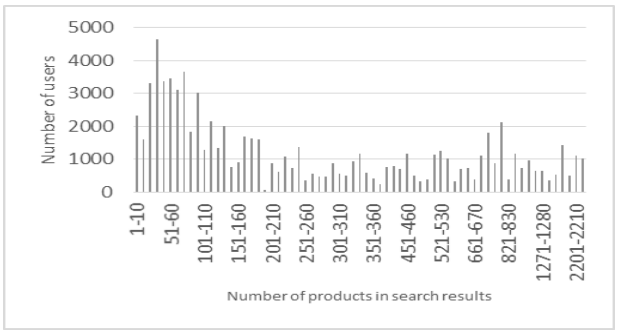


Fig. 10 The number of users obtaining a given number of products as a search results

We also examined how many baskets per customer was created using the same search phrases (Fig 11).

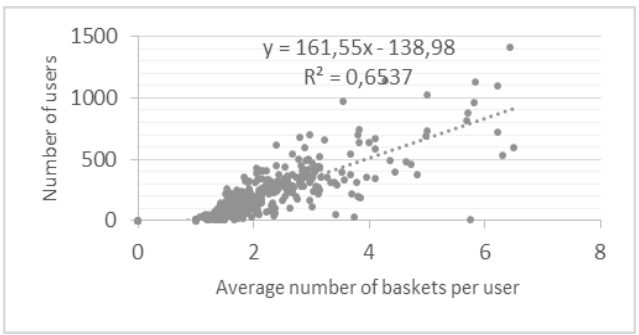


Fig. 11 The average number of baskets per user

As it was shown on Fig. 11, it turns out that during the two-month period it was only from 1 to 4 baskets. As for the b2b shopping it is not a staggering number.

IV. Conclusion

The performed research gave an answer to many questions connected with b2b clients purchasing process. However, were also the source of many new questions. The conducted analysis shows that:

- Transaction logs are source of viable information about customers' behavior and functioning of the b2b platforms. They can be used to investigate customers' purchase decisions and improve the efficiency of b2b tools.
- The examined platform users, specialists in electrics, hydraulics, ventilation used mostly short (from 3 to 11 characters, 1-3 words) text search phrases like pipe, valve, LED, pump, boiler, which as a result gave many pages of items. Thus, the effective platform should include easy-to-use tools that allow for selecting one particular item from an extensive list.
- The performed research was complicated by a set of search phrases resulting in a large number of products found. It turns out that large result lists does not stop customers from finalizing the purchase. However, we weren't able examine how business clients passed from the results list to selecting one specific item. We assume that they used additional manual filtering mechanism rather than reading page after page.
- Our analysis allowed to discover three groups of customers: buyers (few searches and a lot of items in cart), searchers (a lot of searches and few items in cart) and regular ones (few searches and few items in cart).
- The search phrase used by the customer indicates the search engine what subset of offered products should appear on the screen. The user can view presented list and decide whether to make a purchase. The search phrase belongs to customer every-day language and it is not precise. We can say that it is a client's way of naming the class of products. Frequently used short, text phrases gives from 1 to 2720 results. But the most popular phrases returns less than 120 items (3 pages) (Fig. 7, Table 1). In our opinion it is acceptable result for the next manual analysis. The question is the fit and the usefulness of returned results. The conclusion is that most of phrases returns reasonable number of product, so it shouldn't result in inconvenience of shopping and stopping the customer from buying. However we must clearly state the assumption that examined search process was performed with the intention of purchase and there were no obstacles for an effective purchase.
- Fig. 10. shows that users got mostly 1 to 4 results pages. So they were able to read and analyze presented items and finalize the purchasing process. The b2b clients used the same search

phrases only to complete from 1 to 4 shopping baskets.

- The optimal number of products in search results list allowing to finalize the purchase is from 30 to 150 items (1 to 4 pages). However the users with the score of 20 results pages also weren't discouraged from making purchases.
- Search mechanism "clouds of tags" should be reconfigured in order to return less than 200 items (1-4 pages) as a response to the commonly used search phrases.
- If search phrase gives more than 1000 items, client should be supported with additional filtering mechanism or advanced search mechanism allowing for more precise searching.
- The additional extended tests should be performed in order to determine if, in fact, a large number of products in the results discourages customers or makes it difficult to buy the right product (lack of purchases). To perform this research we should use search phrases of similar popularity that gives different number of products found.

V. Future Research

As The Gallup research indicated to build better b2b strategy the companies should combine quantitative (automatically collected) data with a qualitative "small-data" approach [1]. In the future studies we want to obtain interplay between quantitative and qualitative methods so we intend to examine small groups of b2b platform users using in-depth semi-structured interviews and focus groups to measure customers' engagement, and know better client's unique organizational culture. The interesting material for deeper analysis could be obtained also by recording (screen recording) the on-line shopping process [20]. The authors examined single b2b platform that may limit the findings generalizability so in the future research other b2b platforms should be explored to validate the results [2].

References

[1] I. Levey, J. Timmerman, "More Data Doesn't Always Mean Better Insights". *Gallup Business Journal*., p. 1-1, November 2015.

[2] A. Ho, P. Sharma, P. Hosie, "Exploring customers' zone of tolerance for B2B professional service quality", *Journal Of Services Marketing*, vol. 29, no. 5, pp. 380-392, August 2015.

[3] M. Nink, J. Fleming, "Get Better at Knowing Your Client's Decision-Making Process", *Gallup Business Journal*, p. 5, November 2014.

[4] S. U. Stucky, M. Cefkin, Y. Rankin, B. Shaw, "Dynamics of value co-creation in complex IT service engagements", *Information Systems & E-Business Management*, vol. 9, issue 2, pp. 267-281, June 2011.

[5] L. Hsu, K. Wang, W. Chih, "Effects of web site characteristics on customer loyalty in B2B e-commerce: evidence from Taiwan", *Service Industries Journal*, vol. 33, issue 11, pp. 1026-1050, November 2013.

[6] I. Sila, "Factors affecting the adoption of B2B e-commerce technologies", *Electronic Commerce Research*, vol. 13, issue 2, pp. 199-236, May 2013.

[7] L. Taehee, C. Jonghoon, S. Junho, L. Sang-goo, "An Ontology-Based Product Recommender System for B2B Marketplaces", *International Journal Of Electronic Commerce*, vol. 11, issue 2, pp. 125-155, Winter 2006.

[8] J. Bellizzi, "Using Non-Utilitarian Factors to Encourage Business-to-Business Purchases", *Journal Of Global Business Issues*, vol. 3, issue 1, pp. 121-126, Spring 2009.

[9] C. Chlupsa, W. Döhl, J. Lean, Y. Hanoch, "'The impact of implicit motives on the business to business decision making process", *Neuropsychoeconomics Conference Proceedings*, p. 31, 2013.

[10] L. Mingming, M. Parlar, "Free shipping and purchasing decisions in B2B transactions: A game-theoretic analysis", *IIE Transactions*, vol. 37, issue 12, p. 1119-1128, December 2005,

[11] M. Friedman, "Models of Consumer Choice Behavior", in *Handbook of Economic Psychology*, W. F. van Raaij, G. M. van Veldhoven, K.-E. Wärneryd, Ed. Springer, 1988, pp.337-349.

[12] M. Khan, E. Naumann, P. Williams, "Identifying the key drivers of customer satisfaction and repurchase intentions: an empirical investigation of Japanese b2b services", *Journal of Consumer Satisfaction, Dissatisfaction & Complaining Behavior*, vol. 25, pp. 159-178, 2012.

[13] J. Belonax, S. Newell, R. Plank, "The role of purchase importance on buyer perceptions of the trust and expertise components of supplier and salesperson credibility in business-to-business relationships", *Journal of Personal Selling & Sales Management*, vol. 27, issue 3, pp. 247-258, Summer 2016.

[14] C. R. Cano, J. S. Boles, C. J. Bean, "Communication media preferences in business-to-business transactions: an examination of the purchase process", *Journal of Personal Selling & Sales Management*, vol. 25, issue 3, Summer2005.

[15] G. Mallapragada, S. Chandukala, S, L. Qing, "Exploring the Effects of "What" (Product) and "Where" (Website) Characteristics on Online Shopping Behavior", *Journal of Marketing*, vol. 80, issue 2, pp. 21-38, March 2016.

[16] B. Nichols, D. Raska, D. Flint, "Effects of consumer embarrassment on shopping basket size and value: A study of the millennial consumer", Journal of Consumer Behaviour, vol. 14, issue 1, pp. 41-56, January 2015.

[17] S, Ramanathan, Dhar, "The Effect of Sales Promotions on the Size and Composition of the Shopping Basket: Regulatory Compatibility from Framing and Temporal Restrictions", *Journal of Marketing Research (JMR)*, vol. 47, issue 3, pp. 542-552, June 2010.

[18] S. Ramanthan, S. Dhar, "Buy One, Get One Free: How Framing Sales Promotions Affects the Whole Shopping Basket", *Gfk-Marketing Intelligence Review*, vol. 5, issue 1, pp. 49-52, May 2013.

[19] P. Manchanda, A. Ansari, S. Gupta, "The "Shopping Basket": A Model for Multicategory Purchase Incidence Decision", *Marketing Science*, vol. 18, issue 2, p. 95, 1999.

[20] Z. Anesbury, M. Nenycz-Thiel, J. Dawes, J, R. Kennedy, "How do shoppers behave online? An observational study of online grocery shopping", *Journal of Consumer Behaviour*, vol. 15, issue 3, pp. 261-270, May/June 2016.

[21] K. van Ittersum, B. Wansink, J. Pennings, D. Sheehan, "Smart Shopping Carts: How Real-Time Feedback Influences Spending', *Journal Of Marketing"*, vol. 77, issue 6, pp. 21-36, November 2013.

[22] M. Kukar-Kinney, A. Close, "The determinants of consumers' online shopping cart abandonment", *Journal Of The Academy Of Marketing Science*, vol. 38, issue 2, pp. 240-250, Spring 2010.

[23] J. Coppola, K. Sousa, "Characteristics affecting the abandonment of e-commerce shopping carts - a pilot study", *Proceedings For The Northeast Region Decision Sciences Institute (NEDSI)*, pp. 384-389, 2008.

[24] C. Chang-Hoan, K. Jaewon, H. Cheon, "Online Shopping Hesitation", *Cyberpsychology & Behavior*, vol. 9, issue 3, pp. 261-274, June 2006.

[25] M. Hernes, M. Maleszka, N. Thanh Nguyen, A. Bytniewski, "The automatic summarization of text documents in the Cognitive Integrated Management Information System", Proceedings of the 2015 Federated Conference on Computer Science and Information Systems, pp. 1387-1396, 2015.

[26] M. Skuza, A. Romanowski, "Sentiment Analysis of Twitter Data within Big Data Distributed Environment for Stock Prediction", Proceedings of the 2015 Federated Conference on Computer Science and Information Systems, pp. 1349 – 1354, 2015.

[27] G. Suchacka, „Analysis of Aggregated Bot and Human Traffic on E-Commerce Site", Proceedings of the 2014 Federated Conference on Computer Science and Information Systems, pp. 1123 – 1130, 2014.

[28] N. Matta, H. Atifi, F. Rauscher, „Knowledge extraction from professional e-mails", Proceedings of the 2014 Federated Conference on Computer Science and Information Systems, pp. 1407 – 1414, 2014.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1397–1401
DOI: 10.15439/2016F269
ACSIS, Vol. 8. ISSN 2300-5963

Evaluating Business Success Through Social Media Strategies Using AHP

Mervegül Toğlukdemir, Elif Tuygan, Hasan Efe Yeşil
Management of Technology,
Istanbul Technical University
Maçka Istanbul 34367, Turkey
Email: {toglukdemirm, tuygan15, yesil15}@itu.edu.tr

Gülgün Kayakutlu
Industrial Engineering Department,
Istanbul Technical University
Maçka Istanbul 34367, Turkey
Email: kayakutlu@itu.edu.tr

***Abstract*—Social Media has become indispensable for market penetration. It is a beneficial communication platform for understanding the customer focus. Effective use of this media for image creation, customer access, knowledge accumulation and trend analysis, creates competitive advantages. This research is designed to analyze social media strategies of global enterprises and evaluate the value of social media usage. Analytical Hierarchy Process (AHP) is used to model the performance decisions. The model is constructed based on the major evaluation criteria used by the global companies. AHP model expresses cause and effect relationships between companies and social media effects. Cases will be applied for Coca Cola, Turkish Airlines and Starbucks. Enterprise awareness and success of different evaluation criteria are benchmarked.**

***Index Terms*—Analytic Hierarchy Process, Social Networks, Strategic Planning and Management**

I. Introduction

THE Internet has been evolved from a basic tool of communications into an interactive market of products, services to global community and worldwide business transactions. Many enterprises attempt to embrace the digital revolution and using internet is now a necessity for the enterprises. As a result of the increased use of the Internet, using social media also became competitive advantage for the growing companies, they attempt to differentiate their products and services from competitors.

The past decade development and change of the usage of the Internet, it is difficult to differentiate in the global community only with one-way communications from seller to buyer thus, companies aim to interact more with their customers as well as to allow their customers to interact more with them (Winer, 2009). Since social media content has become indispensable to millions of users, it becomes wide source of big data and allows companies to reach more people at a lower costs. Social media is a great opportunity to communicate with customers for the companies, it provides to collect a big data of the users through their expectations, experiences, ideas and reactions for the product. Another benefit of social media is the quick interaction. The ability to receive quick reaction, in terms of ensuring a positive brand image is of great importance. It may take a second customers to respond a message posted on social platforms thus it helps companies to measure the pulse of the customers' ideas based on their positive or negative comments.

The objective of this study is to analyze the value of social media and social web data for businesses to differentiate in the global competition, and propose a theoretical explanation by analyzing social media strategies of the Turkish Airlines, Coca Cola and Starbucks which are global companies. This study involves main titles; firstly, importance of the social media analyzes for the enterprises' success across the globe will be studied. Secondly, Analytic Hierarchy Process method will be explained in steps, which contains explanations and the description of formulations. Thirdly, to analyze companies' awareness on social media, the main goal, criteria and alternatives for the AHP method will be obtained and explained in detail. Then, the sample using AHP method from the literature will be examined. Finally, results will be evaluated and by using AHP method an expressive cause and effect relationships between companies and social media effects are established.

II. Literature Review: Knowledge Management In Social Media Strategies

The widespread use of the social media in recent years it has become an important marketing strategy for the companies to meet directly with the customer in easiest, cheapest and fastest way. Social media helps reshape business models through opinions and emotions owing to fact that it opens up many possibilities to study human interaction and collective behavior. Many companies today are using social media to develop targeted campaigns that reach specific segments and engage their customers.

The past decade development and change of the Internet, numerous social networking sites have been drawing people together and creating new forms of communication. There are various social media forms; people write encyclopedia articles, online marketplaces recommend products via user shopping interactions; and community movements benefit from new forms of collective actions (Tang, Liu, 2010).

According to Feng and Qjan, "Facebook has about one billion users and there are about 3 million photos uploaded by users each day" (2013). LinkedIn is another example, the user can create a professional profile, establish connections to other users, and exchange messages, follow the current news. Tanbeer, Leung and Cameron stated that "social entity is connected to another entity as his or her next-of-kin, friend, collaborator, co-author, classmate, co- worker, team member, and/or business partner" (2014).

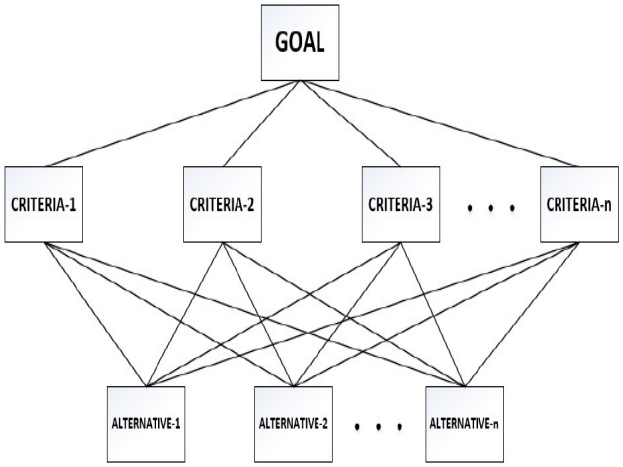


Fig 1. Hierarchical Structure of AHP

Mentioned connection might be used as a power to reach more customer. Thus the social media Websites are the ideal platforms to facilitate the products recommendation and market popularity. Social media has become wide source of big data for the companies. It is significant to analyze the social media entirely for the companies considering their target. Analyzing Social Media includes theories and methodologies from different disciplines such as computer science, data mining, machine learning, social network analysis, sociology, ethnography, statistics, optimization, and mathematics (Zafarani, Abbasi and Huan, 2014). The correct analysis of social media data is considered as the very important step of a successful marketing strategy for the growing companies.

III. Methodology Implemented: Analytic Hierarchy Process (AHP)

People make their decisions in two ways; the first is heuristic approach that develops very fast and is usually not objective. The second is the logical analysis that requires an analytical method. Analytic Hierarchy Process is a quantitative decision-making method according to multiple criteria and alternatives based on mathematics and psychology. Saaty stated that "a hierarchy is a representation of complex problem in a multilevel structure whose first level is the goal followed successively by levels of factors, criteria, sub-criteria, and so on down to bottom level of alternatives" (2006).

AHP lead the way that yield best in order to reach the goal/solve the decision problem instead of stating correct decision. The decision problem dissociated into hierarchy that contain criteria and alternatives. The hierarchy shows cause-effect relations in linear chain. After identifying the problem or stating the goal, hierarchy is created considering main decision point, middle level criteria and the lowest level of the possible alternatives (Figure 1).

Once hierarchy is built, the decision maker evaluate its elements (criteria and alternatives) by comparing their impact/effect on an element in the hierarchy between each other. In comparison, concrete data, judgments, expert opinions or survey results etc. are used as resource by decision maker in order to determine the degree of importance between elements. The importance scale table is used for comparison (Table 1).

Table I.
Importance Of Scale AHP

Definition	Degree of importance
Equally important	1
Moderately important	3
Strongly important	5
Very strongly important	7
Extremely important	9
Intermediate values between two adjacent judgments	2, 4, 6, 8

The importance or numerical weights (according to degree of importance) are obtained for each criteria and alternative in the hierarchy so that benchmarking is made in a consistent and rational way. Finally, according degree of importance numerical importance are calculated for each decision alternatives. This evaluation shows ability of alternatives to achieve the goal (attracting customer through social media).

In this study, computer software "Super Decisions" is used for calculations so mathematical formulas not stated in detail. Steps are required to be resolved in a decision-making problem with AHP are summarized as follows:

Step 1. Model the problem as a hierarchy containing the decision goal, the alternatives for reaching it, and the criteria for evaluating the alternatives.

Step 2. Establish priorities among the elements of the hierarchy by making a series of judgments based on pairwise comparisons of the elements. For example, when comparing potential purchases of commercial real estate, the investors might say that location is five times important than price and price is three times important than timing.

Step 3. Synthesize these judgments to yield a set of overall importance for the hierarchy. This would combine the investors' judgments about location, price and timing for properties A, B, C, and D into overall importance for each property.

Step 4. Check the consistency of the judgments.

Step 5. Come to a final decision based on the results of this process. (Saaty, 2008)

IV. Creating the Model

A. Determining the Goal of AHP Model

In this study, "attracting customers through social media" is chosen as main goal for the hierarchy tree.

B. Determining the Criteria of AHP Model

The criteria will be studied on three of the most common social media tools which provides companies have the opportunity to reach very large data, are chosen; LinkedIn as

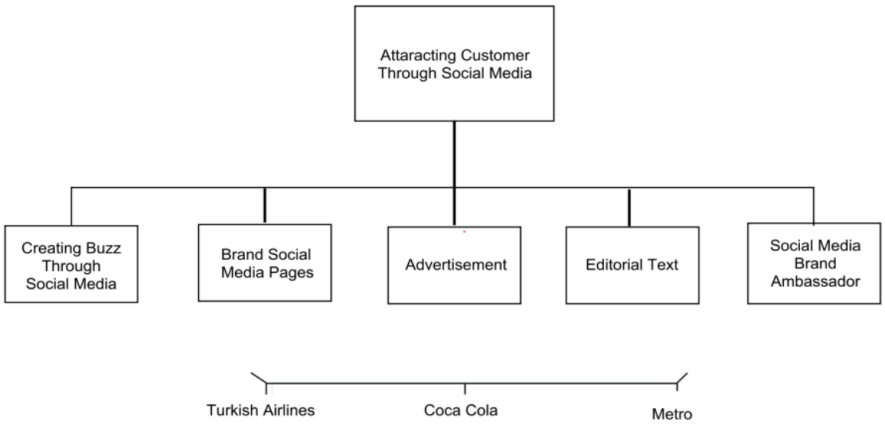


Fig 2. AHP decision tree

professional social media tool since people share their business information, resume and career histories in, Twitter as entertainment and news social media tool since people share their thoughts worldwide and Facebook as associational social media tool since it has become one of the most widely used tools for advertisement and reaching users (Drury, 2008). Determined criteria are as follows:

Brand Social Media Pages: Through social networking profiles of a brand it is possible to analyze demographics of the customers/followers, their interests, location information in order to devise new strategies.

Social Media Advertising: The most commonly adopted approach to the social media advertisement is to advertise on the social networks with the consultation of businesses, which have proved themselves with their experience on social media.

Editorial Text on Social Media: This kind of a communication will not force the customer to buy goods. It is built on arousing curiosity, giving information and offering benefits to the customer.

Social Media Brand Ambassador: Companies find the social network celebrity that suits the brand. It is of course a must for this person to have enough followers on certain networks too.

Creating Buzz Through Social Media: According to this marketing trend, volunteers found by a company convey their experience with a specific product to the people who they encounter with at any time.

C. Determining the Alternatives of AHP Model

The alternatives: three global companies are determined as follows:

Turkish Airlines: Recognized as a global brand Turkish Airlines (Turk Hava Yolları / THY) took 3rd place as one of the most active brands in social media. THY actively use social media, broadcasts in Turkish and English languages. In addition to Company has 8 million over Facebook fans and has over 1 million followers on Twitter. The profile of the brand via Facebook Turkish Airlines Euro league, Barcelona and Manchester United sponsorship or related information, campaigns and current developments can be followed.

Coca Cola: Coca Cola is actively using the Facebook and Twitter both products and campaigns. According to Social Media Principles of the social media strategy of the company is that more than 150,000 associates in more than 200 countries to join conversations take place online about Coca-Cola, represent the company, and share the optimistic and positive spirits of the brand (Coca Cola Company, 2013). Coca-Cola among the best in the world with more than 97 million members on Facebook and it has more than 170 thousand fans from Turkey. A. On Twitter also it has 93 thousand followers.

Starbucks: Starbucks has 640 thousand over Facebook fans and has over 74 thousand followers on Twitter.

After decision of the main goal, criteria, and alternatives; Hierarchical Structure of AHP is created. (Figure 2)

V. Application

A. Applying the Model on a Software

AHP/ANP software that is called Super Decisions is used in the scope of this study. Clusters and nodes are created according to Figure 2 and illustrated in Figure 3.

According to importance level that are shown in Table 1. the nodes of criteria cluster; the nodes of criteria cluster and the nodes of alternative cluster; the nodes of alternative cluster and nodes of criteria cluster are compared two by two (several importance table are shown in Appendix 1).

Companies are utilized according to usage of such social media tools:

- Professional social media tools (LinkedIn),
- Entertainment, new social media tool (Twitter),
- Associational social media tool (Facebook)

Some examples;

With respect to Coca-Cola, creating buzz trough social media is moderately more important than brand social me-

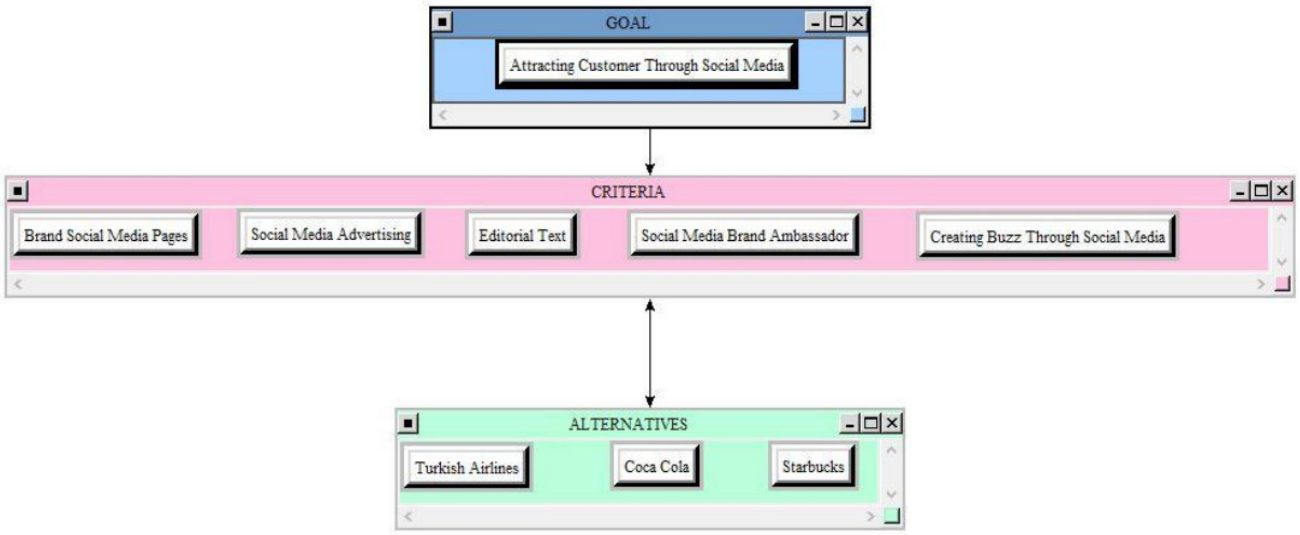


Fig 3. AHP decision tree

dia; brand social media is strongly more important than editorial text and so on.

B. *Results and Discussions*

According to our observations and reviews with alternative companies in Turkey, the importance of the each social media tools is different for each company (approximate values are shown in Table 2).

Software reports are shown in Table 2 and Table 3, accordingly.

As it seems in Table 3, in the scope of social media (LinkedIn, Facebook, Twitter), in order to attract the customer Coca Cola is using social media tool more effective than Turkish Airlines; Turkish Airlines is using social media tool more effective than Starbucks.

VI. Conclusion And Recommendations

Social media is important trend, the company has become an important marketing opportunity to meet directly with the customer. Social media allows businesses to communicate

Table II
Main structure in Super Decision report

Alternative(s) in it	• Coca-Cola • Starbucks • Turkish Airlines
Network type:	Bottom level
Formula:	`Not Applicable`
Clusters/Nodes	• **Alternatives:** ◦ **Coca-Cola:** the importance of LinkedIn %30, Facebook %40, Twitter %40 ◦ **Starbucks:** the importance of LinkedIn %25, Facebook %40, Twitter %35 ◦ **Turkish Airlines**: the importance of LinkedIn %25, Facebook %35, Twitter %40 • **Criteria:** ◦ **Brand Social Media Pages:** social networking profiles ◦ **Creating Buzz Through Social Media:** volunteers found by a company nconvey their experience with a cpecific products ◦ **Editorial Text:** giving information and offering benefits to the customer ◦ **Social Media Advertising:** advertise on the social networks with the consultation of businesses ◦ **Social Media Brend Ambassador:** social network celebrity that suits the brand • **Goal:** ACTSM as Business Success Strategy ◦ **Attracting Customer Through Social Media:** drawing customer attention via social media

TABLE III
ALTERNATIVE RANKING IN SUPER DECISIONS REPORT

Graphic	Alternatives	Total	Normal	Ideal	Ranking
	Coca Cola	0.2170	0.4340	1.0000	1
	Starbucks	0.0781	0.1562	0.3600	3
	Turkish Airlines	0.2049	0.4097	0.9440	2

with the costumers in lower costs and greater efficiency than traditional media tools (Kaplan and Haenline, 2010).Recently one of the most important issues of the businesses involved in social media is to measure the success of social media and events. Therefore, to measure the success of social media marketing, and it has become important to determine the measurement criteria to guide the social media marketing, according to information obtained.

The aim of this study was to analyze social media strategies of the global enterprises and raise enterprises awareness by understanding the value of using social media tools in order to differentiate in the global competition. For this purpose, some global company examples are named which are Turkish Airlines, Coca Cola and Starbucks then social media main criteria are determined and show how the methodology of analytic hierarchy process is executed through main criteria that enterprises are able to use.

As a result, in the scope of social media (LinkedIn, Facebook, Twitter), in order to attract the customer, Coca Cola is using social media tool more effective than Turkish Airlines; Turkish Airlines is using social media tool more effective than Starbucks.

REFERENCES

[1] Barutcu S., Tomas M., (2013) Sürdürülebilir Sosyal Medya Pazarlaması ve Sosyal Medya Pazarlaması Etkinliğinin Ölçümü. IUYD 2013, 4(1), p.6-23.

[2] Coca Cola Company, (2013) , Social Media Principles, retrieved from https://www.coca-colacompany.com , Access:04.03.2016

[3] Culnan M. J., McHugh P. J., Zubillaga J.I. (2010). How Large U.S. Companies Can Use Twitter and Other Social Media to Gain Business Value. MIS Quarterly Executive 9(4).

[4] DeLone W. H., McLean E.R. (2014). Measuring e-Commerce Success: Applying the DeLone & McLeanInformation Systems Success Model. International Journal of Electronic Commerce, 9(1), pp.31-47.

[5] Dhir S., Marinovb M. V., Worsleyb D., (2015). Application of the analytic hierarchy process to identify the most suitable manufacturer of rail vehicles for High Speed 2. Case Studies on Transport Policy, (3), p.431–448

[6] Drury G. (2008). Opinion piece: Social media: Should marketers engage and how can it be done effectively?. Journal of Direct, Data and Digital Marketing Practice, (9), p.274–277.

[7] EticaretMag, E-ticaret Şirketleri İçin Sosyal Medya Kullanım Rehberleri, retrieved from http://eticaretmag.com/e-ticaret-sirketleri-icin-sosyal-medya-kullanim-rehberleri/ , Access: 04.03.2016

[8] Feng H., Qjian X. (2013). Mining user-contributed photos for personalized product recommendation. Neurocomputing, (129), pp.409–420. Department of Information and Communication Engineering, Xi'an Jiaotong University, Xi'an 710049, China.

[9] Güner, H., (2005), Bulanık AHP ve bir İşletme İçin Tedarikçi Seçimi Problemine Uygulanması, Pamukkale University Institute of Science Industrial Engineering Department, p.133

[10] Ho W., (2008), Integrated Analytic Hierarchy Process and its Applications-A literature Review, European Journal of Operational Research, (186), p.211-228

[11] Kahya E., (2016). Analyzing unstructured Facebook social network data through web text mining: A study of online shopping firms in Turkey. Information Development, 32(1), pp.70-80. doi: 10.1177/0266666914528523

[12] Kaplan A.M. ve Hainlein M. (2010). Users of the World, The challanges and Opportunities of Social Media, Business Horizons, 53, 59-68

[13] Kong F., Liu H. (2005). Applying Fuzzy Analytic Hierarchy Process to Evaluate Success Factors of E-Commerce. International Journal of Information and Systems Sciences, 1(3-4), pp.406-412.

[14] Linda S., (2010). Social Commerce – E-Commerce in social Media Context. World Academy of Science, Engineering and Technology International Journal of Social, Behavioral, Educational, Economic, Business and Industrial Engineering, 4(12).

[15] Saaty T. L., (2006). Fundamentals of Decision Making and Priority Theory. RWS Publications 4922 Ellsworth Avenue, Pittsburg, PA 15213, p. 94

[16] Saaty, T. L., Niemira, M.P., (2006), A framework for making a better decision, Research Review, p. 13

[17] Saaty, T.L (2008). Decision Making for Leaders: The Analytic Hierarchy Process for Decisions in Complex World. Pittsburg, Pennsylvania: RWS publications. 8

[18] Supciller A. A., Capraz O., (2011), AHP-TOPSIS Yöntemine Dayalı Tedarikçi Seçimi Uygulaması. Ekonometri ve İstatistik, (13), p.1–22

[19] Tanbeer S.K., Leung C.K., Cameron J.J. (2014). Interactive Mining Of Strong Friends From Social Networks And Its Applications In E-Commerce. Journal of Organizational Computing and Electronic Commerce, (24), pp.157–173, doi:10.1080/10919392.2014.896715

[20] Tang L., Liu H., (2010). Community Detection and Mining in Social Media. Synthesis Lectures on Data Mining and Knowledge Discovery. Morgan & Claypool Publishers, doi:10.2200/S00298ED1V01Y201009DMK003

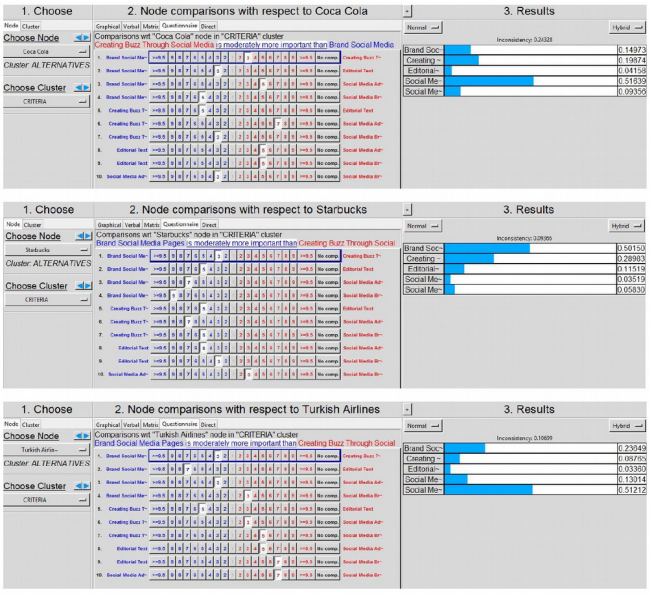
[21] Winer R.S. (2009). New Communications Approaches in Marketing: Issues and Research Directions. Journel of İnteractive Marketing, 23, pp.108-117. Marketing Szience Institute, Stern School of Business, New York University, USA.

[22] Zafarani R., Abbasi M.A., Huan L. (2014). Social Media Minging. Cambridge University Press, April 20, 2014. draft version:

VII. APPENDIX

Some of importance comparison table are as follows.

With respect to Coca-Cola, creating buzz trough social media is moderately more important than brand social media.



2nd International Workshop on Ubiquitous Home Healthcare

POPULATION aging is a phenomenon affecting many countries around the world. For example in Europe the life expectancy increased from 45 years in the early twentieth century, to 80 years now. Significantly longer life leads to age-related problems and diseases. In parallel, the cost of hospital care is increasing and additionally, a lack of the qualified caregivers is observed. Development of ubiquitous healthcare technologies can improve the quality of life of assisted citizens and can curtail growth in healthcare spending fueled by aging populations, and the prevalence of obesity, diabetes, cancer and chronic heart and lung diseases. In particular, information systems integrated with wearable, mobile devices and sensor networks at home can continuously assist persons while moving out or staying at home. Ubiquitous healthcare systems used as assisted living solutions will not only help to prevent, detect and monitor health conditions of a person but will also support of elderly, sick and disabled people in their independent living.

The goal of the UHH 2016 workshop is to gather researchers and engineers working in the field of ubiquitous healthcare to present and discuss new ideas, methods, and applications of assisted living IT technologies.

TOPICS

The workshop welcomes all work related to ubiquitous healthcare, but with a focus on the following themes (this list is not exhaustive):

- Ubiquitous healthcare information systems,
- Information processing algorithms for UHH,
- Ubiquitous services for home and mobile applications,
- Human-system interaction in UHH,
- Wearable sensors and systems,
- Smart glasses and smart watches in UHH,
- Data mining and knowledge discovery in ubiquitous healthcare,
- Integration of sensors and devices for UHH,
- Security of ubiquitous healthcare systems,
- Ensuring the Availability, Transparency, Seamlessness, Awareness, and Trustworthiness (A.T.S.A.T.) of home and mobile systems,
- Standardization in ubiquitous home healthcare,
- Applications of UHH for elderly, sick, and disabled people,
- Elderly care monitored dosage systems,
- Welfare technology for UHH.

The proposed papers should emphasize at least one of the following aspects:

- Assisted living,
- Home care,
- Self care,
- Mobile care.

BEST PAPER AWARD

During the closing ceremony the best paper award will be presented. Selection criteria will be based on results of the reviewing process and the quality of the presentation.

EVENT CHAIRS

- **Biallas, Martin,** iHomeLab, Hochschule Luzern, Switzerland
- **Wtorek, Jerzy,** Gdańsk University of Technology, Poland

PROGRAM COMMITTEE

- **Andrushevich, Alexey,** iHomeLab
- **Augustyniak, Piotr,** AGH University of Science and Technology
- **Bujnowski, Adam,** Gdansk University of Technology, Poland
- **Cavallo, Filippo,** Filippo Cavallo, The BioRobotics Institute, Scuola Superiore Sant'Anna
- **Haller, Michael,** University of Applied Sciences Upper Austria
- **Kaczmarek, Mariusz,** Gdansk University of Technology, Poland
- **Kistler, Rolf,** iHomeLab, Switzerland
- **Louventain, Nicolas,** SnT, University of Luxembourg
- **Martin, Benoit,** Université de Lorraine, France
- **McCall, Roderick,** Luxembourg Institute of Science and Technology
- **Pecci, Isabelle,** Université de Lorraine
- **Polinski, Artur,** Gdansk University of Technology, Poland
- **Popletev, Andrei,** SnT, University of Luxembourg
- **Rosell, Javier,** Universitat Politecnica de Catalunya
- **Ruminski, Jacek,** Gdansk University of Technology
- **Sincak, Peter,** Technical University of Kosice
- **Strumiłło, Paweł,** Lodz University of Technology
- **Svitek, Miroslav,** Czech Technical University in Prague
- **Tkacz, Ewaryst,** Silesian University of Technology, Poland
- **Truyen, Bart,** Vrije Univ. Brussels
- **Vogl, Anita,** MIL, University of Applied Sciences Upper Austria

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1405–1408
DOI: 10.15439/2016F495
ACSIS, Vol. 8. ISSN 2300-5963

Smart Glasses: A semantic fisheye view on tiled user interfaces

Ilyasse Belkacem, Isabelle Pecci, Benoît Martin
LCOMS, Université de Lorraine
Ile du Saulcy, CS 50128
57045 Metz Cedex 01, France
Email: {ilyasse.belkacem, isabelle.pecci, benoit.martin}
@univ-lorraine..fr}

Abstract— **With the evolution of mobile technology, many devices are introduced with very limited screen sizes like smart glasses. This technology must be accompanied with new visualization techniques. A classic interface can't meet the expectations of the user who becomes increasingly hard to please. The challenge is to display information and allow the user a better navigation with less effort especially in situation of mobility. This paper explores a fisheye view on tiled user interfaces for smart glasses that uses the semantic relationships of items of information contained in the tiles. We propose a reformulation of the degree of interest function and a semantic model for tiled interfaces that supports this reformulation. We developed a prototype to demonstrate the feasibility of our approach and to improve our design approach in our future work.**

keywords— *smartglasses, information visualization, semantic fisheye view, degree of interest*

I. Introduction

SMART glasses are glasses with a wearable technology, including advanced electronic and IT components (embedded processor, display screen, sensors, camera ..). They allow the visualization and interaction with a large information space. Ergonomics, user-friendly interface and comfortable viewing should be made on these glasses to minimize the time and effort provided by the user during navigation. As smart glasses offer many services, it is necessary to find a display mode that convinces the user by giving him easy access to any service with more information.

The concept of tiles [1] allows the creation of a customized user interface with high flexibility [2]. In use, if we need to make changes to configure the tiles for the user, adjusting the position and size will be easily implemented according to our need. The tiles can contain the services provided by smart glasses, a service can be graphics, text or other visual data. Also, the user can point to a tile and make it active. In addition, the tiles can provide users with fast and direct access to launch the services available through the mobile device [1][2].

We consider that the concept of tiles is suitable for smart glasses. However, a large number of tiles (i.e. services) cause some problems. First of all, it is very difficult to represent all that information on very limited space like the screen of smart glasses and navigate easily. In addition, we cannot link the services together easily if we cannot see all the tiles in the same view. It is also very important to have the global view of services and bind them.

We can overcome these problems with the different techniques that address the access to large data spaces on a limited display area [3] but efficient solution is needed.

In this paper, we present a new concept for a user interface on smart glasses based on the concept of tiles and visualization techniques from a wide data space on a small screen. The rest of the paper is organized as follows. We present in the next section (section 2) the fisheye technique and a related works to our work.

In Section 3 we present our approach and applying it after to a prototype in section 4. We expose after perspectives and our future work that we consider interesting in section 5 and we will finish with section 6 that will conclude our work.

II. State of the art

A. Presentation techniques for large data spaces

There are different techniques of presentation and visualization for large data spaces. They can be categorized into two main techniques: distortion-oriented and non-distortion-oriented [3].

The technique of non-distortion can display some information with scrolling or paging to access to the rest of the information. This technique is less effective if the data space is very large, because the user can be lost during navigation.

On the other hand, distortion technique allows the user to simultaneously visualize a local part with a high level of detail, and to have a vision of the global context with less detail on the same screen. In this technique, there are different types of deformations available that differ in their transformation function [3]: polyfocal display, bifocal display, fisheye view and perspective wall.

Fisheye view technique is the technique adopted by our approach. We will focus on this technique in the following section. The strength of this technique is based on the degree of interest (DOI) that measures the interest of each element of information to present to the user. It enables to present the relevant information in detail and irrelevant information with less detail.

B. Fisheye view

The fisheye view is initially proposed by Furnas 1986 [4]. Like other distortions techniques mentioned in the previous section this technique can give a detailed view while keeping the global context. This technique is naturally useful because it is based on the importance and relevance of the information presented to the user.

This technique is inspired by the fisheye lens camera that magnifies near objects and reduces distant objects for a local view in detail and a view of the global context.

1. DOI function:

The fisheye technique uses the degree of interest (DOI) function given by:

$$DOI_{fisheye}(i|.=f_u) = API(i) - D(i,f_u) \quad (1)$$

where $DOI_{fisheye}$ is the degree of interest for a user to an element i given that the current focus element is f_u.

The degree of interest function assigns to each item i of the information a value composed of *A Priori Importance* $API(i)$ of the item i and the distance $D(i,f_u)$ between the item i and the current focus item f_u. The importance $API(i)$ is static and doesn't depend on the current focus item.

The value of the degree of interest increases with the importance and decreases with distance. We can set a threshold k and display only the information items that have a higher degree of interest than the threshold $DOI_{fisheye}(i|.=f_u) \geq k$.

This technique was extended by Sarkar and Brown 1992 [5] defining the mathematical formulas for graphics applications. They proposed four functions: the position, size, the amount of detail to display and visual worth of each vertex of a graph.

2. Emphases algorithm:

The degree of interest function of Furnas quantifies just the importance of each item of information but without detailing how to present it. It is important for the user to distinguish between the different degrees of interest visually. Emphases algorithm makes the mapping between the degree of interest attributed to an item of information and its encoding in graphics visualization variables such as size, color...

Many techniques exist in the literature. Noik [6] has distinguished 4 types:

- Implicit: an order in the placement of the items, for example. It is generally static.
- Filtered: the items that have a lower degree of interest than the threshold will be filtered.
- Distorted: this technique deforms the size, shape, and position.
- Adorned: emphasizes an item of information using other graphic variables such as color, thickness …

Many applications used the fisheye view was developed but few of these that integrates semantics. Research work related to our problem is the semantic zoom introduced in several applications such as Pad ++ [11]. It displays more details on an item of information according to the meaning.

The combination of semantic zoom and fisheye technique was introduced in some research work. Zizi and Beaudouin-Lafon (1995) [12] used the web of documents and information retrieval techniques to provide an interactive map. Van Ham and Van Wijk (2004) [13] used a clustering algorithm to represent semantical distortions for interactive visualization of small world graphs. Janecek & Pu (2002, 2005) [14] [15] developed a framework for a flight itinerary using relationships between itineraries. Our research has a similar goal of using semantic in a fisheye view with a different approach that we present in the next section for tiled user interface on a promising area application: smart glasses.

III. PROPOSED APPROACH

We propose in this section a reformulation of the degree of interest function of the fisheye proposed by Furnas[4] and transform the a priori importance $API\ (i)$ to that takes into account the current point of focus. We also propose an approach that uses this function to compute the degree of interest based on a rich semantic model that considers several concepts.

A. DOI function

The new proposed formula of the degree of interest is given by:

$$DOI_{fisheye}(i|.=f_u) = API(i,f_u) - D(i,f_u) \quad (2)$$

where $DOI_{fisheye}$ is the degree of interest for a user to an item i given that the current focus item is f_u.

Unlike the Furnas formula cited above, the API of an information item is not static. It depends of the current user item focus. In other words, the importance of an item of information i will be dynamic over the item of current focus f_u. It does not have the same importance if the user point out two different items.

The main purpose of this reformulation is to provide useful additional information under certain user focus conditions. Thus, the user could discover new items which are strongly linked to the current focus item. This formula can also increase the usability; the user will discover surprising and interesting information which will be useful for him.

We expect that users using a fisheye interface based on this formula can achieve their tasks faster than a using typical fisheye based on previous formula of Furnas.

The DOI function is typically used for hierarchical data structures, structured text, calendars. We want to apply this function to a tiled interface for our smart glasses. Through our approach, we want to access to tiles with a semantic view and meaning.

B. Semantic model

Each tile provides a service to the user; our goal is to magnify the tiles services that are semantically related when the user hovers on a given tile.

So we need to make comparisons between different services. For this we need a semantic description based on

the analysis of each service to detect what they have in common in terms of content. We can make this description based on metadata and attributes of these services, but the main challenge is to select attributes that are meaningful to the user.

In figure 1, we propose a semantic model focused on the service that defines the main concepts that allow computing our measure of semantic similarity between the different services.

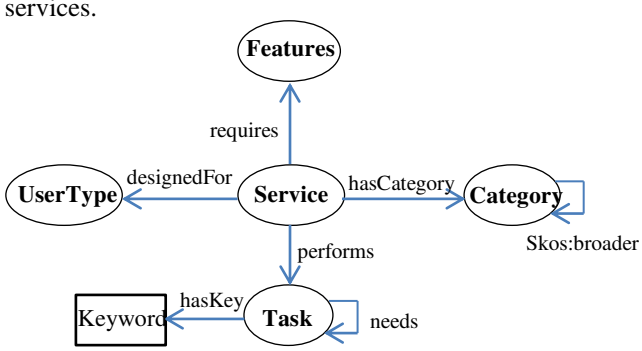


Fig. 1 Semantic service model

Our semantic model uses 4 concepts:

- Category: modeled by a tree structure, each category can have several sub categories (e.g. multimedia category can have two sub categories: audio and video).
- Task: performed by some services. A service can perform one or more tasks, so two services that have several common tasks to execute are probably similar. The goal is not just to identify similar services, but also services that are connected in some way to a service.
- User type: Two services can have the same category and perform the same tasks but are not designed for the same type of users.
- Features: A service may require authentication, internet connection and a camera to operate. Such requirement can give meaning to the proximity between the two services.

After building the model, we must define a similarity measure that quantifies a semantic relatedness with a numeric value between two services from our model. The similarity between two services can be defined by common properties. The services that have more common properties are more similar [7].

Each property has a type; it can be organized hierarchically (e.g. Category) or not organized hierarchically (e.g. Tasks) or a literal (we don't have this type in our model). Different measures of semantic proximities according to property nature are present in the literature [8].

Our calculation has to integrate different types of properties to have a single numeric value that represents the similarity between two services. We describe here the similarity measures that we have chosen for each type:

Properties hierarchically organized:

This type is based on hierarchies of properties. Wu and Palmer measure [9] uses the depth of the two properties and the depth of the least common subsummer (lcl). It's defined as:

$$sim_{Wu,Palmer}(p_1, p_2) = \frac{2\times depth(lcl(p_1,p_2))}{depth(p_1)+depth(p_2)} \quad (3)$$

Properties not hierarchically organized:

The similarity is computed with Jaccard measure [10] which takes into account the number of common properties compared to the total number of properties. It's defined as:

$$sim_{Jaccard}(p_1, p_2) = \frac{\{p_{11}\dots p_{12}\}\cap\{p_{21}\dots p_{22}\}}{\{p_{11}\dots p_{12}\}\cup\{p_{21}\dots p_{22}\}} \quad (4)$$

where $\{p_{11}\dots p_{12}\}$ and $\{p_{21}\dots p_{22}\}$ are respectively the values of the proporities p_1 and p_2.

After computing a similarity value for each property (hasCategoy, needs, performs, requires). We can make a similarity vector based on these similarities. Each component of the vector contains a similarity value of a property. We need to aggregate this vector into a single value using an aggregate function. We choose the weighted average as function on the hypothesis that the semantic concepts don't have necessarily the same importance.

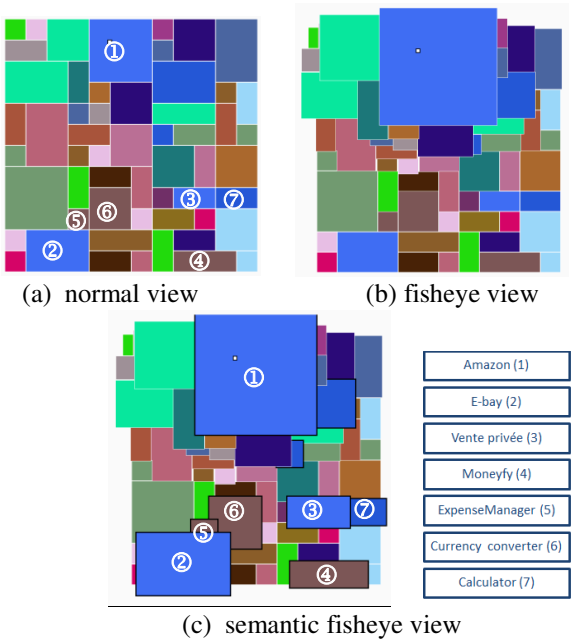
The similarity between two services S_1 and S_2 is given by:

$$Sim(S_1, S_2) = \frac{\sum_k w_k \times Sim_k(p_{S_1}, p_{S_2})}{\sum k} \quad (5)$$

where w_k is the weight of the property p for S_1 and S_2. And the DOI function becomes:

$$DOI_{fisheye}(i|. = f_u) = sim(i, f_u) - D(i, f_u) \quad (6)$$

IV. A PROTOTYPE: PROOF OF CONCEPT



(a) normal view (b) fisheye view

(c) semantic fisheye view

Fig. The proof of concept

Our prototype developed is a tiled interface on android platform for smart glasses. The semantic model is built by Protégé[1] integrating 56 services, each tile in the interface contain a service. We use Jena[2] Framework to be able to compute different similarities from the model. We consider

[1] https://jena.apache.org/
[2] http://protege.stanford.edu/

in our prototype only the category to compute similarities from the model.

Figure 2.a shows the tiled view without any fisheye view. Figure 2.b takes into account only distance to magnify tiles for a fisheye view.

The semantic model is activated in figure 2.c. The associated service to the tile (1) is Amazon. From the computation in our model, it has a similarity of $sim = 1$ with e-bay (2) and Vente-privee (3) because they belong to the same sub-category Shopping. The services: Monefy (4), ExpenseManager(5), CurrencyConverter (6) and the service: Calculator (7) belong respectively to Banking and Shopping categories whose main category is Business which is also the main category of Shopping. So they have a similarity of $sim = 0.5$ with the Amazon service.

We have taken a similarity threshold $k = 0.5$ to emphases a tile. We are aware of the need of building a good emphases algorithm that will be the subject of another work in the future. For this prototype, when the user hovers over a tile, we compute all sizes of tiles to allocate more space for important tiles on relation with the focus item. The size of a tile in the fisheye view depends on the size of the tile in the normal view, the distance from focus tile and the importance of the item that the tile contains. We notice that the tile (3) and (4) had the same size in normal view but was not magnified with the same degree because their services don't have the same similarity with Amazon services.

V. PERSPECTIVES

In this paper, we described a reformulation of the degree of interest introduced by Furnas [4] for fisheye view. The importance of an item is dynamic and depends also on the current focus item. Building a semantic model that takes into account the user context (location, time ...) and preferences to determine the degree of interest of an item is our major improvement axis that has a strong impact on the navigation of the user with smart glasses. Also, a more sophisticated aggregate function that replaces the weighted average is necessary for computing the global similarity.

In terms of visualization, fisheye view must support changes in the context and user preferences with flexibility using different emphases techniques.

We are planning an evaluation of our approach through the recruitment of users for an experiment by creating two interfaces: one with a classic fisheye view and the second with our semantic fisheye view to accomplish tasks on smart glasses and make a comparison on different criteria.

VI. CONCLUSION

Using of wearable devices like smart glasses is continually increasing. Smart glasses have a wide services area on a limited screen size. This involves viewing problems and interaction.

It becomes important to offer carefully designed interfaces that fit with small screens and overcome their limitations in order to improve the user experience. They should help users to easily navigate and understand the information presented to perform their tasks quickly and efficiently.

We aimed to address this problem by proposing a semantic fisheye view on tiled interfaces which allows enriching the user navigation using the relationships between the different services contained in the tiles.

Research on these aspects has a strong impact on the support of the mobile technology and the rise of using smart glasses.

ACKNOWLEDGMENT

This work was performed within the eGLASSES project, which is partially funded by NCBiR, FWF, SNSF, ANRand FNR under the ERA-NET CHIST-ERAII framework. The authors thank Arthur DOQUET for his help with the development part.

REFERENCES

[1] Flynt, David Wayne, et al. "Tile space user interface for mobile devices." U.S. Patent No. 7,933,632. 26 Apr. 2011.

[2] Walter, Wolfgang E., and Christoph Persich. "Active Tiled User Interface." U.S. Patent Application No. 11/829,025.

[3] Leung, Y. K., & Apperley, M. D. (1994). A review and taxonomy of distortion-oriented presentation techniques. ACM Transactions on Computer-Human Interaction (TOCHI), 1(2), 126-160.

[4] Furnas, G. W. (1986). Generalized fisheye views (Vol. 17, No. 4, pp. 16-23). ACM.

[5] Sarkar, M., & Brown, M. H. (1992, June). Graphical fisheye views of graphs. In Proceedings of the SIGCHI conference on Human factors in computing systems (pp. 83-91). ACM.

[6] Noik, E. G. (1994, May). A space of presentation emphasis techniques for visualizing graphs. In Graphics Interface (pp. 225-225). CANADIAN INFORMATION PROCESSING SOCIETY.

[7] Pirró, G., & Euzenat, J. (2010). A feature and information theoretic framework for semantic similarity and relatedness. In The Semantic Web–ISWC 2010 (pp. 615-630). Springer Berlin Heidelberg.

[8] Petrakis, E. G., Varelas, G., Hliaoutakis, A., & Raftopoulou, P. (2006). X-similarity: computing semantic similarity between concepts from different ontologies. JDIM, 4(4), 233-237.

[9] Wu, Z., & Palmer, M. (1994, June). Verbs semantics and lexical selection. In Proceedings of the 32nd annual meeting on Association for Computational Linguistics (pp. 133-138). Association for Computational Linguistics.

[10] Niwattanakul, S., Singthongchai, J., Naenudorn, E., & Wanapu, S. (2013, March). Using of Jaccard coefficient for keywords similarity. In Proceedings of the International MultiConference of Engineers and Computer Scientists (Vol. 1, p. 6).

[11] Bederson, B. B., & Hollan, J. D. (1994, November). Pad++: a zooming graphical interface for exploring alternate interface physics. In Proceedings of the 7th annual ACM symposium on User interface software and technology (pp. 17-26). ACM.

[12] Zizi, M., & Beaudouin-Lafon, M. (1995). Hypermedia exploration with interactive dynamic maps. International Journal of Human-Computer Studies, 43(3), 441-464.

[13] Van Ham, F., & Van Wijk, J. J. (2004, October). Interactive visualization of small world graphs. In Information Visualization, 2004. INFOVIS 2004. IEEE Symposium on (pp. 199-206). IEEE.

[14] Janecek, P., & Pu, P. (2002, May). A framework for designing fisheye views to support multiple semantic contexts. In Proceedings of the Working Conference on Advanced Visual Interfaces (pp. 51-58). ACM.

[15] R Janecek, P., & Pu, P. (2005). An evaluation of semantic fisheye views for opportunistic search in an annotated image collection. International Journal on Digital Libraries, 5(1), 42-56.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1409–1412
DOI: 10.15439/2016F512
ACSIS, Vol. 8. ISSN 2300-5963

Cardiovascular data analysis using electronic wearable eyeglasses - preliminary study

Adam Bujnowski, Jacek Ruminski, Mariusz Kaczmarek, Krzysztof Czuszynski, Piotr Przystup
Gdansk University of Technology
ul. Narutowicza 11/12, 80-233 Gdansk, Poland
Email: bujnows@biomed.eti.pg.gda.pl

***Abstract*—The paper presents an alternative approach to the monitoring of the cardiovascular system. The study depicts configurations of the utilized system and preliminary results of electrical and mechanical parameters of the cardiac system which can be measured using a head-worn device.**

I. Introduction

THE continuously progressing miniaturization in modern electronics results in the creation of miniature, yet powerful devices that can be battery powered and network connected. Many of such devices are wearable. The permanently increasing computational power, characterised by a reasonable runtime and fitted with batteries, creates new possibilities.

Wearable electronic eyeglasses can be regarded as an example of such devices. In fact, the eyeglasses are a miniaturized computer that fits in the frame resembling traditional glasses. Many of such devices are available - equipped with various configurations. Most of them have a local near to eye display. These eyeglasses can be either semi-transparent or completely isolating (non-transparent). Most of these electronic glasses are equipped with a single camera for observing the front scene in relation to the person wearing the device.

A major disadvantage of the majority of the commercially available platforms is the limited potential for future expansion. In fact, most of the modifications are limited to writing new applications that will utilize the hardware available on the platform or the hardware wirelessly connected by means of Wi-Fi or Bluetooth interfaces [6], [7].

As a result of the ERA-NET-CHIST-ERA II project *The interactive eyeglasses for mobile, perceptual computing (eGlasses)* a new open platform in the form of the glasses was designed and developed [1]. It is considered as an open platform that can be extended by means of adding local hardware using various interfaces.

The eGlasses consist of a powerful computer running the Linux or Android system. Expansions might regard various aspects of the device's usage. As an example, an Infrared camera allows them to operate in the darkness or can help the user to estimate the object's temperature. The eyetracker camera enables user interface navigation in a hands-free mode [2]. Additional proximity sensors might be used for obstacle detection [3].

One type of possible platform extensions are biomedical measuring units for measuring biosignals such as ECG, EEG, EMG, and the body temperature of the person wearing such a platform [4].

This paper focuses on the measurement of the cardiovascular system. The measurement of the electrical biopotentials from the cardiac muscle (ECG) is usually measured from the contacting electrodes located on the chest - close to the cardiac muscle. A wearable platform such as eGlasses should not be large nor complicated in application. We are limit measuring electrodes location to close to the ordinary glasses frame. There exist known works related to the estimation of the cardiac potentials with ballistic data collected on the head [8].

This paper is organized as follows. In section II, the eGlasses platform is briefly discussed and an experimental set-up is proposed. Section III presents the results of measurements, and section IV outlines the conclusions.

II. Materials and Methods

A. The eGlasses platform

The eGlasses are designed to be expandable. They consist of three major boards referred to as the "base board", "side board" and "T-board". The connection between the boards is made using flat flexible connectors (FFC) and is shown in Fig. 1. The main unit is the base board. Currently, it is based on the DART4460 board from Variscite [5]. The board is based on the dual-core OMAP 4460 with 1GB RAM and internal FLASH memory (8GB). The DART 4460 module is located on the larger board where level converters, position sensors, DC-DC power sources and peripheral connectors are located. In the future, this board might be replaced by a more powerful one. There is a QUAD core board with IMX6 SoC on board in its preparatory stage. The main board is designed to be located in the right panel of the eGlasses. On the opposite side - the right panel - there is a battery holder with an additional board carrying mainly the USB hub with a USB-to-serial converter. In the front panel, in the centre, the T-shaped board carrying the front camera is located. This board additionally holds the IR camera and is used as an interconnection for signals between the base board and the side board.

The major advantage of the eGlasses platform is the amount of available IO ports. User can have several USB ports,

This work has sponsored by the ERA-NET-CHIST-ERA II project *The interactive eyeglasses for mobile, perceptual computing (eGlasses)* and partly from Gdansk University if Technology - Faculty of Electronics, Telecommunications and Informatics statue funds.

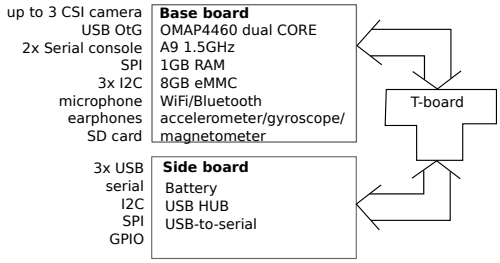


Fig. 1. Block diagram of the eGlasses platform

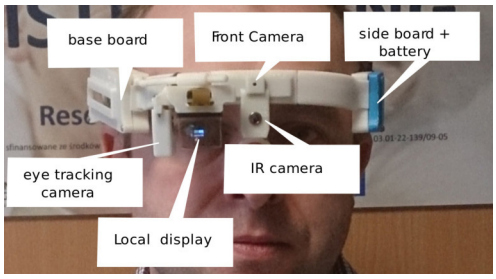


Fig. 2. The eGlasses and location of major components

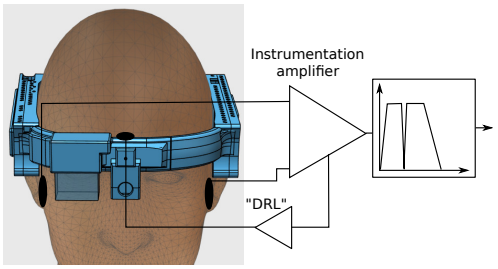


Fig. 3. Electrode's location for the ECG measurement using eGlasses

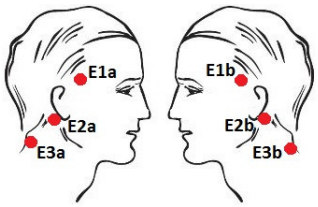


Fig. 4. ECG electrode's placement on the head

serial, SPI, I2C and USB ports for interfacing with additional extensions. The eGlasses prototype is shown in fig. 2.

B. Cardiac measurements

Self-diagnostics can play an important role in several fields of life. One of the most frequently explored types of vital systems are those related to the cardiovascular system. There is a variety of different methods for cardiac measurements. One of the simplest and cheapest techniques is an electrical cardiac potentials measurement. The ECG signal is generated by dedicated cell groups inside the heart. Outside it can be regarded as a current dipole. A conventional ECG test measures a difference of potentials between specific locations on the chest. In the case of the eGlasses we raise the question if it is possible to measure the ECG signal on the head. Therefore, we have selected an electrode location on the both sides of the face in front of the ears. In order to improve the CMRR, the ECG measurement system uses a dedicated electrode that returns the inverted common potential back to the body, i.e. the DRL (Driven Right Leg) solution is utilized. A middle point between measurements electrodes was selected as the location of the DRL electrode (Fig. 3). The ECG signal observed is characterised by a weak amplitude. In order to measure such a signal, we have designed a custom single channel ECG measurement system with five times greater gain when compared to an ordinary chest-based system.

Another signal coming out of the cardiac system is blood pulsation in vessels. A piezoelectric transducer was used to measure blood pulsation. We have used a temporal artery to access the mechanical pulsation. Moreover, the optimal position of the electrodes has been tested. We have prepared one more module, allowing simultaneous measurements of four differential signals. The first channel has been used for traditional ECG measurement, thus two electrodes were placed on the examined person's chest. The other three channels were used to measure the ECG on the head. Electrodes were placed on the temples (ECG1), behind the ears (ECG2) and on the neck (ECG3). Signals were recorded with 24-bit resolution and 250 Hz sampling rate. Each channel has been filtered with a 50 Hz notch filter in digital domain. The base line drift has been removed using a median filter (Fig. 4).

C. Experiment setup

In order to conduct the experiment, we have prepared two modules of ECG and a piezoelectric sensor (Fig. 5). The module marked as ECG1 with ordinary amplification was used for the chest as a reference module, and ECG2 with five times greater amplification was used for head measurements. We have used a piezoelectric sensor and a simple amplifier. The signals were connected to the digital storage oscilloscope and recorded on a USB stick. We have collected data from five volunteers aged 22-46 to prove the system's performance. To acquire ECG signals, standard adhesive and disposable electrodes were used.

III. Results

An experimental set-up has been created on the prototyping bread-board (Fig. 6). The ECG1 and ECG2 amplifiers were assembled on separate printed circuit boards (PCB). The piezoelectric transducer amplifier was assembled using trough hole components located directly on the bread-board. All operational amplifiers enabled the operation from a single power supply. We used a 5V DC power supply with medical class separation from the mains.

As a piezoelectric transducer we have used FT-27T-4.0A transducer with 4kHz resonant frequency, 200Ω of resonant resistance and 25nF capacitance. The unfiltered signals were

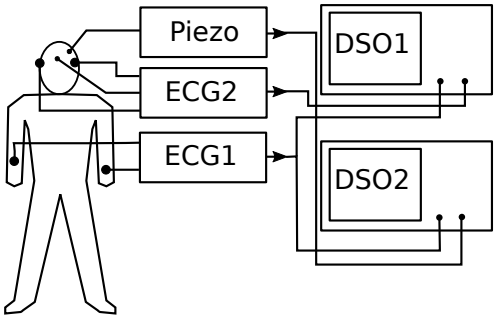


Fig. 5. Block diagram of the experiment setup

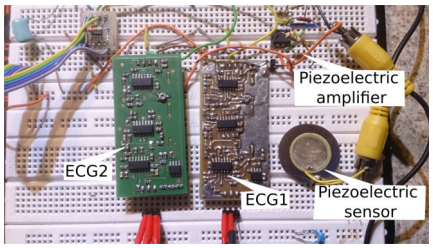


Fig. 6. ECG measurement boards

connected to two digital storage oscilloscopes (DSO), simultaneously recording data (Fig. 7). In order to synchronize the data streams, we connected output of the ECG1 signal to both instruments. It allowed us to estimate the time shift between the instruments.

Based on the raw data, one might notice a similarity between the ECG recorded using the standard procedure and the signal obtained from the head. Additionally, the data recorded from the pulsating vessel resembles arterial pulse (Fig. 8). The signals recorded reflect the standard Wiggers diagram, where atrial pressure peaks are in relation to the ECG signal. In Fig. 8, T-peaks of the electrocardiogram are coinciding with the arterial pressure peaks related to the opening of the mitral valve.

The relations between the ECG signals recorded by elec-

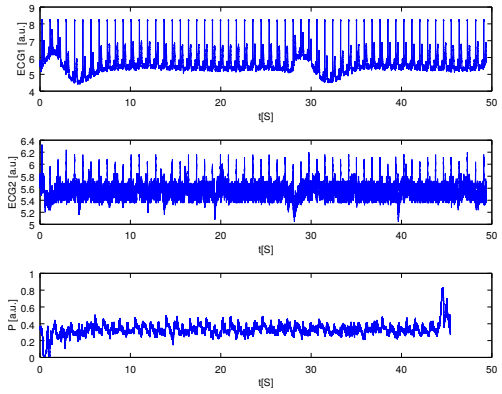


Fig. 7. Results of measurements - raw signals from the top: ECG signal from the chest, ECG signal measured on the head and piezoelectric pulse

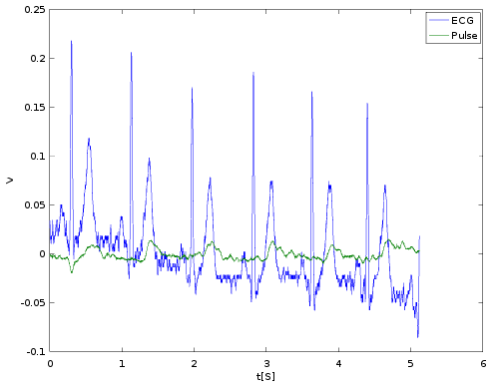


Fig. 8. Results of measurements - ECG data with recorded pulse signal measured by the piezo sensor

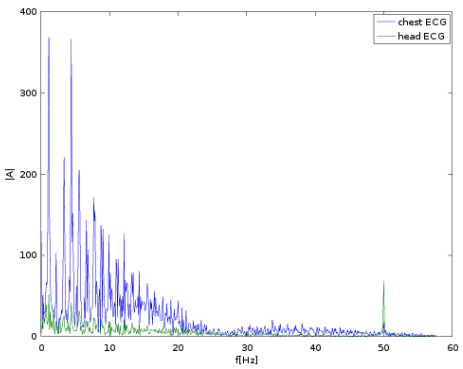


Fig. 9. Simultaneous measurements of the chest ECG with ECG recorded from the head spectrum of modulus

trodes located in different places on the body (chest and head) were examined (Fig. 11). The amplification of the head ECG is about five times greater than the chest ECG. Moreover, the recorded QRS amplitude is much smaller for the head-measured signal.

The components of both signals were examined by means of the Fourier transformation (Fig. 9). The spectrum of the signals was restricted to 60 Hz. The spectrum data showed a peak of the component at 50Hz. After filtration, the 50Hz component disappeared and simultaneously the head ECG became more ”clear” (Fig. 10). When analysing the signal from Fig. 10, QRS peaks are visible and the heart-rate can be calculated, however, the ratio of the detected QRS periods is about 10% smaller than that from the signal acquired on the chest.

Finally, the optimum position of the electrodes has been selected in order to obtain the highest amplitude of the QRS complex (Fig. 11). The amplitudes of the QRS complexes for the examined leads can be found in Tab. I.

IV. Conclusion

The measurement of the ECG signal using electrodes located on the head is possible, however, a custom design of the acquisition system should be prepared. The ECG signal

TABLE I
AVERAGE QRS COMPLEX AMPLITUDE MEASURED ON THE VARIOUS POSITION ON THE HEAD

	Chest ECG	**ECG1**	**ECG2**	**ECG3**
QRS amplitude	$650\mu V$	No QRS	$30\mu V$	$30\mu V$

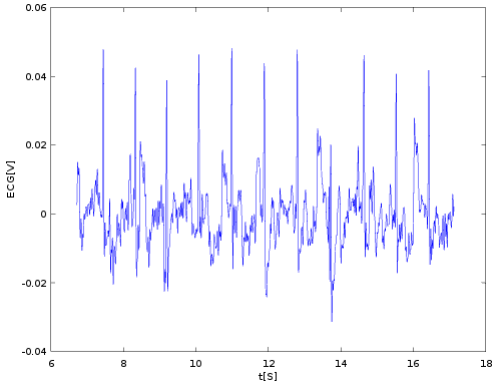


Fig. 10. Low-pass filtered ECG signal measured on the head (bandwidth 0.05-30Hz)

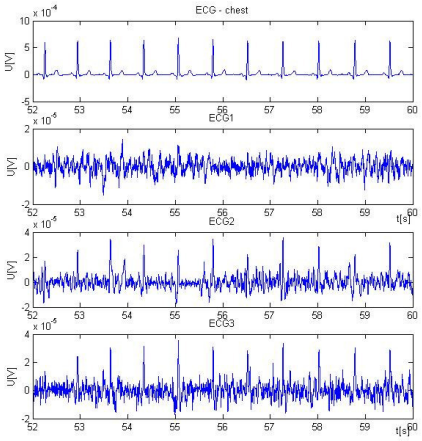


Fig. 11. ECG signals recorded on chest (upper plot) and on the head (ECG1-3)

measured on the head is nosier, but it is still possible to recognize the QRS periods and calculate the heart rate.

It is not possible to acquire the proper ECG signal from any electrode combination over the head. A detailed analysis should be performed to choose the best electrodes set-up. A preliminary study shows that it is possible to record the QRS complex using electrodes located on the neck, or behind the ears. However, the recorded signal is approximately 20 times smaller than the signal measured in a conventional way. Assuming that the ECG will be measured using the eGlasses platform, the preferable position of the electrodes is just behind the ears. In this case, the electrodes could be embedded in the earpiece of the glasses.

Additionally, it is possible to measure the blood pulsation component on the head using a temporal artery. Unfortunately, the localization of this artery on the head is difficult. In the study presented, traditional disposable and adhesive ECG electrodes were used. In the case of a wearable platform, long-term electrodes should be considered.

REFERENCES

[1] Homepage of the eGlasses project; http://www.e-glasses.net [accessed 15.03.2016]
[2] Kocejko, T., Ruminski, J., Wtorek, J., Martin, B., Eye tracking within near - to - eye display, in Human System Interactions (HSI), 2015 8th International Conference on, pp.166, - 172, 25 - 27 June 2015, doi: 10.1109/HSI.2015.7170661
[3] Bujnowski, A.,Czuszynski, K., Ruminski, J., Wtorek, J., McCall, R., Popleteev, A., Louveton, N., Engel, T., Comparison of active proximity radars for the wearable devices, in Human System Interactions (HSI), 2015 8th International Conference on , pp.158 - 165, 25 -27 June 2015 doi:10.1109/HSI.2015.7170
[4] Bujnowski A., Ruminski J., Przystup P., Czuszynski K., Kocejko T. Self Diagnostics Using Smart Glasses - Preliminary Study, in Human System Interactions (HSI), 2016 9th International Conference on (accepted)
[5] The DART4460 specification http://www.variscite.com/products/system-on-module-som/cortex-a9/dart-4460-cpu-ti-omap-4-omap4460 [accessed 14.03.2016]
[6] J. Ruminski, M. Smiatacz, A. Bujnowski, A. Andrushevich, M. Biallas, and R. Kistler, Interactions with recognized patients using smart glasses, in Proceedings of the 8th International Conference on Human System Interactions (HSI 15), pp. 187194, June 2015.
[7] J. Ruminski, A. Bujnowski, J. Wtorek, A. Andrushevich, M. Biallas, and R. Kistler, Interactions with recognized objects, in Proceedings of the 7th International Conference on Human System Interactions (HSI 14), pp. 101105, Costa da Caparica, Portugal, June 2014.
[8] D. Da He, E. S. Winokur and C. G. Sodini, "A continuous, wearable, and wireless heart monitor using head ballistocardiogram (BCG) and head electrocardiogram (ECG)", Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE, Boston, MA, 2011, pp. 4729-4732.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1413–1416
DOI: 10.15439/2016F501
ACSIS, Vol. 8. ISSN 2300-5963

Accuracy analysis of the RSSI BLE SensorTag signal for indoor localization purposes

Mariusz Kaczmarek
Gdansk University of Technology Narutowicza 11/12, 80-233 Gdansk, Poland
Email: mariusz.kaczmarek@eti.pg.gda.pl

Jacek Ruminski, Adam Bujnowski
Gdansk University of Technology Narutowicza 11/12, 80-233 Gdansk, Poland
Email: {jwr, bujnows}@biomed.eti.pg.gda.pl}

Abstract—In t**his paper we describe possibility of use the RSSI signal (Radio Signal Strength Indication) from Texas Instruments SensorTag CC2650 for indoor positioning purposes. This idea is not a new but in our opinion it is possible to use SensorTags with Bluetooth LE wireless interface for positioning inside buildings in such applications as people findings in hospitals, senior come care, etc. RSSI is mostly selected as the sensor localization method in the indoor circumstances. In this paper, we aim to analyze accuracy, calibrate and map RSSI to distance by doing a series of the experiments. Obtained results are very promising and shows possibility of use this technique for position estimation.**

I. Introduction

IDEA of position localization basing on wireless networks is widely known for mobile phones where using information about signal strength from BTS (Base Transceiver Station) one can determine the position of mobile phone speaker [1]. However, this is a coarse location, which is not suitable for indoors use.

Many technologies have been investigated to bridge the gap and bring positioning indoors, such as a combination of AGPS, accelerometer and magnetometer [2], Bluetooth [3], Ultrawideband [4], ZigBee [5]. Wi-Fi is one of most discussed of them, and is considered as the most promising one as the infrastructure and user equipment is already widely available, e.g. in public buildings, public area like parks, airports or railway stations, and it is able to deliver accuracies in the range of a few meters. An exemplary distribution of wireless networks in the building of the Faculty of Electronics, Telecommunications and Informatics of Gdansk University of Technology is shown in Fig. 1. One can choose the best configuration for signal strength scanning. Wi-Fi fingerprinting was pioneered in [6], and has since attracted considerable interest, mainly focused on increasing the accuracy of the technique.

But the use of Wi-Fi network access points can be limited due to different artifacts such as different sensitivity of chipsets in mobile devices [7][8]. Some efforts have also been made in the literature to reduce the effects of RSS variations due to channel impediments by using a compressive sensing (CS) principle such as in [9].

For private home position localization more suitable will be use of bluetooth devices called beacons or like e.g. Texsas Instruments BLE SensorTag CC2650 [10]. This is cheap devices for controlling environmental parameters like ambient temperature, humidity, air pressure, luxometer data and accelerometer data. A good example of implementation of the indoor positioning system is Nashvile project: Mayor, Music City Center Unveil Wayfinding App [11].

II. Material and Methods

A. Bluetooth LE

The Bluetooth 1.0 standard was introduced by SIG in 1999 [12]. The new specification of Bluetooth 4.0LE improved technology that helps everyday gadgets stay paired longer while using less power. Bluetooth 4.0 enables a new class of gadgets such as fitness trackers, medical devices, key fobs for car, beacons sensors and even home lighting controls.

B. Bluetooth and Wi-Fi possible interference

Because both Wi-Fi and Bluetooth wireless technology share 2.4GHz frequency and spectrum and will often be located in close physical proximity to one another, there is concern for how they may interfere with one another. Fig. 1. shows the Wi-Fi networks and spectrum at Faculty of Electronics, Telecommunication and Informatics GUT. Wi-Fi and Bluetooth fail gracefully in the presence of interference. By this is meant that the communication protocols are very robust and include mechanisms for error checking and correcting, as well as requesting that corrupted packets be resent. Therefore the result of increasing levels of

This work was supported by ERA-NET-CHIST-ERA II eGLASSES – The interactive eyeglasses for mobile, perceptual computing; and by European Regional Development Fund concerning the project: UDA-POIG.01.03.01-22-139/09-03 -"Home assistance for elders and disabled – DOMESTIC", Innovative Economy 2007-2013, National Cohesion Strategy

interference is almost always confined to a slowing of the data rate as more packets need to be resent.

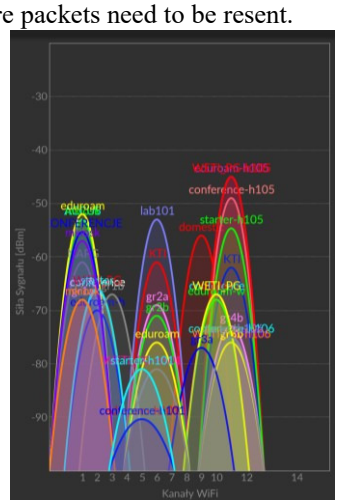


Fig. 1 Wi-Fi network spectrum measurements

C. RSSI measurements and distance calculations

Received signal strength indication (RSSI) is a measurement of the power level received by sensor. Because radio waves propagate according to the inverse-square law, distance can be approximated based on the relationship between transmitted and received signal strength (the transmission strength is a constant based on the equipment being used), as long as no other errors contribute to faulty results. Various experts have given nice theoretical ideas of the model propagation. These theoretical ideas prepares ground for starting and estimating the parameters.

$$RSSI = -10\mathrm{n} \cdot \log(d) + A\,, \qquad (1)$$

where d = distance, A = txPower at distance of 1m, n = signal propagation constant, usually in free space n = 2, but it will vary based on local geometry - for example, a wall will reduce RSSI by ~3dBm and will affect n accordingly and [RSSI] = dBm.

However, practically you need to measure RSSI value at different known and convenient distances and prepare chart. This measurements should be done in various seasons (day, night, dry, cold, rainy). These charts will provide you procedure of extrapolation or intrapolation of the distance.

D. Position estimations

Based on the coordinates of three reference nodes: ST#1(x1, y1), ST#2(x2, y2), and ST#3(x3, y3), and the corresponding distances (calculated from RSSI measured signal) from each reference node to the target node: R1, R2, and R3, we can obtain the following equations [13]:

$$\begin{aligned} (x_1 - x)^2 + (y_1 - y)^2 &= R_1^2 \\ (x_2 - x)^2 + (y_2 - y)^2 &= R_2^2\,, \\ (x_3 - x)^2 + (y_3 - y)^2 &= R_3^2 \end{aligned} \qquad (2)$$

where (x, y) denotes the (unknown) coordinates of the target **P**.

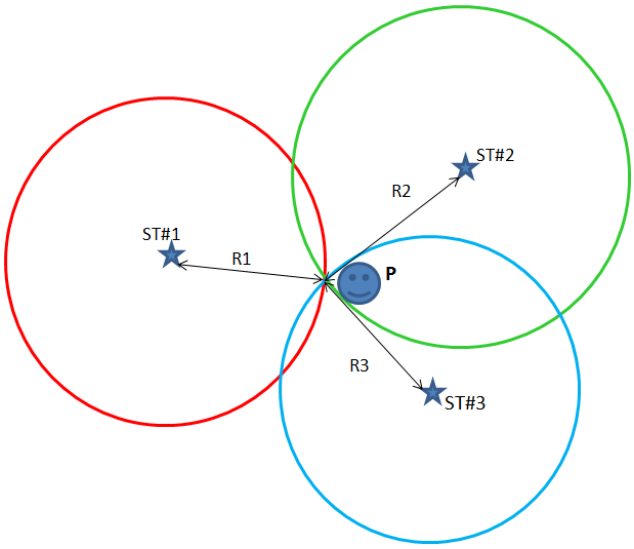


Fig. 2 Position estimation basing on three distances from SensorTags ST#1. ST#2, ST#3

III. EXPERIMENTS AND RESULTS

We perform several experiments in different configurations. Two of them are shown in Fig. 3. Firs case is called direct view because SensorTags array was in direct view with receiving smartphone without any obstacles, the second one was called with wall obstacle because the sensor array was hidden behind the wall's corner.

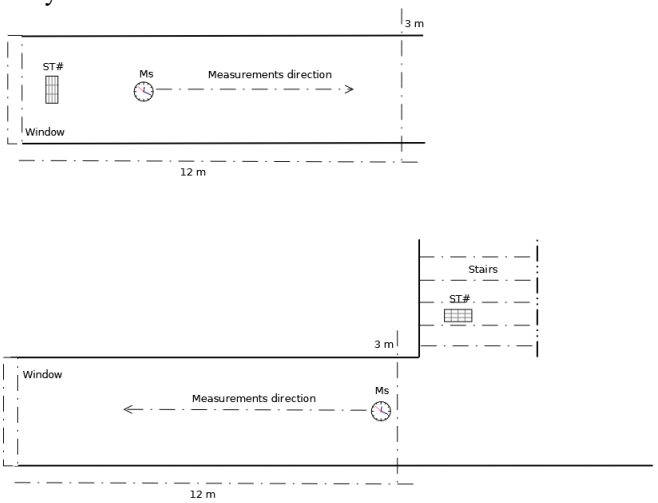


Fig. 3 Configuration of measurements procedure of RSSI (CC2650); a) direct view, b) with wall obstacle

Measurements was done using the written application (Fig. 8) - equipment: smartphone: One Plus One, Android 5.1.1, API 22 TI Sensor Tag CC2650. For fixed distance (range: 0.1m up to 6m) it was read RSSI value for the three devices that radiate towards the receiver (n=10 readings for each device). Results for both configuration are shown in Fig. 4 and Fig. 5.

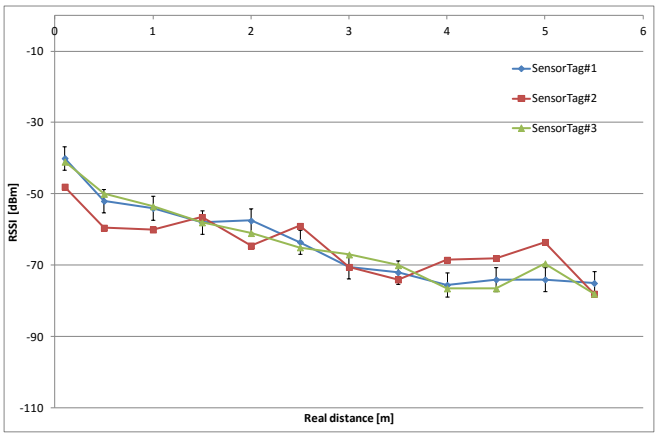


Fig. 4 RSSI vs distance measurements for three TI CC2650 SensorTags - direct view

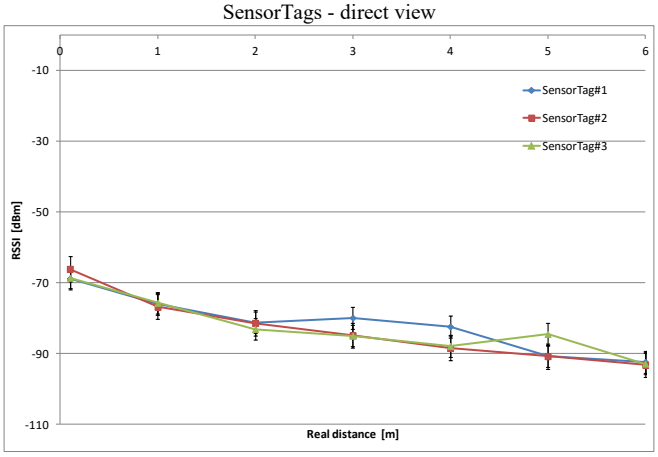


Fig. 5 RSSI vs distance measurements for three TI CC2650 SensorTags with wall obstacle

Basing on obtained readings the mean values and standard deviation values were calculated and plotted (RSSI vs "known" distance) on the charts - Table I. Next the data were fitted to polynomial model according to equation (3).

$$RSSI = \mathrm{A} \cdot d^2 + B \cdot d + C\,, \qquad (3)$$

where: A, B, C - parameters, d - distance.

It seems to be the best way to calibrate the system according to flat/home configurations such as walls, furniture and other obstacles. Example of fitted curve for measured data for SensorTag#3 in direct view experiment is shown in Fig.6.

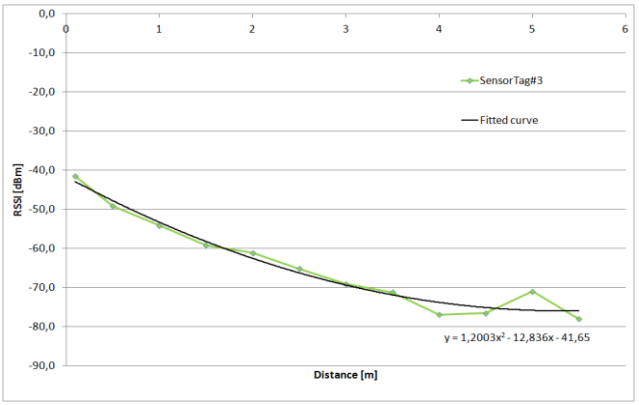


Fig. 6 Dependency between distance and RSSI. The continuous line represents the linear regression model. Dots are mean value of 10 measurements for each distance

IV. Application

Possible scenarios for the use of the proposed method include: a hospitals, nursing homes, senior homes but also museums, airports or train stations.

One example [14] is shown in Fig. 7. and Fig. 8. Inside the senior's flat one can place TI CC2650 sensors for remote monitoring of the environmental conditions (ambient temperature, humidity, luminance). Added value of such Bluetooth LE (BLE) system is possibility of indoor target localization. Dedicated application (Fig. 8) with possibility of import flat schema and localization of sensor allows for calibration measurements of RSSI signal. Basing on this procedure it is possible to track target among short distances.

As there is no fixed standard which manufacturers are required to follow, signal strength indications are to be used for indication only and do not indicate the true absolute signal strength received. These values are reported by a piece of software which allows the operating system to use the wireless card – i.e. the drivers. These drivers feature the role of controlling and reporting the status of the card, and

TABLE I.
EXPERIMENTS RESULTS FOR DIRECT VIEW PROCEDURE, ONEPLUS ONE SMARTPHONE WITH ANDROID 6.0.1+ TI CC2650

Distance [m]	RSSI mean value [dBm]			RSSI median value [dBm]			Standard deviation [dBm]			Estimated distance - equation (1) [m]			Estimated distance - calibrated model, equation (3) [m]		
0.1	-42.0	-40	-48	-41	-48.0	-41.5	4.7	3.5	1.8	0.08	0.12	0.07	-0.02	-0.60	-0.01
0.5	-51.5	-52	-59.5	-50	-60.0	-49.1	2.2	5.8	2.1	0.60	1.12	0.38	0.80	1.48	0.62
1	-54.2	-54	-60	-53.5	-59.3	-54.1	3.3	4.7	1.7	1.00	1.00	1.00	1.06	1.34	1.08
1.5	-57.7	-58	-56.5	-58	-58.0	-59.1	4.5	6.9	4.8	1.57	0.87	1.89	1.42	1.08	1.60
2	-58.7	-57.5	-64.5	-61	-64.2	-61.1	2.6	3.7	1.4	1.78	1.77	2.41	1.53	2.42	1.83
2.5	-63.0	-63.5	-59	-65	-58.8	-65.2	1.7	1.9	3.3	2.98	0.95	3.90	2.05	1.24	2.35
3	-73.0	-70.5	-70.5	-67	-69.1	-69.1	6.9	5.0	5.9	9.05	3.03	6.04	3.84	3.85	2.95
3.5	-73.9	-72	-74	-70	-72.5	-71.2	8.7	5.4	4.3	9.93	4.35	7.59	4.13	5.48	3.35
4	-75.9	-75.5	-68.5	-76.5	-68.0	-76.9	6.2	2.5	7.3	12.18	2.70	13.65	6.23	3.48	6.89
4.5	-73.0	-74	-68	-76.5	-70.4	-76.5	4.2	6.5	6.2	9.05	3.49	13.12	3.84	4.36	6.89
5	-73.6	-74	-63.5	-69.5	-65.4	-71.0	5.7	4.5	4.0	9.63	2.03	7.43	4.03	2.73	3.31
5.5	-76.0	-75	-78	-78	-77.2	-78.0	5.2	3.6	4.8	12.30	6.99	15.22	6.20	6.99	6.99

therefore the strengths reported by the card are highly dependent on the mapping which is established between hardware analog to digital converse values and RSSI values reported by the driver.

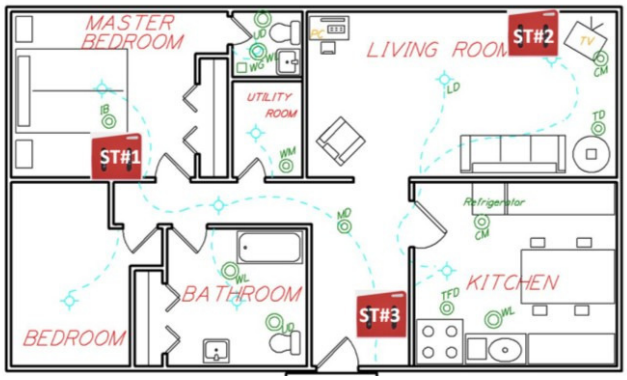


Fig. 7 Example of possible implementation in home environment for three TI CC2650 SensorTags

Different device design and usage by end users could also lead to different signal levels due to human influences. Furthermore, differences in the environment from interfering access points and devices, as well as human traffic and changes in furniture layout will cause different RSSIs to be received in the same location.

Fig. 8 GUI of mobile application for BLE indoor positioning

Disadvantages of the method:

- acceptable minimum version of Android is 4.2 (API 19)
- CC2560 - lack of continuous operation mode,
- very unstable RSSI readings.

V. Conclusion

It was argued that there are many factors which can affect the RSSI returned by a BLE devices, including the antenna design, hardware design, drivers and the environment. Given the large number of factors governing the received RSSI, calibration is unlikely to be able to compensate for all of them, leading us to conclude that there is an inherent limit to the accuracy of a BLE positioning system especially when multiple devices are used.

Small scale signal variations (e.g. multipath) may greatly affect the RSS measurement. Variations of up to 30-40dB have been reported [15]. We have measured at least 2-6dB variations in indoor deployments of low power Bluetooth LE networks. Therefore, we would suggest that instead of using a single RSSI measurement to estimate distance, try using the average or median value of N measurements collected on the same spot (at least N>20) so that you can reduce the effect of small scale fading. Then you can use the log-distance model with more accuracy. If you have more measurements, extract the basic characteristics of the propagation environment first (like path loss exponent etc), to achieve better results. Another interesting issue is the question of the deployment of BLE tag's in the home environment for optimal readings and determination of the position.

References

[1] M. Sauter (2010). "3.7.1 Mobility Management in the Cell-DCH State". From GSM to LTE: An Introduction to Mobile Networks and Mobile Broadband. John Wiley & Sons. ISBN 9780470978221.

[2] D. Gusenbauer, C. Isert, and J. Krosche, “Self-contained indoor positioning on off-the-shelf mobile devices”, Indoor Positioning and Indoor Navigation 2010, pp. 1-9.

[3] S. Feldmann, K. Kyamakya, A. Zapater, and Z. Lue, “An indoor Bluetooth-based positioning system: concept, implementation and experimental evaluation”, International Conference on Wireless Networks, 2003.

[4] C. Zhang, M. Kuhn, B. Merkl, A.E. Fathy, and M. Mahfouz, “Accurate UWB indoor localization system utilizing time difference of arrival approach”, IEEE Radio and Wireless Symposium, pp. 515-518, 17-19th Oct. 2006.

[5] M. Sugano, “Indoor localization system using rssi measurement of wireless sensor network on zigbee standard”, Wireless and Optical Communications, pp. 1-6, 2006.

[6] P. Bahl, V.N. Padmanabhan, “RADAR: an in-building RF-based user location and tracking system”, Proc. of Infocom, pp. 775-784, 2000.

[7] A. Parameswaran, H. Thottam, S. Upadhyaya, "Is RSSI a Reliable Parameter in Sensor Localization Algorithms – An Experimental Study" (PDF). September 2009. 28th International Symposium On Reliable Distributed Systems, New York. Retrieved 17 March 2013..

[8] L., Gough; Gallagher, Thomas; Binghao, Li. "Differences in RSSI readings made by different Wi-Fi chipsets: A limitation of WLAN localization". Localization and GNSS (ICL-GNSS), 2011 International Conference on.

[9] C. Feng, W. Au, S. Valaee, and Z. Tan, “Compressive sensing based positioning using rss of wlan access points,” in INFOCOM, 2010 Proceedings IEEE, 2010, pp. 1–9.

[10] http://www.ti.com/sitesearch/docs/universalsearch.tsp?searchTerm=cc2650&linkId=10&src=top&m=dd#linkId=3

[11] http://www.nashville.gov/News-Media/News-Article/ID/3477/Mayor-Music-City-Center-Unveil-Wayfinding-App.

[12] http://standards.ieee.org/about/get/802/802.15.html.

[13] G. Han, D. Choi, W. Lim, “A Novel Reference Node Selection Algorithm Based on Trilateration for Indoor Sensor Networks”, IEEE Intl Conf. on Computer and Inform. Technology, 2007,p.1003-1008.

[14] M. Kaczmarek, A. Bujnowski, J. Wtorek, A. Poliński: Multimodal Platform for Continuous Monitoring of the Elderly and Disabled// Journal of Medical Imaging and Health Informatics.. -Vol. 2., issue 1 (2012), p.56-63.

[15] C. Feng, W. Au, S. Valaee, and Z. Tan, “Compressive sensing based positioning using rss of wlan access points,” in INFOCOM, 2010 Proceedings IEEE, 2010, p. 1–9.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1417–1422
DOI: 10.15439/2016F538
ACSIS, Vol. 8. ISSN 2300-5963

Enhanced Eye-Tracking Data: a Dual Sensor System for Smart Glasses Applications

Paweł Krzyżanowski, Tomasz Kocejko, Jacek Ruminski, Adam Bujnowski
Gdansk University of Technology, Department of Biomedical Engineering,
Faculty of Electronics, Telecommunications and Informatics, Poland
Email: pawkrzyz@pg.gda.pl, tomkocej@pg.gda.pl, jacek.ruminski@pg.gda.pl, bujnows@biomed.eti.pg.gda.pl

Abstract – **A technique for the acquisition of an increased number of pupil positions, using a combined sensor consisting of a low-rate camera and a high-rate optical sensor, is presented in this paper. The additional data are provided by the optical movement-detection sensor mounted in close proximity to the eyeball. This proposed solution enables a significant increase in the number of registered fixation points and saccades and can be used in wearable electronics applications where low consumption of resources is required. The results of the experiments conducted here show that the proposed sensor system gives comparable results to those acquired from a high-speed camera and can also be used in the reduction of artefacts in the output signal.**

I. Introduction

THE rapid development of wearable electronics has meant that this technology is ever-present in everyday life; the number of such applications and solutions for human-machine interactions is constantly increasing. However, the miniaturization of these devices and the growing array of functionalities require the development of new interfaces. For many applications, for example septic-type interactions [1], the eye-tracking interface is a very convenient option. With just a glance from the user, it is possible to execute commands. This technology has many possible applications [2, 3]; however, it is still under continuous study [4, 5]. In addition, research to enhance video-based gaze tracking is needed to increase both the quality and the amount of data acquired by the system [6] as well as to increase the speed of real-time processing [7].

Since modern wearable electronics like smart glasses are multitasking devices, the design of interfaces within the constraints of the processor unit's computational power is a challenge. In modern eye-tracking interfaces, the camera frame rate is a factor that can increase the potential number of applications of the interface. With the advanced image processing required for reliable estimation of eye and gaze position, the eye-tracking interface is the device which consumes the majority of the power resources. In addition, a need for high computational power and the excessive power consumption are difficult requirements to fulfil.

In gaze-tracking interfaces, it is important to achieve a correct image acquisition, an accurate calculation of the fixation point and the appropriate processing of these data. In practical applications, the time required by the CPU/GPU to process the data is significant [8] and the reduction of this time results in lower consumption of power, which is also important when considering the design of a battery charging circuit. It should be emphasized that studies which have presented new integrated circuits (ICs) have achieved effective results with software and hardware data processing, but these ICs are not in widespread use [9].

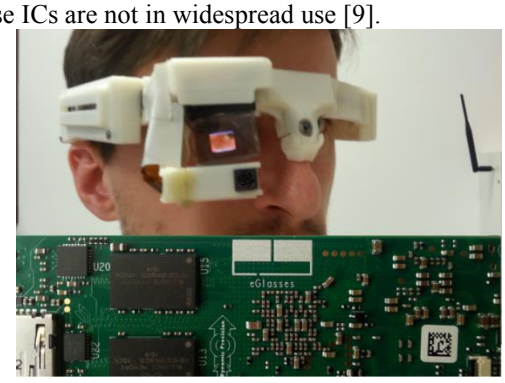


Fig. 1 Prototype of the eGlasses platform

The eGlasses (Fig. 1) electronic eyewear is a multisensory platform capable of running on both Linux and Android operating systems. The eye-tracking module is one of the eGlasses input interfaces. It allows operation by graphic user interface, controlled by gaze, and also provides data that can be used in interaction with everyday objects. The core elements of the eye-tracking module hardware are infrared LED-based eye surface lights, the camera that registers eye movements (eye camera) and the camera that registers the scene images (scene camera). In the current prototype, the eye-tracking module allows a 30 Hz sampling frequency at a resolution of 320x240 pixels. This sampling rate significantly decreases for higher resolutions (e.g., it is only 15 Hz for an image of 640x480 pixels resolution). The

This work has been partially supported by NCBiR, FWF, SNSF, ANR and FNR in the framework of the ERA-NET CHIST-ERA II, Project eGLASSES – The interactive eyeglasses for mobile, perceptual computing and by statutory funds from the Faculty of Electronics, Telecommunications and Informatics Faculty, Gdansk University of Technology.

use of an optical sensor with low power consumption allows a high frame rate to be maintained, while preserving a high resolution in the eye-tracking camera. This study presents a technique for increasing the eye-tracker sampling frequency by combining the low-rate camera with the additional optical sensor. So far such an idea has been used mostly in non-wearable electronic equipment to minimize errors caused by head rotation [10].

The rest of the paper is organized as follows: Section 2 describes the methods, experimental stand and algorithm. The experimental set-up and results are presented in Section 3. Section 4 contains a discussion and potential applications for this new concept. Section 5 presents the conclusion.

II. Methods and Materials

The underlying concept of this research was to use an eye-tracking camera and an additional simple optical sensor to increase the number of registered pupil positions. The principle of operation of the additional sensor was based on measurements of eyeball displacements. To evaluate this idea, a custom-built eye-tracker allowing measurement of pupil movements with a 180 Hz sampling rate was utilized. The optical sensor was then added to complement the eye-tracker. An appropriate test stand was built to conduct the experiments in a controlled environment.

A. Experimental Stand and Data

The experimental hardware equipment consisted of the three main components: the eye-tracking camera, the optical motion sensor, and a model of an eyeball made of plastic. The model was of spherical shape and contained an artificial pupil (Fig. 2).

The stand was equipped with two bearing servomotors (HD3688MG) and the frame was constructed using 3D printer technology. The construction had two degrees of freedom: horizontal (x) and vertical (y). The servomotors were controlled by a PWM signal from the microprocessor. This approach allowed independent movement in both the x- and y-directions.

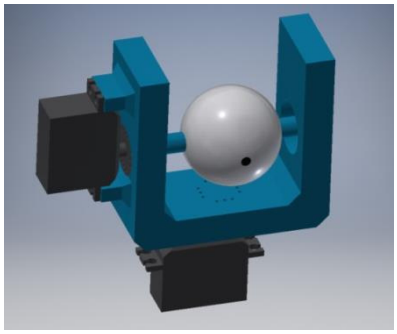


Fig. 2 The 3D frame (Fr) with the eyeball model controlled by servomotors (Sv)

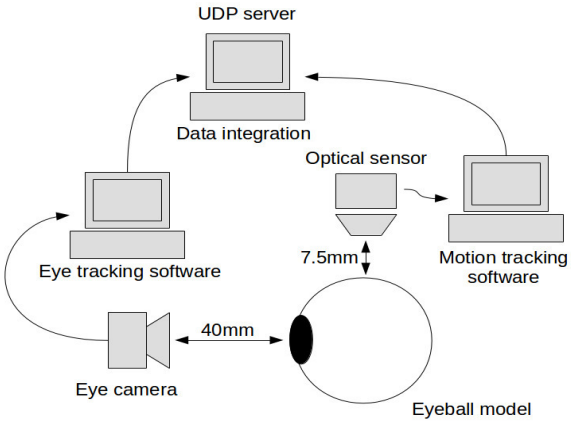
The servomotors allowed the eyeball position to be changed up to 25 times per second. With this frequency it was possible to simulate saccades that lasted 40 ms. Healthy eye movements were therefore able to be simulated [11]. The camera captured the pupil position with a frequency of 180 Hz. The basic parameters of the eye-tracker are presented in Table 1.

Table I
Basic parameters of the eye tracker used

Parameter	Value
Eye-tracking technique:	Dark Pupil, Pupil Centre Detection
Eye-tracking:	Monocular (left eye)
Data rate:	180 Hz
Accuracy:	0.7°
Precision:	0.5°
Light condition restrictions:	No
Eye-tracking camera resolution:	320x240 px

The eye camera was connected to a PC computer with the eye-tracking software installed and running on Linux OS. The main classes of the eye-tracking software covered eye and scene camera image acquisition, pupil-detection algorithms, video for Linux camera support, gstreamer and opencv camera support and fixation and gaze-tracking algorithms. The eye-tracking software offered three algorithms for pupil centre detection; this study was conducted using the most accurate one, based on the selection of pupil boundary candidate points and ellipse fitting.

a)



b)

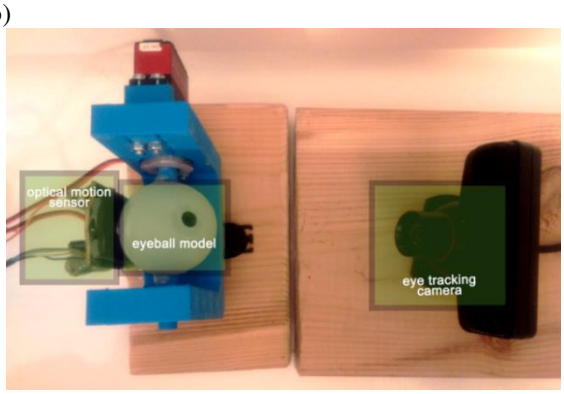


Fig. 3 Configurations of a) the experimental stand and b) its implementation

To track the eyeball motion, the PAN3101, a low cost CMOS optical sensor integrated with DSP, was used. The

DSP serves as an engine for the estimation of non-mechanical motion. The optical navigation technology used in this optical sensor enabled investigators to measure changes in position by optically acquiring sequential surface images (frames) and mathematically determining the direction and magnitude of movement. The current x and y movements can be read from the registers via a serial port. The sensor required additional optics. The one used in this research (HDNS-2100) allowed application of the sensor at a distance of 7.5 mm from the eyeball model. In fact, this is currently a primary impediment to *in vivo* use of this sensor. However, it appears to be possible to increase the distance between the optical sensor and the eyeball, although this would require some changes to the angle between the mirrors and the focal length of the lens. The IR LED diode power should also be adjusted in accordance with safety limits [12]. During experiments, the sampling rate of the sensor was set to 125 Hz. The optical sensor was connected to the additional PC unit. All data acquired by the sensors were transferred using the UDP protocol to the external server for further integration and processing.

There were certain discrepancies in the parameters measured by each sensor. The camera was used to acquire pupil positions frame by frame in an absolute coordinate system, while the optical sensor measured the total eyeball movement (total shift between the last two frames). The shift was the sum of the movement before and after camera sampling, so the value of this shift had to be split proportionally between the frames (Fig. 4).

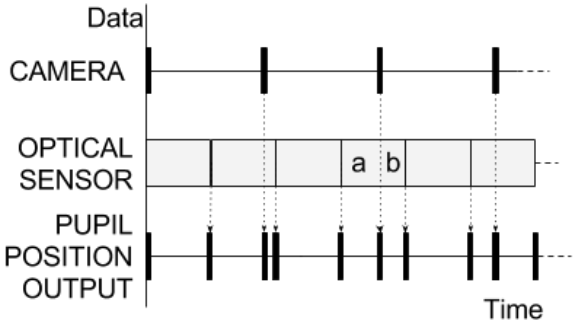


Fig. 4 Output pupil positions and incoming data from both sensors: at a point (the eye camera) and proportional to a shift in time (the optical sensor)

B. Data Acquisition and Fusion Algorithm

The concept underlying of this project was to insert a high frequency signal into a low frequency one (Fig. 5). The algorithm developed for this comprised two major parts. The first was a calibration procedure, while the second was devoted to calculating the additional pupil positions from data registered by the optical sensor, with reference to those captured by the camera. A description of the algorithm is as follows:

1. Start, set the first pupil position from the camera and reset the shift register in the optical sensor (timestamp = 0).
2. Get data from the optical sensor (timestamp, x and y movement).
3. Calculate a new pupil position by adding to the previous pupil position the scaled shift obtained from the optical sensor.
4. Repeat points 2-3 until new data from the eye-tracker is ready.
5. Get data from the eye camera (pupil position and timestamp t_1) and update to a new fixation point.
6. Get data from the optical sensor and split information proportionally to time before and after timestamp t_1.
7. Calculate a new pupil position based on data after t_1.
8. Go to point 2.

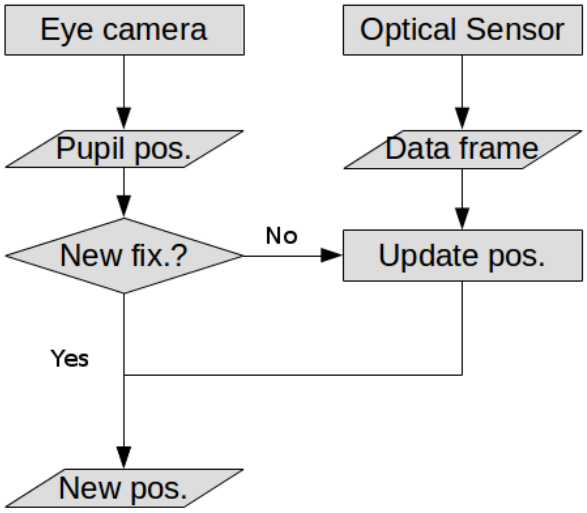


Fig. 5 Block diagram of the proposed algorithm

C. Experiments

To evaluate the concept of extending a set of pupil positions recorded by the eye tracking camera using data acquired by the optical sensor, several experiments were conducted. In the first experiment, a set of pupil positions were recorded over a certain period by both the reference eye-tracking camera (180 Hz sampling rate) and the optical sensor (125 Hz sampling rate). This was done in order to check the similarities of the paths recorded by the sensors. Pearson correlation coefficients were calculated separately for the x and y coordinates:

$$corr_x = \frac{\sum_i (px_i - p\bar{x})(mx_i - m\bar{x})}{\sqrt{\sum_i (px_i - p\bar{x})^2}\sqrt{\sum_i (mx_i - m\bar{x})^2}}, \quad (1)$$

$$corr_y = \frac{\sum_i (py_i - p\bar{y})(my_i - m\bar{y})}{\sqrt{\sum_i (py_i - p\bar{y})^2}\sqrt{\sum_i (my_i - m\bar{y})^2}}, \quad (2)$$

where px_i, py_i are the x and y coordinates of a pupil position, mx_i, my_i are the x and y coordinates of the sample registered by the optical motion sensor, $p\bar{x}$, $p\bar{y}$ are the average values of pupil position samples along the x- and y-axes, and $m\bar{x}$, $m\bar{y}$ are average values of eyeball positions along the x- and y-axes, registered by the optical motion sensor.

Following this, a set of experiments exploring different patterns of movement were performed. Square-shaped

movements (Fig. 7) were used to calculate the proper constant ratio between the measurements from both sensors. A set of pupil positions from the camera was recalculated by using perspective transformation to compensate for the viewing angle. After calibration, various sequences of the eyeball movement were tested (ellipse, triangle and random) with a fixation duration of between 140 and 530 ms. Each session lasted for ten repeated sequences, and the measurements were stored with a synchronized timestamp.

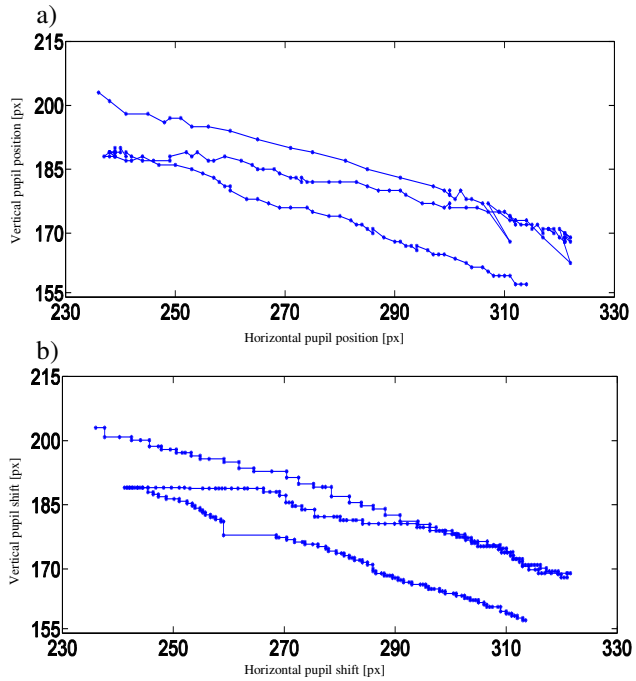


Fig. 6 Similarities between the eyeball movement recorded by the camera and optical sensor: a) The plot of consecutive pupil centre positions detected by the eye tracker b) The plot of the eyeball motion (x, y) coordinates detected by the optical sensor

The sampling rates remained the same; 180 Hz for the camera and 125 Hz for the optical sensor. Next, the original 180 Hz raw data were reduced to simulate recording with 5, 10, 30 and 60 Hz sampling rates. This allowed the creation of a set of data that could be extended by the optical sensor readings and which fully corresponded to the reference data. The results were compared to both: the high-speed 180Hz camera reference as well as the theoretical, programmed ideal eyeball movement. The signals from the optical sensor, the camera and the new combined set had different lengths, and all data were therefore resampled to 1ms using interpolation. The similarity between signals was checked by calculating the mean square error:

$$MSE_x = \frac{1}{n}\sum_{i=1}^{n}(\hat{X}_i - X_i)^2 \,, \quad (3)$$

$$MSE_y = \frac{1}{n}\sum_{i=1}^{n}(\hat{Y}_i - Y_i)^2 \,, \quad (4)$$

where n is number of samples, $\hat{X}_i$and $\hat{Y}_i$ are signal pupil positions in the horizontal and vertical dimensions, and X_i and Y_i are pupil positions from the reference signal (ideal movement or high speed camera).

III. Results

The similarities between the pupil positions recorded by the camera and the eyeball movement acquired by the optical sensor are presented in Fig. 6. The calculated correlation between waveforms recorded by the eye camera at 180 Hz and the optical sensor at 125 Hz were strongly positive:

$$corr_x = 0.91$$

$$corr_y = 0.89$$

In the next experiment, data square movement was implemented to calculate the ratio between sensors. The results are presented in Fig. 7 and numerical values are given in Table 2.

Following this, experiments were performed to check for dependencies between the information gain and the signal variability. The eyeball motion was programmed for fixation duration of 140 ms and data from 30, 10 and 5 Hz camera frequencies were enhanced by the optical sensor. The calculated pupil positions were compared to both reference signals: the high-speed camera and the theoretical route. For greater clarity, the results are presented in Figs. 8-10 and in Table 3. The axes of the coordinate system are pixels (except for the optical sensor figure) and the values plotted are the centred pupil positions.

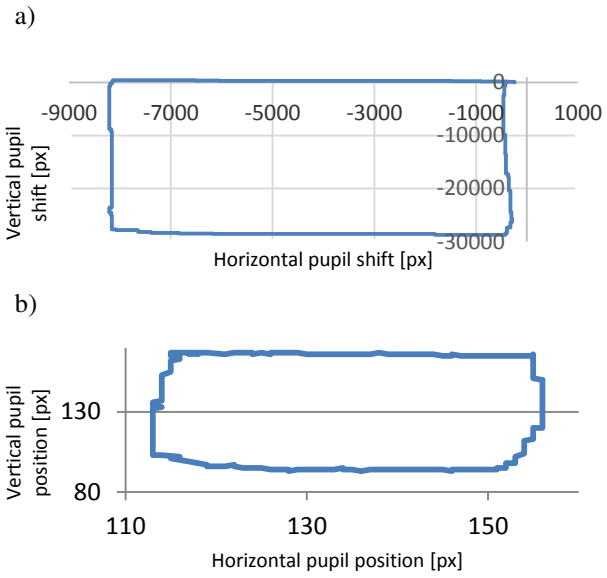


Fig. 7 Calibration of data from both sensors: a) movement recorded by the mouse sensor; b) pupil positions recorded by the high-speed camera after perspective transformation

Table II
Calculated scale constants

Name	Value
X ratio:	$4.58*10^{-3}$
X standard deviation:	$1.59*10^{-4}$
Y ratio:	$2.49*10^{-3}$
Y standard deviation:	$2.08*10^{-5}$

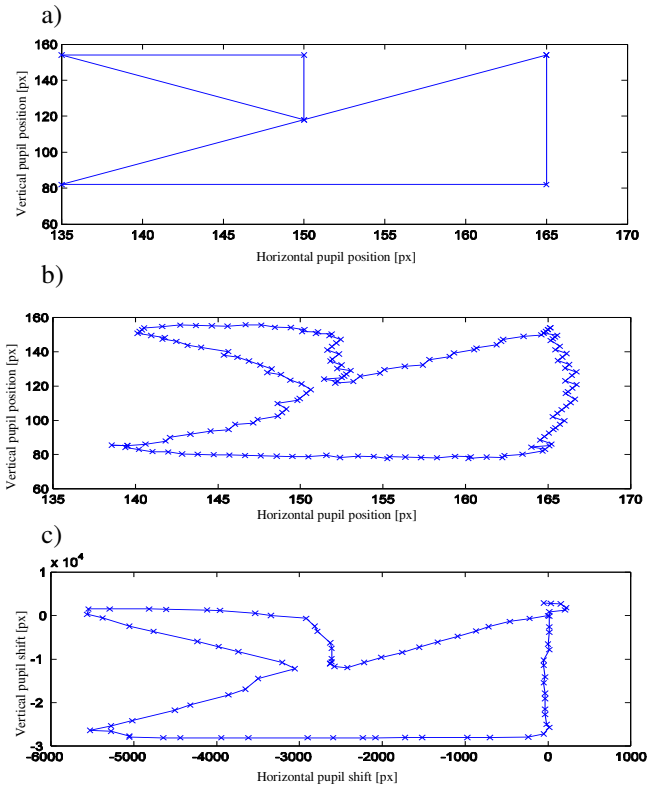


Fig. 8 Raw data before processing: a) ideal movement from the eyeball model; b) pupil positions recorded by the camera; c) eyeball shift measured by the optical sensor

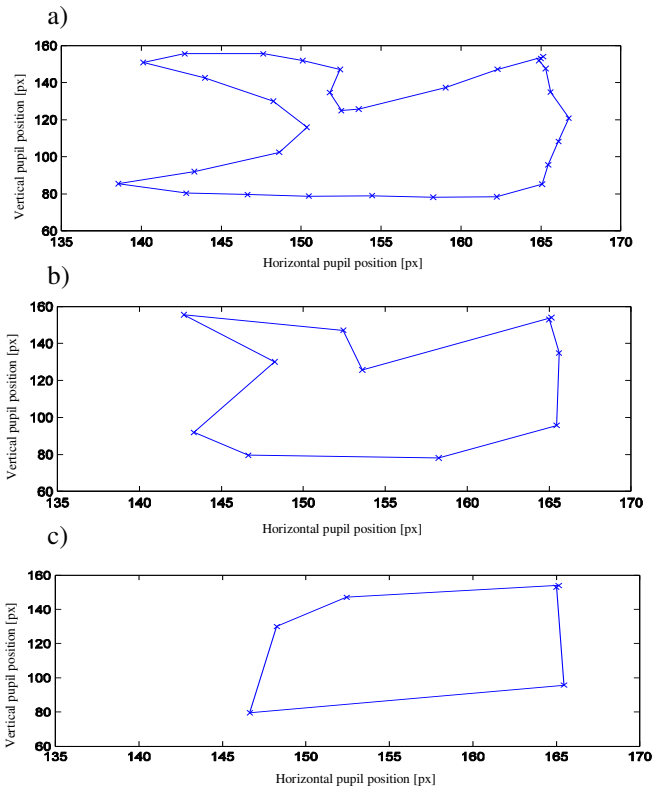


Fig. 9 The effect of decreasing information during reducing of the camera sampling rate: a) 30 Hz camera showing good reproduction of the signal; b) 10 Hz camera showing sufficient reproduction but with some information already lost; c) 5 Hz camera showing high distortion of signal

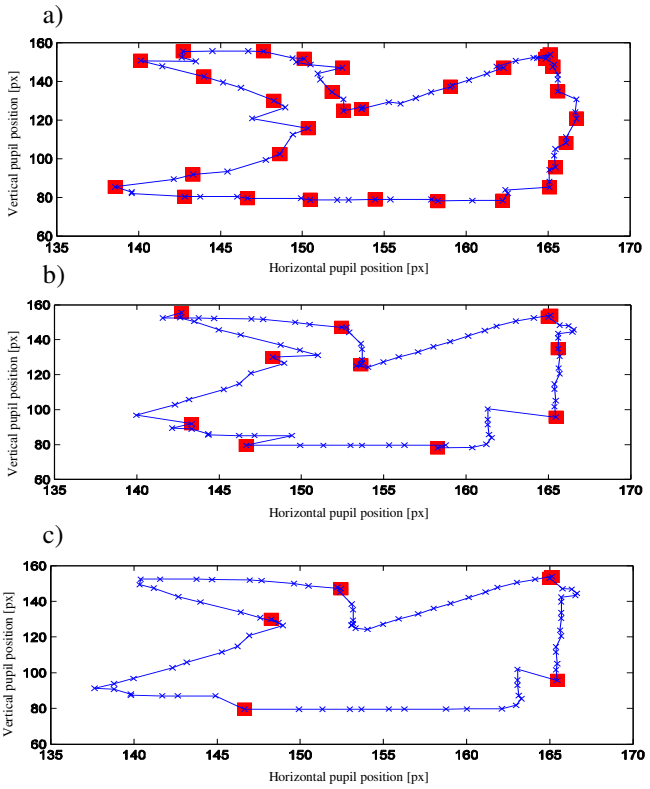


Fig. 10 Enhanced signal after combining data from both sensors. Red squares are pupil positions from camera, and blue crosses are fixations calculated by the system a) 30Hz camera showing no new data compared to the camera system only; b) 10Hz camera signal showing some curves between points; c) 5Hz camera showing the largest improvement compared to using only the camera system

In the last experiment (Fig. 11), a random motion was tested and the system also gave additional data where the sampling rate was insufficient.

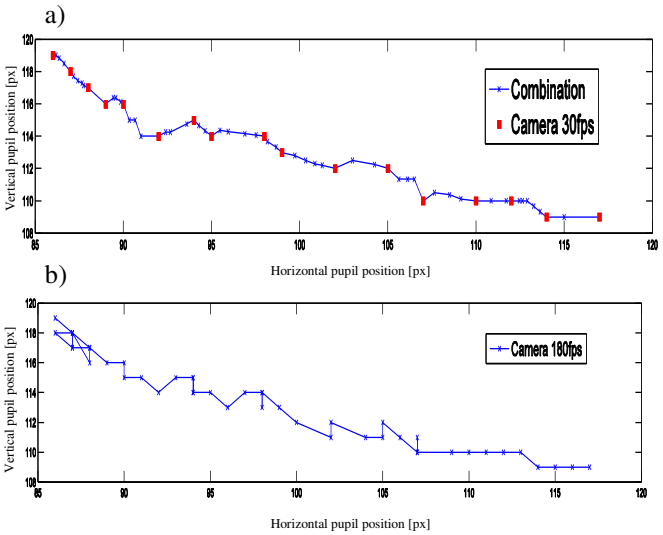


Fig. 11 Experiment combining data from both sensors: a) Low speed camera (red squares) and calculated pupil positions (blue ‘x’) from optical sensor; b) reference high-speed camera

Table III
EXPERIMENT WITH ENHANCED DATA USING THE DUAL SENSOR SYSTEM

Fixation period: 140 ms	MSE			
	Camera - Reference		Combined - Reference	
	X	Y	X	Y
Shape 60 Hz - 180 Hz cam	0.09	0.30	0.25	0.81
Shape 60 Hz - ideal move	7.91	26.00	8.32	24.25
Shape 30 Hz - 180 Hz cam	0.16	0.63	0.71	2.91
Shape 30 Hz - ideal move	8.14	25.73	8.65	19.76
Shape 10 Hz - 180 Hz cam	3.56	20.33	3.40	16.16
Shape 10 Hz - ideal move	8.83	24.42	9.28	15.89
Shape 5 Hz - 180 Hz cam	12.65	124.06	2.66	15.55
Shape 5 Hz - ideal move	30.71	113.25	9.00	20.57

IV. DISCUSSION

These experiments confirmed the assumption that data captured with the eye-tracking camera can be expanded using additional samples gathered by a faster optical sensor.

Some older studies state that the minimum sampling frequency to record saccades should be 300 Hz [13], but newer research shows that a 50-60 Hz sampling rate is sufficient [14]. The additional data acquired from the optical sensor (which can be set up to 3 kHz) should be sufficient for measurement of both fixations and saccades. The system developed here can be used in applications where a low computing time is required, e.g., in systems containing many peripherals that involve computational complexity [15].

Analytical study of the data obtained by means of the dual sensor system gives the expected results. The eyeball movement direction changed 9 times per second. When the camera sampling rate was three times larger than the signal variability, there was no new information gain after combining the data. When the signal fluctuates fast enough when compared to the camera sampling rate, the information gain increases. This can be used to find the optimal usage of computational resources with an acceptable quality of eye tracking in any mobile device.

The additional value of this system emerges from the first and last experiments conducted (Fig. 6 and Fig. 11). In these recordings, the camera eye-tracking algorithm shows small fluctuations or artefacts. In contrast, the optical sensor during the same timestamp shows no distortions. This feature can be used as redundancy and may result in increased reliability of the system.

V. CONCLUSIONS

The study performed here shows that it is possible to extend the number of pupil positions captured by the camera based on data provided by a much faster optical sensor. A high level of similarity between the expanded and reference datasets proves that the proposed algorithm is correct and reliable. The algorithm developed here may influence the implementation of the eye-tracing procedures utilised in wearable electronic devices such as smart glasses. The possibility of extending the number of pupil positions using the optical sensor allows for a significant reduction in the eye-tracking camera sampling rate, and thus enables reduction of the required computational power and power consumption.

REFERENCES

[1] K. Czuszynski, J. Ruminski, T. Kocejko and J. Wtorek, "Septic safe interactions with smart glasses in health care," in Engineering in Medicine and Biology Society, Milan, 2015, pp. 1604-1607, 10.1109/EMBC.2015.7318681

[2] T. Kocejko, A. Bujnowski and J. Wtorek, "Eye-mouse for disabled," in Human-Computer Systems Interaction, Springer Berlin - Heidelberg, 2009, pp. 109-122, 10.1109/HSI.2008.4581433

[3] T. Kocejko, A. Bujnowski and J. Wtorek, "Complex human computer interface for LAS patients," in Human System Interactions, IEEE Press, pp. 272-275, 10.1109/HSI.2009.5090991

[4] Z. Yücel, A. Salah and C. Mericli, "Joint attention by gaze interpolation and saliency," in Cybernetics vol. 43, 2013, pp. 829-842, 10.1109/TSMCB.2012.2216979

[5] B. R. Pires, M. Devyver and A. Tsukada, "Unwrapping the eye for visible-spectrum gaze tracking on wearable devices," in Applications of Computer Vision, 2013, pp. 369-376, 10.1109/WACV.2013.6475042

[6] F. Li, S. Kolakowski and J. Pelz, "Using structured illumination to enhance video-based eye tracking", in Image Processing, ICIP, San Antonio, 2007, pp. 373-376. 10.1109/ICIP.2007.4378969

[7] M. Mehrubeoglu, L. M. Pham and H. T. Le, "Real-time eye tracking using a smart camera," in Applied Imagery Pattern Recognition Workshop (AIPR), IEEE, Washington, 2011, pp. 1-7, 10.1109/AIPR.2011.6176373

[8] A. Al-Rahayfeh and M. Faezipour, "Enhanced frame rate for real-time eye tracking using circular Hough transform," in Systems, Applications and Technology Conference, Farmingdale, 2013, pp. 1-6, 10.1109/LISAT.2013.6578214

[9] D. Kim, J. Cho and S. Lim, "A 5000S/s Single-Chip Smart Eye-Tracking Sensor", in: Solid-State Circuits, ISSCC 2008, pp. 46-59, 10.1109/ISSCC.2008.4523049

[10] D. Beymer and M. Flickner, "Eye gaze tracking using an active stereo head," in Computer Vision and Pattern Recognition vol.2, Madison, USA, 2003, pp. 451-458, 10.1109/CVPR.2003.1211502

[11] W. M. King, S. G. Lisberger and A. F. Fuchs, "Oblique Saccadic Eye Movements of Primates," in Journal of Neurophysiology vol. 56, 1986, pp. 769-784,

[12] *ICNIRP guidelines on limits of exposure to laser radiation of wavelengths between 180 nm and 1,000 μm*, in Health Physics 105(3), 2013, pp. 271-295.

[13] M. Juhola, V. Jantti and I. Pykko, "Effect of sampling frequencies on computation of the maximum velocity of saccadic eye movements", in Biological Cybernetics, Springer-Verlag, 1985, pp. 67-72, 10.1007/BF00337023

[14] R. Wierts, M. Janssen and H. Kingma, "Measuring Saccade Peak Velocity Using a Low-Frequency Sampling Rate of 50 Hz", in Biomedical Engineering, 2008, pp. 2840-2842, 10.1109/TBME.2008.925290

[15] T. Kocejko, J. Ruminsk, J. Wtorek and B. Martin, "Eye tracking within near-to-eye display," in Human System Interactions, Warsaw, 2015, pp. 166-172, 10.1109/HSI.2015.7170661

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1423–1426
DOI: 10.15439/2016F540
ACSIS, Vol. 8. ISSN 2300-5963

Medical simulation center as a model for testing mHealth concepts in prehospital emergency medicine

Dr. med. Bibiana Metelmann, M.D.
Klinik fuer Anaesthesiologie, Universitaetsmedizin Greifswald, Ferdinand Sauerbruch Straße 17475 Greifswald, Germany
Email: bibiana.metelmann@uni-greifswald.de

Dr. med. Camilla Metelmann, M.D.
Klinik fuer Anaesthesiologie, Universitaetsmedizin Greifswald, Ferdinand Sauerbruch Straße 17475 Greifswald, Germany
Email: camilla.metelmann@uni-greifswald.de

***Abstract*—Newly developed mHealth tools need to be tested in standardized conditions without possible patient harm before implementation. One possible approach to secure these two conditions is to analyze the mHealth tool in a medical simulation center. Medical simulation centers create realistic routine or emergency scenarios with the aid of computer-operated mannequins. Medical simulation is widely established, especially in emergency medicine, because it combines theoretical knowledge and practical skills. To evaluate a mHealth concept in the field of prehospital emergency medicine in a medical simulation center, distinctive scenarios should be used. The mHealth concept in this study was a mobile, high definition, real-time video connection between the emergency site and a remote medical expert. Since all participants in this study confirmed that the chosen scenarios were realistic and relevant, a medical simulation center appears to be a suitable model for testing mHealth concepts in prehospital emergency medicine.**

I. Introduction

NEWLY developed mHealth tools need to be tested in standardized conditions without possible patient harm before implementation. One possible approach to secure these two conditions is to analyze the mHealth tool in a medical simulation center.

As AMMENWERTH and co-workers have explained [1], there are three ways of testing a new health information technology. The first way is to evaluate it in a laboratory, but the results are limited by a low external validity. The second way is a field evaluation test, but for this, both software and hardware have to be sufficiently mature to not possibly harm any person. So the solution is often the third way: a study in a medical simulation center, which combines good internal and external validity [1]. Usability studies in medical simulation centers are ideal for objective, structured analysis of new technical devices [2].

The present article has been structured in the context of the LiveCity ("Live Video-to-Video Supporting Interactive City Infrastructure") European Research Project and has been supported by the Commission of the European Communities - DG CONNECT (FP7-ICT-PSP, Grant Agreement No.297291).

Simulations in general provide the facility to evaluate complex systems and the interaction of the different variables [3], [4]. MHealth concepts in emergency scenarios can be tested in computer generated simulation or in medical simulation centers. It could be shown that the former, for instance in the form of virtual reality, is a successful way of analyzing for instance fire emergencies [5]. This paper aims to analyze how the latter can be used to test mHealth concepts. In medical simulation centers (study) participants encounter realistic routine or emergency scenarios, which are created with the aid of computer-operated mannequins [6]. These scenarios are based on predefined, structured protocols, but are adapted dynamically, depending on the action of the participants. Thus, the participants see and feel the consequences of their individual actions [7]. Because medical simulation combines theoretical knowledge and practical skills, it is widely established, especially in emergency medicine [8], [9], [10]. It is frequently and successfully used in training at all stages of medical education and in research [11], [12]. It can be used for both individual settings and group settings and offers the opportunity to standardize while minimizing negative consequences of potential errors [9], [13].

Simulation studies offer the opportunity to conduct experimental cross-over trials with high internal validity. The external validity depends on how realistic the simulated scenarios are. The perception of how realistic a scenario in a simulation centre is, is influenced by three different aspects: the equipment fidelity, the environment fidelity and the psychological fidelity [14]. The equipment fidelity is characterized by the used hardware and software. The environment fidelity is mostly created by the appropriate surrounding for every scenario. Psychological fidelity is the ability of the individual participant to immerse into the simulated situation. Psychological fidelity can be increased by enhancing equipment and environment fidelity [15].

To evaluate a mHealth concept in the field of prehospital emergency medicine in a medical simulation center, distinctive scenarios should be used. Hence, it is important to generate scenarios of emergencies, which occur often and in

which an early start of the right therapy has a huge impact on morbidity and mortality. Three paramount examples are stroke, myocardial infarction and trauma. These emergencies belong to the "First Hour Quintet", termed by the European Resuscitation Council. It describes five emergencies, which are life-threatening diseases, which require fast treatment [16], [17]. Stroke, myocardial infarction and trauma are also among the main causes of death in Europe [18]. Worldwide, they belonged to the group of top 10 leading causes of death in 2004 and prognosis for 2030 predicts them to be within the top 5 leading causes of death worldwide [19]. Thus, there are many approaches to improve the therapy of these emergencies; among them the application of mHealth. It could be shown, that mHealth for stroke in prehospital emergency medicine is feasible and beneficial [20].

To reflect the broad spectrum of emergencies, it is also important to include emergencies, which are especially challenging. Examples for those emergencies could be rare diseases and difficulties during pregnancy. The treatment of rare diseases often lacks standard operating procedures. Additionally, the emergency personnel might not have encountered a similar situation before, which increases the stress level. The complexity of difficulties during pregnancy is caused by the fact that the unborn child has to be considered, too. For instance, there are only a limited number of pharmaceuticals, for which it could be proven, that they can be administered during pregnancy without teratogenic or negative long-term effects for the unborn child [21]. Additionally, pregnancy changes many physiological parameters in women, which have to be kept in mind.

II. MATERIAL AND METHODS

To test the hypothesis, that medical simulation centers are ideal for testing mHealth concepts in prehospital emergency medicine, ten typical emergency scenarios were prepared from five different categories: "Stroke", "Myocardial infarction", "Trauma", "Rare diseases" and "Difficulties during pregnancy". All scenarios were structured to be handled according to the worldwide used "ABCDE"-approach to rapidly evaluate the emergency situation: "A" for airway, "B" for breathing, "C" for circulation, "D" for disability, "E" for exposure. Relevant details of the case have been included to be recognised by "SAMPLE"-history: "S" for symptoms, "A" for allergies, "M" for medication, "P" for past medical history, "L" for last oral intake and "E" for events leading to the illness/injury [22], [23].

The mHealth concept in this study was a mobile, high definition, real-time video connection between the emergency site and a remote medical expert. For this purpose a camera called LiveCity camera was used, which was developed in the European Union funded research project LiveCity [24], [25], [26].

According to usual guidelines in German emergency medicine two paramedics worked together as a team. Together they handled ten scenarios, five of them with a remote medical expert and five of them without. The sequence of the case scenarios and the assignment to the two cross-over categories was randomized.

When paramedics are alerted by the emergency dispatcher, they are provided with information about the location of the emergency, age and gender of the patient and the operation key word. The operation key word is based on the callers description of the emergency and indicates the kind of emergency [27], [28]. For all 10 scenarios in this study operation key words were developed. In general, every scenario followed this pattern:

The paramedics were presented with the operation key words and age and gender of the patient. After that, they entered the simulation room, where the mannequin was postured with additional props, to illustrate the specific emergency scenario. The paramedics then started to ask the mannequin questions regarding the emergency, took the medical history, measured the vital signs and examined the patient. According to the predefined case description, which was developed specifically for every scenario, questions were answered and the vital signs and findings in the physical examination shown. They were dynamically adjusted according to the actions of the paramedics. In the simulation cases with a remote medical expert, the paramedics could decide at which point during the simulation they wanted to start the video consultation. Depending on features of the emergency case (for instance the severeness and the urgency to start the treatment) and on traits of the paramedic (like level of experience and wish for reassurance) the consultation could start either right after hearing the operation key word or after finishing the diagnostics and initial treatment just for confirmation or at any given moment between that, as would be the case after the implementation of this mHealth system. All the scenarios ended, when the paramedics decided to start the transport to the hospital. That endpoint was chosen, because before the beginning of the transport, all major decisions regarding diagnostics and treatment have to be made. If the paramedics did not start transport within 27 minutes, the scenarios were terminated. This time limit was set, because in Germany in 95% the emergency doctor reaches the emergency site within 26.6 minutes [29]. Thus, in the majority of cases there would be an emergency doctor at the emergency site after 27 minutes to take over from the remote medical expert.

For the simulation in this study, the equipment of the paramedics contained all medical devices, that are statutory for an ambulance car in Germany according to DIN EN 1789 Typ C [30], [31]. These are, for instance, monitor of medical parameters (including body temperature, blood sugar level and 12-lead-ECG) and defibrillator, ventilator machine, medical suction machine, stiff neck and medical bag with

drugs, dressing and devices to secure the airway and blood circulation.

The study was performed in the fully equipped high fidelity medical simulation center of the Department of Anaesthesiology at Greifswald University Medicine. To assess the outcome in practical, technical and psychological aspects, paramedics and doctors were interviewed by use of semi-structured questionnaires.

III. RESULTS

A total of 10 emergency doctors and 21 paramedics took part. Among the many aspects assessed by questionnaires in the LiveCity project, three key findings for this study will be presented. Table I shows how doctors and paramedics rated the statement “The scenarios were realistic." Table II shows how doctors and paramedics rated the statement “The scenarios were relevant." Table III shows how doctors and paramedics rated the statement “I took the simulation work seriously.”

Table I: “The scenarios were realistic.”

	Agree	Partly agree	Partly disagree	Disagree
Doctors (10)	4	5	1	0
Paramedics (21)	9	11	1	0

Table II: “The scenarios were relevant.”

	Agree	Partly agree	Partly disagree	Disagree
Doctors (10)	7	3	0	0
Paramedics (21)	14	7	0	0

Table III: “I took the simulation work seriously.”

	Agree	Partly agree	Partly disagree	Disagree
Doctors (10)	6	4	0	0
Paramedics (21)	18	3	0	0

IV. DISCUSSION

In this study, the majority of emergency doctors and paramedics confirmed that the simulation of the scenarios was realistic. They also acknowledged that the chosen scenarios were relevant examples of emergency situations. All paramedics and emergency doctors agreed or partly agreed that they took the simulation work seriously.

Great emphasis was put on high verifiability, fidelity and validity, which are three of the most important concepts of evaluating the quality of simulation [32]. These three concepts are interconnected. To achieve high verifiability, all 10 developed scenarios were tested and adapted beforehand. As mentioned above, the fidelity can be divided into equipment fidelity, environment fidelity and psychological fidelity. To increase the equipment fidelity in this study, the Laerdal mannequin Resusci Anne® (Laerdal Medical GmbH, Puchheim, Germany) was used and the vital signs were dynamically simulated with the monitor iSimulate ALSi® (Skillqube GmbH, Wiesloch, Germany). The Laerdal Resusci Anne is globally used in education and research [33], [34], [35]. To enhance environment fidelity, every scenario had different characteristic accessories, e.g. in one case of simulated heart attack, a patient was watching sports sitting on a sofa with a football flag while eating potato crisps. Psychological fidelity is the ability of the individual participant to immerse into the simulated situation. Because the psychological fidelity depends on equipment and environment fidelity, huge emphasis has to be on increasing both of the latter. In the concept, that the behaviour in the simulated scenario mirrors the behaviour in a real case, high authenticity is essential [36]. In this study all participants took their work very seriously during the simulation, and the majority rated the simulated scenarios as realistic. Thus, the possibility of the participants behaving in the study environment similar to their normal behaviour is very high. This implies a good external validity. Furthermore, all emergency doctors and paramedics partly agreed or agreed that the chosen scenarios were relevant. This is also an indicator for a good external validity.

V.CONCLUSION

Since all paramedics and emergency doctors confirmed that the chosen scenarios were realistic and relevant, a medical simulation center appears to be a suitable model for testing mHealth concepts in prehospital emergency medicine. It offers the opportunity to evaluate mHealth concept without potential patient harm.

ACKNOWLEDGMENT

The authors would like to thank Professor Michael Wendt, Professor Konrad Meissner and Professor Klaus Hahnenkamp for their continuous support.

REFERENCES

[1] Ammenwerth, E., Hackl, W.O., Binzer, K., Christoffersen, T.E., Jensen, S., Lawton, K., Skjoet, P., and Nohr, C.: ‘Simulation studies for the

evaluation of health information technologies: experiences and results', The HIM journal, 2012, 41, (2), pp. 14-21,

[2] Landman, A. B., Redden, L., Neri, P., Poole, S., Horsky, J., Raja, A. S., Pozner, C. N., Schiff, G., and Poon, E.G.: 'Using a medical simulation center as an electronic health record usability laboratory', Journal of the American Medical Informatics Association : JAMIA, 2014, 21, (3), pp. 558-563, doi:10.1136/amiajnl-2013-002233

[3] Cossentino, M., Lodato, C., Ribino, P., and Seidita, V.: 'A heuristic for Problem Formalization in Agent Based Simulation studies', in Editor (Ed.)^(Eds.): 'Book A heuristic for Problem Formalization in Agent Based Simulation studies' (2015, edn.), pp. 1733-1743, doi:10.15439/2015F287

[4] Preisler, T., Dethlefs, T., and Renz, W.: 'Simulation as a service: A design approach for large-scale energy network simulations', in Editor (Ed.)^(Eds.): 'Book Simulation as a service: A design approach for large-scale energy network simulations' (2015, edn.), pp. 1765-1772, doi:10.15439/2015F116

[5] Kinateder, M., Ronchi, E., Nilsson, D., Kobes, M., M, M., x00Fc, ller, Pauli, P., A, M., x00Fc, and hlberger: 'Virtual reality for fire evacuation research', in Editor (Ed.)^(Eds.): 'Book Virtual reality for fire evacuation research' (2014, edn.), pp. 313-321, doi:10.15439/2014F94

[6] Johannsson, H., Ayida, G., and Sadler, C.: 'Faking it? Simulation in the training of obstetricians and gynaecologists', Current opinion in obstetrics & gynecology, 2005, 17, (6), pp. 557-561,

[7] Gredler, M. E.: 'Games and simulations and their relationships to learning', Handbook of research on educational communications and technology, 2004, 2, pp. 571-581,

[8] Gaba, D.M.: 'The future vision of simulation in healthcare', Simulation in healthcare : journal of the Society for Simulation in Healthcare, 2007, 2, (2), pp. 126-135, doi:10.1097/01.sih.0000258411.38212.32

[9] Kyle, R., and Murray, W. B.: 'Clinical Simulation' (Elsevier Science, 2010. 2010), [10] Lestander, Ö., Lehto, N., and Engström, Å.: 'Nursing students' perceptions of learning after high fidelity simulation: Effects of a Three-step Post-simulation Reflection Model', Nurse Education Today, 2016, 40, pp. 219-224, doi:http://dx.doi.org/10.1016/j.nedt.2016.03.011

[11] Melbye, S., Hotvedt, M., and Bolle, S.: 'Mobile videoconferencing for enhanced emergency medical communication - a shot in the dark or a walk in the park? -- A simulation study', Scandinavian journal of trauma, resuscitation and emergency medicine, 2014, 22, (1), pp. 35,

[12] Cannon-Diehl, M. R.: 'Simulation in healthcare and nursing: state of the science', Critical care nursing quarterly, 2009, 32, (2), pp. 128-136, doi:10.1097/CNQ.0b013e3181a27e0f

[13] Muller-Juge, V., Cullati, S., Blondon, K.S., Hudelson, P., Maitre, F., Vu, N.V., Savoldelli, G.L., and Nendaz, M. R.: 'Interprofessional collaboration between residents and nurses in general internal medicine: a qualitative study on behaviours enhancing teamwork quality', PloS one, 2014, 9, (4), pp. e96160, doi:10.1371/journal.pone.0096160

[14] Fritz, P. Z., Gray, T., and Flanagan, B.: 'Review of mannequin-based high-fidelity simulation in emergency medicine', Emergency Medicine Australasia, 2008, 20, (1), pp. 1-9, doi:10.1111/j.1742-6723.2007.01022.x

[15] Bauman, E. B.: 'Game-based Teaching and Simulation in Nursing and Healthcare' (Springer Publishing Company, 2013. 2013),

[16] Krafft, T., Garcia Castrillo-Riesgo, L., Edwards, S., Fischer, M., Overton, J., Robertson-Steel, I., and Konig, A.: 'European Emergency Data Project (EED Project): EMS data-based health surveillance system', European journal of public health, 2003, 13, (3 Suppl), pp. 85-90,

[17] Nilsen, J. E.: 'Improving quality of care in the Emergency Medical Communication Centres (EMCC) '. Proc. Konferanse for medisinsk nødmeldetjeneste 7. - 8.nov. 2012, Sola, Norway, 8.11.2012 2012

[18] Fischer, M.: 'Factors influencing outcome after prehospital emergencies - a european perspective'. Proc. EUROANESTHESIA 2007, München, Deutschland, 9.6.2007 2007.

[19] WHO: 'Injuries and violence: the facts', in Editor (Ed.)^(Eds.): 'Book Injuries and violence: the facts' (2010, edn.).

[20] Hubert, G., Müller-Barna, P., and Audebert, H.: 'Recent advances in TeleStroke: a systematic review on applications in prehospital management and Stroke Unit treatment or TeleStroke networking in developing countries', International Journal of Stroke, 2014, 9, (8), pp. 968-973, doi:10.1111/ijs.12394

[21] Grzeskowiak, L. E., Gilbert, A. L., and Morrison, J. L.: 'Methodological challenges in using routinely collected health data to investigate long-term effects of medication use during pregnancy', Therapeutic advances in drug safety, 2013, 4, (1), pp. 27-37, doi:10.1177/2042098612470389

[22] Thim, T., Krarup, N. H., Grove, E. L., Rohde, C. V., and Lofgren, B.: 'Initial assessment and treatment with the Airway, Breathing, Circulation, Disability, Exposure (ABCDE) approach', International journal of general medicine, 2012, 5, pp. 117-121, doi:10.2147/ijgm.s28478

[23] Henry, M. C., Stapleton, E. R., and Edgerly, D.: 'EMT Prehospital Care' (Mosby JEMS/Elsevier, 2011. 2011).

[24] Metelmann, B., and Metelmann, C.: 'M-Health in Prehospital Emergency Medicine: Experiences from the EU funded Project LiveCity', in Anastasius, M. (Ed.): 'M-Health Innovations for Patient-Centered Care' (IGI Global, 2016), pp. 197-212, doi:10.4018/978-1-4666-9861-1.ch010

[25] Goncalves, J., Cordeiro, L., Batista, P., and Monteiro, E.: 'LiveCity: A Secure Live Video-to-Video Interactive City Infrastructure', in Iliadis, L., Maglogiannis, I., Papadopoulos, H., Karatzas, K., and Sioutas, S. (Eds.): 'Artificial Intelligence Applications and Innovations' (Springer Berlin Heidelberg, 2012), pp. 260-267, doi:10.1007/978-3-642-33412-2_27

[26] Palma, D., Goncalves, J., Cordeiro, L., Simoes, P., Monteiro, E., Magdalinos, P., and Chochliouros, I.: 'Tutamen: An Integrated Personal Mobile and Adaptable Video Platform for Health and Protection', in Papadopoulos, H., Andreou, A., Iliadis, L., and Maglogiannis, I. (Eds.): 'Artificial Intelligence Applications and Innovations' (Springer Berlin Heidelberg, 2013), pp. 442-451, doi:10.1007/978-3-642-41142-7_45

[27] Ellinger, K.: 'Kursbuch Notfallmedizin: orientiert am bundeseinheitlichen Curriculum Zusatzbezeichnung Notfallmedizin' (Dt. Ärzte-Verlag, 2011. 2011).

[28] Wölfl, C.G., and Matthes, G.: 'Unfallrettung: Einsatztaktik, Technik und Rettungsmittel ; mit 32 Tabellen' (Schattauer, 2010. 2010).

[29] Schmiedel, R., and Behrendt, H.: 'Leistungen des Rettungsdienstes 2008/09', in Editor (Ed.)^(Eds.): 'Book Leistungen des Rettungsdienstes 2008/09' (2011, edn.).

[30] Kemper, H.: 'Fahrzeugkunde: Arten und Ausführungen der genormten Feuerwehrfahrzeuge' (Ecomed Sicherheit, 2010. 2010).

[31] Kühn, C.: 'Patienten- und Personenbeförderung: Zulassungs-, betriebs- und sicherheitstechnische Begutachtung aktueller Fahrzeugumbauten ; Gutachten über die Zulassung und den Betrieb von sogenannten Multifunktionsfahrzeugen (Liegetaxi, Sondermietwagen, Tragestuhlwagen, Selbstfahrermietliegewagen, Funkmietliegewagen, Liegendtransportwagen)' (Hüthig Jehle Rehm, 2008. 2008).

[32] Feinstein, A.H., and Cannon, H.M.: 'Fidelity, verifiability, and validity of simulation: Constructs for evaluation', Developments in Business Simulation and Experiential Learning, 2001, 28.

[33] Yasuda, Y., Kato, Y., Sugimoto, K., Tanaka, S., Tsunoda, N., Kumagawa, D., Toyokuni, Y., Kubota, K., and Inaba, H.: 'Muscles used for chest compression under static and transportation conditions', Prehospital emergency care : official journal of the National Association of EMS Physicians and the National Association of State EMS Directors, 2013, 17, (2), pp. 162-169, doi:10.3109/10903127.2012.749964

[34] Monsieurs, K.G., De Regge, M., Schelfout, S., D'Hondt, F., Mpotos, N., Valcke, M., and Calle, P.A.: 'Efficacy of a self-learning station for basic life support refresher training in a hospital: a randomized controlled trial', European journal of emergency medicine : official journal of the European Society for Emergency Medicine, 2012, 19, (4), pp. 214-219, doi:10.1097/MEJ.0b013e32834af5bf

[35] Abelairas-Gomez, C., Rodriguez-Nunez, A., Casillas-Cabana, M., Romo-Perez, V., and Barcala-Furelos, R.: 'Schoolchildren as life savers: At what age do they become strong enough?', Resuscitation, 2014, 85, (6), pp. 814-819, doi:10.1016/j.resuscitation.2014.03.001

[36] Cumin, D., Weller, J.M., Henderson, K., and Merry, A.F.: 'Standards for simulation in anaesthesia: creating confidence in the tools', British journal of anaesthesia, 2010, 105, (1), pp. 45-51, doi:10.1093/bja/aeq095

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1427–1430
DOI: 10.15439/2016F525
ACSIS, Vol. 8. ISSN 2300-5963

Estimation of blood pressure parameters using ex-Gaussian model

Artur Poliński
Gdansk University of Technology
in Gdańsk
Narutowicza 11/12, 80-233 Gdańsk, Poland
Email: apoli@eti.pg.gda.pl

Tomasz Kocejko
Gdansk University of Technology
in Gdańsk
Narutowicza 11/12, 80-233 Gdańsk, Poland
Email: tomasz.kocejko@pg.gda.pl

***Abstract*—The paper presents an example of model-based estimation of blood pressure parameters (onset, systolic and diastolic pressure) from continuous measurements. First, the signal was low pass filtered and its quality was estimated. Good quality periods were divided into beats using an electrocardiogram. Next, the beginning of each beat of the blood pressure signal was approximated basing on the function created from the sum of two independent distributions: Gaussian and exponential. The nonlinear least square method was used to fit measurement data to the model. The initial conditions for the fitting procedure were selected for each beat on the basis of its parameters. Finally, the diastolic and systolic values of blood pressure and onset were determined.**

I. Introduction

THE BLOOD pressure analysis allows for a certain characterization of the cardiovascular system and thus the patient state. Therefore, it is very important to accurately estimate its characteristic values from noisy data (measurements). The methodology requires that certain steps are performed. First, the signal quality should be estimated in order to skip the fragments of the data which are too noisy. This can be done by using the methods proposed in [1], [2]. Next, the characteristic points of a full beat of blood pressure signal should be estimated. A different method of estimation can be utilized including a windowed and weighted slope sum function [3], a filter bank with variable cutoff frequencies, rank-order nonlinear filters, and decision logic [4], inflection and zero-crossing points of blood pressure, and then combinatorial amplitude and interval criteria to select the onset and systolic peak [5], wavelets [6], principal components [7], waveform descriptor compared with a customized template [8], determined lines and polynomial approximation [9] or Fourier series interpolation [10].

In our studies we have decided to use a model-based approach to obtain signal parameters. In such an approach the approximation results correspond to the selected model. Therefore, it is important that the model has a similar shape to the real signal. A simple solution is to use a polynomial model

This work was partially supported by European Regional Development Fund concerning the project: UDA-POIG.01.03.01-22-139/09-00 -"Home assistance for elders and disabled - DOMESTIC", Innovative Economy 2007-2013, National Cohesion Strategy and by Statutory Funds of Electronics, Telecommunications and Informatics Faculty, Gdansk University of Technology.

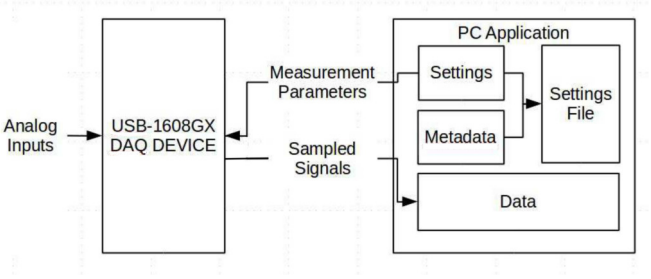


Fig. 1. Communication between DAQ and data receiving software

(like in R-wave estimation [11], which does not guarantee, however, correct results for longer parts of a signal. We have decided to use the more complicated exponentially modified Gaussian function. The exponentially modified Gaussian function was used in different applications like chromatography [12] or cell proliferation and differentiation [13], , but to our knowledge, it has not been used in blood pressure modelling. The practical limitation of such a simple model is that the pressure waveform is a combination of incident and reflected waves. It affects the estimation of systolic pressure. In such an approach, the estimation of model parameters is crucial. Since there is a nonlinear dependence of the model parameters on measurement data, the nonlinear least square approach was used. The model parameters obtained allow to extract diastolic and systolic pressure, but also to compare different pulses using reconstructed parameters. These parameters allow a further data analysis including the dependence between heart rate and blood pressure analysis [14] or estimation of patient condition [15], [16].

II. Material and Method

The continuous blood pressure and ECG were measured using a custom ECG module and CNAP monitor. Both devices were connected to the 16-channel USB Data Acquisition DAQ Module. All data relayed by the DAQ module were recorded by custom software. The block diagram (Fig. 1) presents a general overview of communication between DAQ and the proposed custom recording software.

The rapid measurement and analysis of the ECG signal was enabled by the dedicated ECG optimized pre-amplifier de-

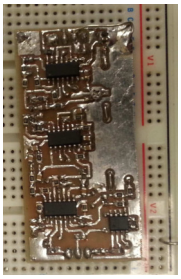


Fig. 2. The prototype of the custom ECG measuring module

signed and manufactured for the purpose of this study (Fig. 2). The overall amplification of the unit was set to approximately 5000. The bandwidth was limited to 0.05 Hz-60 Hz with a dedicated 50 Hz notch filter. A similar solution was used in [17] and [18]. To improve the CMRR, a dedicated DRL circuit was implemented. Blood pressure was measured by means of a standalone CNAP monitor. It enabled the continuous monitoring of blood pressure and pulse rate. All utilized hardware and software came from different vendors which made the integration of the measurements slightly challenging. The CNAP monitor analogue output was connected, next to the custom ECG unit, to the additional PC by means of the DAQ Module. Dedicated software was used for reading the data from the data logger. The measurements were triggered over the network by the UDP protocol. Because ECG and CNAP were registered by means a DAQ data logger there was no delay between measurements. The use of external trigger (over network) allows to extend the measurements by additional parameters like respiration rate, etc. The UDP protocol ensured simple and quick data transfer. Moreover, the UDP allowed for multicast traffic which is very convenient for synchronized measurements of different biosignals.

The data were collected from three healthy men at the age of 28, 33 and 47. The duration of measured signals (including blood pressure, ECG, respiration and eye movement - not used in the present study) was 120 seconds for each person.

The measured signals were sampled with 1 kHz frequency. This relatively high frequency is required to obtain a large number of data for each beat of pressure. This allows to increase the accuracy of the approximation results. First, the continuous blood pressure signal was filtered using a low pass FIR filter (order 1256) with cut off frequency equal to 16 Hz. This made it possible to remove the high frequency noise without modifying the shape of the signal and respiratory influence. The Matlab *filtfilt* function was used, and consequently there was no delay between the raw and filtered data. After blood pressure preprocessing, the R-wave of the electrocardiogram was detected using [19] algorithm. The shape of the blood pressure wave was good enough to skip the above-mentioned quality measure calculations. For each R-wave detected, the minimum of the blood pressure signal was sought. The search was carried out up to 200 ms from the R-wave detected. If there was more than one point of the minimum, than the latest minimum was chosen as a reference (t_{min}). Next, the time of the maximum of the blood pressure was sought at the 400 ms window. If there was more than one point, the latest value was chosen (t_{max}). The last step was choosing the time interval for signal approximation (t_{min}-10 ms, t_{max}+150 ms). Each beat was approximated using the model derived from convolution of two independent additive processes: Gaussian and exponential (ex-Gaussian model)

$$f(t) = C + M\frac{\lambda}{2}\exp\left(\frac{\lambda(2\mu+\lambda\sigma^2-2t)}{2}\right)\times \left(1-erf\left(\frac{\mu+\lambda\sigma^2-t}{\sqrt{2}\sigma}\right)\right) \quad (1)$$

where M is a constant required to stretch the function to the desired range of beat pressure, C is a constant which determines the level of the signal (diastolic blood pressure), μ, σ, λ are parameters describing the properties of the function (μ and σ come from the Gaussian model and λ from the exponential model). The $erf(\cdot)$ is an error function

$$erf(x) = \frac{2}{\pi}\int_0^x e^{-t^2}dt \quad (2)$$

Having t_{min} and t_{max} initial values for fitting were calculated as follows

$$\mu_{init} = t_{max} \quad \sigma_{init} = (t_{max}-t_{min})/2 \quad \lambda_{init} = 10 \quad (3)$$
$$M_{init} = (BP(t_{max}) - BP(t_{min}))/4 \quad C_{init} = BP(t_{min})$$

where BP is the registered continuous blood pressure signal. The *lsqnonlin* Matlab function was used to find the global minimum for the nonlinear least square problem. The systolic and diastolic blood pressure was determined from the minimum and maximum values of each estimated beat. A more complicated issue is onset calculation. In the approach assumed, it was represented by the point of the maximum curvature of the estimated model, where the curvature was determined from

$$curve = \frac{|f"(t)|}{(1+(f'(t))^2)^{3/2}} \quad (4)$$

To allow comparison with other methods, the algorithm was tested on publicly available data. The 037 record from the MIMIC database [20] was used. Since the data were sampled at 125 Hz, the FIR filter was redesigned (order 157). The analysed data were restricted to the beats with SBP<180 mmHg and DBP>20 mmHg. This simple operation was conducted to remove the extremely high and low peaks. The median error for SBP, DBP and onset as well as the RMS error ($RMSE$)

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(data(i)-approximation(i))^2} \quad (5)$$

were calculated for the first 15000 beats. In equation 5, $data$ represents either SBP, DBP or onset obtained from annotations, while $approximation$ represents the same parameters obtained from the algorithm proposed. The $data$ values are integers so $approximation$ values were convert integers as well.

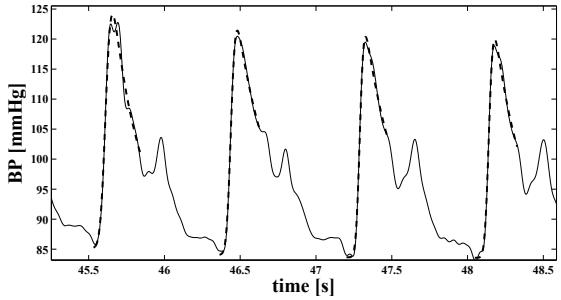


Fig. 3. Filtered signal and its approximation (dotted line)

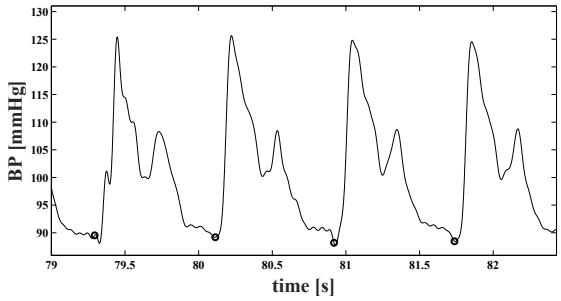


Fig. 4. An example of the onset points (circles)

III. Results

Example of the fitting results (Fig. 3) and founded onset points (Fig. 4) as well as the influence of sampling frequency on DBP and SBP are presented (Figs. 5 - 6).

There were certain difficulties with the analysis of the reference data for 298 of 15000 beats. Modification of one initial data from σ_{init} to $0.6\sigma_{init}$ succeeded in 187 cases, while for the last 111 cases, further modification of one initial data from $0.6\sigma_{init}$ to $0.5\sigma_{init}$ succeeded in 26 cases. Convergence was not obtained for 85 beats. The median error for SBP, DBP and onset was equal to 2 mmHg, 2 mmHg and 0 samples, respectively, while the $RMSE$ error was equal to 4.97 mmHg, 3.23 mmHg and 2.40 samples. Removing 1% of the worst approximation results reduced the $RMSE$ error to 3.08 mmHg, 2.64 mmHg and 2.28 samples, respectively.

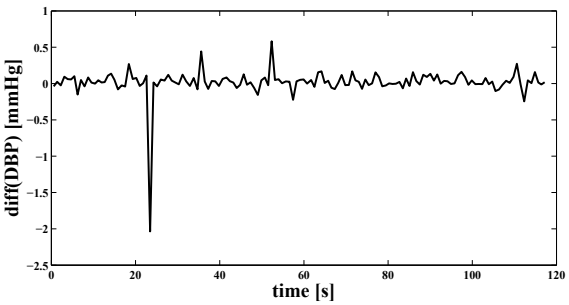


Fig. 5. The difference in estimated diastolic blood pressure (for 1 kHz and 125 Hz sampling frequency)

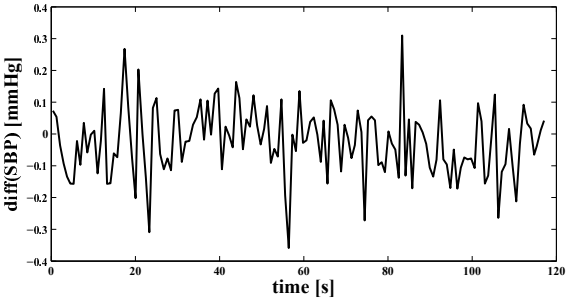


Fig. 6. The difference in estimated systolic blood pressure (for 1 kHz and 125 Hz sampling frequency)

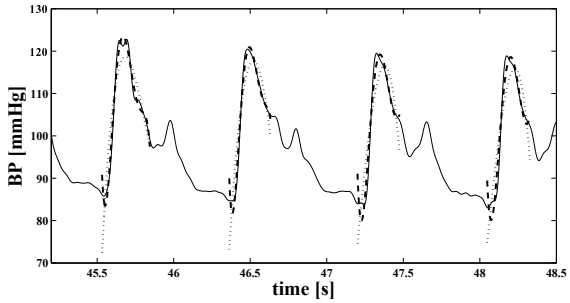


Fig. 7. Filtered signal and its polynomial approximation (dotted line - 3-rd, and dashed line 5-th degree)

IV. Discussion and Conclusions

The results of approximation were highly satisfactory (Fig. 3). In the case of polynomial approximation, the results were much worse. An example of such approximation using 3-rd and 5-th degree polynomial is presented in Fig. 7. The difference between systolic and diastolic blood pressure calculated from low pass filtered data was similar to the approximated one. To reduce the problem of onset detection, the ECG signal was used as a reference. The main problem in nonlinear least square fitting is choosing the starting point for every iteration. Its influence was estimated by perturbing initial conditions. The results of fitting were most sensitive to the μ parameter. Even a change of 5% in the initial value caused poor fitting results. The distortions present in the recorded signal caused problems with the estimation of the fitting error. Thus, the results were compared to the fitting with the initial values. The quality of fitting using the initial values was estimated by visual inspection and a comparison of diastolic and systolic blood pressure with the values calculated for the low pass filtered blood pressure signal (defined as the minimum of blood pressure after R-wave - for the diastolic, and the maximum after DBP for the systolic). The fitting results were much less sensitive to the initial value of σ. Satisfactory results were obtained for values which were as much as ten times smaller and five times greater. Similar results were obtained for the initial λ estimation. The method presented was more

sensitive for the initial values of C and M. Based on the experiment conducted we have assumed that a deviation of 5% is acceptable for C. In the case of parameter M, satisfactory results could be obtained when the value was in the range between 50% and 120% of the initial value.

The nonlinear least squares fitting can use the Jacobian calculated analytically or numerically. We did not notice any significant differences in the results obtained regarding the method of Jacobian calculation. The difference in systolic and diastolic values was lower than 10^{-3} mmHg, while the difference in onset was lower than 0.1 ms.

The sampling frequency did not influence the results obtained. Down sampling of the blood pressure signal from 1 kHz to 125 Hz did not modify the estimation of DBP and SBP significantly (except for one sample) (Figs. 5 and 6).

The diastolic blood pressure can be easily obtained from the model (DBP=C). The analytic estimation of systolic blood pressure requires solving a nonlinear equation. The advantage of the proposed approach is that the calculated parameters: μ, σ, λ can be used to estimate the arteries' condition and their changes.

In our approach the onset estimation bases on finding the maximum curvature of the estimated model curve. To prevent finding the maximum value of the model as the point of maximum curvature, the search was limited to 50 ms after the beginning of each beat. The sampling frequency did not have a large influence on onset detection. The difference between values obtained for 1 kHz and 125 Hz sampling frequency was lower than 7 ms (except for three peaks). The approach adopted for onset detection can be modified by introducing an additional parameter p, which will allow the onset point to be shifted. The curvature will then be defined as

$$\frac{|f"(t)|}{(p+(f'(t))^2)^{3/2}}. \tag{6}$$

In general, the reference ECG signal is not required. The beat localization can be obtained by using only the blood pressure waveform (for example by using the algorithm proposed in [3]), and the proposed model can be then used to obtain signal parameters.

The results obtained from the reference data are satisfactory. Some problems may be due to the relatively low sampling frequency and possible distortions in the analysed signal. Better results can be obtained by further modifying the initial parameters.

The analysis of all the results obtained enables the conclusion that the proposed method of blood pressure parameters estimation using the ex-Gaussian model is suitable to the requirements and reliable.

REFERENCES

[1] W. Zong, G. B. Moody, and R. G. Mark, "Reduction of false arterial blood pressure alarms using signal quality assessment and relationships between the electrocardiogram and arterial blood pressure", Med. Biol. Eng. & Comput., vol. 42, pp. 698–706 September 2004. doi: 10.1007/BF02347553

[2] J. X. Sun, A. T. Reisner, R. G. Mark, "A signal abnormality index for arterial blood pressure waveforms", Computers in Cardiology, vol. 33, pp. 13–16, 2006.

[3] W. Zong, T. Heldt, G. B. Moody, R. G. Mark, "An Open-source Algorithm to Detect Onset of Arterial Blood Pressure Pulses", Computers in Cardiology, vol. 30, pp. 259–262, 2003. doi:10.1109/CIC.2003.1291140

[4] M. Aboy, J. McNames, T. Thong, D. Tsunami, M. S. Ellenby, and B. Goldstein, "An Automatic Beat Detection Algorithm for Pressure Signals", Trans. on Biomed. Eng. , vol. 52, no. 10, pp. 1662–1670 October 2005. doi: 10.1109/TBME.2005.855725

[5] B. N. Li, M. C. Dong, M. I. Vai, "On an automatic delineator for arterial blood pressure waveforms", Biomedical Signal Processing and Control, vol. 5, pp. 76–81, 2010. doi:10.1016/j.bspc.2009.06.002

[6] A. Pachauri and M. Bhuyan, "Wavelet Transform Based Arterial Blood Pressure Waveform Delineator", International Journal of Biology and Biomedical Engineering, Issue 1, vol. 6, pp. 15–25, 2012.

[7] P. Xu, M. Bergsneider, X. Hu, "Pulse onset detection using neighbor pulse-based signal enhancement", Med. Eng. & Phys., vol. 31, pp. 337–345, 2009. doi:10.1016/j.medengphy.2008.06.005

[8] L. Yanga, M. Zhaoa, C. Penga, X. Hub, H. Fengc, Z. Ji, "Waveform descriptor for pulse onset detection of intracranial pressure signal", Med. Eng. & Phys., vol. 34, pp. 179–186, 2012. doi:10.1016/j.medengphy.2011.07.008

[9] E. Kazanavicius, R. Gircys, A. Vrubliauskas, S. Lugin, "Mathematical methods for determining the foot point of the arterial pulse wave and evaluation of proposed methods", Information Technology and Control, vol. 34, no. 1, pp. 29–36, 2005.

[10] A. Fanelli, T. Heldt, "Signal quality quantification and waveform reconstruction of arterial blood pressure recordings", *in Conf. Proc. IEEE Eng. Med. Biol. Soc.* 2014, pp. 2233–6. doi:10.1109/EMBC.2014.6944063

[11] P. Augustyniak, "Recovering The Precise Heart Rate From Sparsely Sampled Electrocardiograms", *in Computers in Medicine*, Łódź 23-25.09.1999, pp. 59-65.

[12] K. Lan, J. W. Jorgenson, "A hybrid of exponential and gaussian functions as a simple model of asymmetric chromatographic peaks", Journal of Chromatography A, vol. 915, pp. 1–13, 2001. doi:10.1016/S0021-9673(01)00594-5

[13] A. Golubev, "Exponentially modified Gaussian (EMG) relevance to distributions related to cell proliferation and differentiation", Journal of Theoretical Biology, vol. 262, pp. 257–266, 2010. doi:10.1016/j.jtbi.2009.10.005

[14] A. Poliński, J. Kot, A. Meresta, "Analysis of correlation between heart rate and blood pressure", Federated Conference on Computer Science and Information Systems (FedCSIS), pp. 417–420, 2011.

[15] M. Kaczmarek, A. Bujnowski, J. Wtorek, A. Poliński, "Multimodal Platform for Continuous Monitoring of the Elderly and Disabled", J. Med. Imaging Health Inform., vol. 2, no. 1, pp. 56–63, March 2012. doi: http://dx.doi.org/10.1166/jmihi.2012.1061

[16] J. Wtorek, A. Bujnowski, J. Rumiński, A. Poliński, M. Kaczmarek, A. Nowakowski, "Assessment of Cardiovascular Risk in Assisted Living", Metrol. Meas. Syst. vol. 19, issue 2, pp. 231–244, May 2012. doi: 10.2478/v10178-012-0020-0

[17] A. Bujnowski, J. Rumiński, P. Przystup, K. Czuszyński,T. Kocejko, "Self Diagnostics Using Smart Glasses—preliminary study", 9th International Conference on Human System Interaction (HSI2016), pp. 511–517, DOI: 10.1109/HSI.2016.7529682

[18] A. Bujnowski, J. Rumiński, M. Kaczmarek, K. Czuszyński, P. Przystup, "Cardiovascular data analysis using electronic wearable eyeglasses—preliminary study", Federated Conference on Computer Science and Information Systems (FedCSIS), pp. 1409–1412, 2016.

[19] J. Pan and W. J. Tompkins, "A Real-Time QRS Detection AlgorithmA Real-Time QRS Detection Algorithm" IEEE Trans. on Biomed. Eng. , vol. 32, no. 3, pp. 230–236 March 1985. doi: 10.1109/TBME.1985.325532

[20] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. Ch. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C-K. Peng, H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals" Circulation 101(23):e215-e220 [Circulation Electronic Pages; http://circ.ahajournals.org/cgi/content/full/101/23/e215]; 2000 (June 13).

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1431–1434
DOI: 10.15439/2016F329
ACSIS, Vol. 8. ISSN 2300-5963

Estimation of respiration rate using an accelerometer and thermal camera in eGlasses

Jacek Ruminski *Member, IEEE*, Adam Bujnowski, Krzysztof Czuszynski, Tomasz Kocejko
Gdansk University of Technology
Gdansk, Poland
jacek.ruminski@pg.gda.pl

***Abstract*— Respiration rate is a very important vital sign. Different methods of respiration rate measurement or estimation have been developed. However, especially interesting are those that enable remote and unobtrusive monitoring. In this study, we investigated the use of smart glasses for the estimation of respiration rate especially useful for indoors applications. Two methods were analyzed. The first one is based on measurements of respiration-related body movements using an accelerometer. The second one uses the thermal camera to observe temperature changes in the nostril region. For both methods signals were extracted, filtered and processed using two different respiration rate estimators. Both methods were validated during experiments with the participation of volunteers using the respiration belt as a reference measurement method. Results proved that for both methods it is possible to reliable estimate the respiration rate with Root Mean Square Error lower than 2 breaths per minute, which is sufficient for medical screening.**

Keywords—smart glasses; respiration rate estimation, thermal imaging

I. Introduction

Smart glasses are wearable devices that can extend human senses and capabilities of information processing. Additionally, the near-to-eye display could provide graphical information with much higher privacy than a smartphone or a tablet. Smart glasses can be equipped with different sensors (e.g. accelerometers, gyroscopes, cameras), communication interfaces (e.g. WiFi, Bluetooth), etc. Recently, many devices have been proposed on the consumer market, including Google Glass, Epson Moverio BT-200, Recon Jet, Lumus DK-40, etc. [1]. Many ideas and demonstrations of potential roles of smart glasses in healthcare have been presented. Some include improved visualization of veins locations (Evena Medical [2]), access and visualization of data from medical records [3][4][5], presentation of vital signs on the display of smart glasses (Philips [6]), etc.

In this paper we analyze the possible role of smart glasses in estimation of respiration rate. Smart glasses were previously used to estimate some vital signs [7][8]. Typically, the pulse rate was estimated using photoplethysmography [9] or respiration rate (and pulse rate) using data collected by an accelerometer or gyroscope of the Google Glass [7]. Additionally, for the identified patient (e.g. using face recognition [10] or using graphical markers like QR-code [11][12]) the measured vital signs or other health-related data can be automatically uploaded to the healthcare information system. The goal of this paper is also to analyze two different respiration rate estimators applied to short-time data collected using the accelerometer or the thermal camera. Respiration rate is typically monitored using masks with thermistors, analyzing ECG drifts or using clinical observations. It is especially important in quality of sleep analysis (e.g. for sleep apnea detection [13]). Some remote methods use analysis of video sequences recorded from the chest region [14] looking for respiration-related movements. Those recordings are typically performed using visible light cameras [15] and infrared cameras [16][17]. Thermal recordings were typically performed using cameras with good spatial resolution [18][19] performing respiratory rate analysis in the frequency domain.

The rest of the paper is structured as follows: Section II presents the proposed methods. Results are described in Section III. Section IV present a discussion of results and concludes the paper.

II. Methods

A. Measurement systems

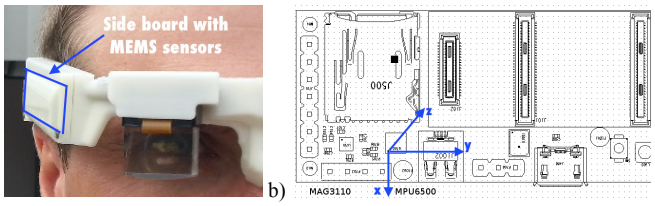
Respiration is important measurement of the body's most fundamental function. Smart glasses can use different sensors to measure or estimate respiration rate (for the observed person and for the wearer). Here, we assume that respiration rate can be estimated using analysis of data captured: with accelerometer/gyroscope sensors - for the user of smart glasses and with the thermal camera - for the observed person. Both measurement methods were implemented in the eGlasses prototype, developed under the eGlasses project (www.eglasses.eu). This experimental platform is dedicated to research activities, so different electronic modules can be changed; it is possible to print another cover using 3D printer, add sensors or electrodes, change the display, etc. The current prototype uses OMAP 4460 processor with 1GB RAM, 1024x768 transparent display from Elvision Company, 5MPx camera, WiFi and Bluetooth 4 wireless interfaces, different sensors (accelerometer, gyroscope, magnetometer, OMRON D6T thermal sensor, etc.), eye-tracker and extension slots. The Android 4.1 OS and Linux Ubuntu OS have been already tested. For the goals of this paper the accelerometer and the thermal camera module were used.

B. Accelerometer and respiration rate

The used single-chip MPU-6500 (Invesense) integrates the 3-axis accelerometer, the 3-axis gyroscope, and the onboard Digital Motion Processor™ (DMP) in a small, 3 mm x 3 mm x 0.9mm package. The device can operate from 1.8V and

This work has been partially supported by NCBiR, FWF, SNSF, ANR and FNR in the framework of the ERA-NET CHIST-ERA II, project *eGLASSES – The interactive eyeglasses for mobile, perceptual computing* and by Statutory Funds of Electronics, Telecommunications and Informatics Faculty, Gdansk University of Technology.

consumes 6.1mW in full operating mode (33µW in low-power mode). The typical offset of the accelerometer is ±60 mg and 300 µg/√Hz of noise. The location of the chip on the base board (Fig. 1) enables the measurement of acceleration in *x* direction (head top to down), *y* direction (back of the head towards face) and *z* direction (ear-to-ear direction).



a) b)

Fig. 1. a) The eGlasses prototype with the indicated location of the (side) base board with sensors. b) The layout of the (side) base board with the indicated axes of the accelerometer and gyroscope.

It is assumed that respiration activities influence body (head) movements so it should be possible to estimate the respiration rate. Typically, the acceleration module is calculated as:

$$|a| = \sqrt{a_x^2 + a_y^2 + a_z^2}\ , \tag{1}$$

where: a_x, a_y, and a_z are the measured acceleration values for particular directions.

In this work, measured signals ($|a|$ or directional a signals) are resampled (to f_s=11Hz) and normalized (mean removal). Next, filtration is used to remove noise and not-respiratory related signal components. The moving average is used (window size $f_s/2$) and the band-pass Butterworth filter (4^{th} order) to pass frequencies between 6bpm and 40bpm. Finally, two respiratory rate estimators are applied (described later).

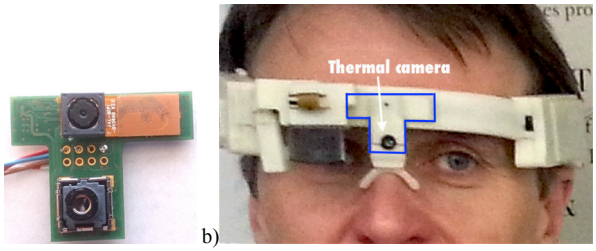
C. Thermal camera and respiration rate

The respiration rate was also analyzed using sequences of thermal images recorded for nostril regions. The FLIR Lepton thermal camera module, located in the designed front board, was used for thermal recordings (Fig. 2). It is characterized by high dynamic range (14bits), small size (<1cm^2) and relatively small spatial resolution (80x60).

The multistep procedure was used for the estimation of respiration rate. First, a sequence of thermal frames was captured during the short time period (30s windows were used in the experiments). Next, in this preliminary study, the nostrils region (a rectangle with width = nose width) was manually selected directly around/below the nose. Then, the nostril ROI was used to calculate the average pixel value inside that ROI. The operation was repeated for each frame producing a 1-D signal (time series) of infrared (thermal) radiation changes. Next the same filtration was used as for signals obtained for the accelerometer (i.e. moving average and Butterworth band-pass filter).

D. Respiration rate estimators

Two respiration rate estimators were analyzed for signals obtained for the accelerometer and for the thermal camera. The first estimator, *eRR_sp,* often used in other studies, identifies the frequency in the frequency domain of the analyzed signal, for the dominating peak in the frequency spectrum.



a) b)

Fig. 2. a) The module with cameras: visible ligth camera (top) and Lepton infrared camera (bottom). b) The location od the module in the eGlasses.

The second estimator, *ePR_ac,* analyzes the periodicity of peaks locations for the autocorrelation function in time domain as a function of time lags. The autocorrelation for different time lags is calculated and the period is determined calculating an average time differences between detected peaks. As a peak detector we used a method looking for a local minimum and a local maximum, for which their difference is grater than the threshold value T:

$$d=s_{fn}(t_{j+1})-s_{fn}(t_j),\ d>T, \tag{2}$$

where: $s_{fn}(t_{j+1})$ - filtered signal value of the local minimum at j, $s_{fn}(t_j)$ - filtered signal value of the local maximum at $j+1$.

The threshold value T was calculated in two phases. In the first phase (T= T_1) as:

$$T_1=T_{K1}*(max(s_{fn}(t))-min(s_{fn}(t))), \tag{3}$$

where T_{K1} was the scaling value set to 0.33.

Then the median of the detected peak-to-peak gradient values was calculated. The scaled (T_{K1}=0.25) median value was used as the threshold value (T=T_2) in the second pass of the algorithm to detect final peaks of each signal.

The calculated frequencies for both estimators were multiplied by 60 to obtain results in breaths per minute (*bpm*, e.g. $ePR\_ac=f_{ac}*60$).

E. Experiments and validation

The analyzed methods for the estimation of respiration rate were validated during experiments with the participation of 11 volunteers (mean age: 39.73y±11.98). Subjects were asked to seat comfortable and not move except natural breathing. Then, separately, data were measured during 1min, by two devices. In parallel, the pressure, chest belt (Vernier RMB) was used as a reference. To synchronize signals subjects were asked to hold the breath at the beginning of data recording. Measured signals were analyzed manually to calculate reference respiration rate (RR) values. According to the RR definition (number of respiration events in time) the complete periods between successive inspirations were indicated and counted in 30s long time windows. The RR was calculated as:

$$RR=(N_{RR}*60)/(t_{le}-t_{fs}), \tag{4}$$

where: N_{RR} – number of respiration events (inspiration to inspiration), t_{le} – time of the end of the last event, t_{fs} - time of the start of the first event.

III. Results

A. Accelerometer and respiration rate

Table I presents results of the respiration rate estimation for measurements performed using the accelerometer of smart glasses and using the reference respiration belt.

TABLE I. RESULTS OF RESPIRATION RATE ESTIMATION FOR THE ACCELEROMETER

	Chest Belt			Accelerometer	
Subject	*RR [bpm]*	*eRR_sp*	*eRR_ac*	*eRR_sp*	*eRR_ac*
S01	15.400	14.650	14.990	10.950	14.160
S02	21.400	21.970	21.650	20.480	20.160
S03	10.000	10.254	9.900	9.580	9.470
S04	15.670	16.110	15.740	8.020	11.160
S05	20.000	20.510	20.000	17.977	18.310
S06	15.000	13.184	13.470	13.840	13.760
S07	14.060	13.180	13.920	16.450	15.600
S08	13.240	13.180	13.370	13.450	13.370
S09	11.900	11.720	12.040	9.350	10.680
S10	13.640	13.180	13.890	13.610	13.030
S11	13.640	13.180	13.590	12.520	12.840

The values of RMSE for particular estimators are: for the belt: *RMSE(eRR_sp)*=0.73, *RMSE(eRR_ac)*=0.50; for the accelerometer: *RMSE(eRR_sp)* =2.99, *RMSE(eRR_ac)*=1.73. Signal examples are presented in Fig. 3: measured signals by the accelerometer (subject S05) and obtained results (spectrum for the filtered signal, and autocorrelation as a function of time lags with the detected peaks).

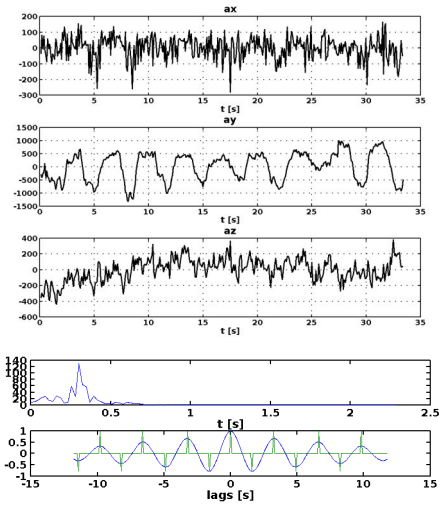


Fig. 3. Top: Signals from the accelerometer for subject S05. Bottom: Spectrum of the filterred signal and autocorrelation signal as a function of time lags with indicated detected peaks.

Table II presents results of the respiration rate estimation for measurements performed using the thermal camera of smart glasses and using the reference respiration belt.

The values of RMSE for particular estimators are: for the belt: *RMSE(eRR_sp)*=0.55, *RMSE(eRR_ac)*=0.24; for the thermal camera: *RMSE(eRR_sp)* =0.68, *RMSE(eRR_ac)*=0.69.

TABLE II. RESULTS OF RESPIRATION RATE ESTIMATION FOR THE THERMAL CAMERA

	Chest Belt			Thermal camera	
Subject	*RR [bpm]*	*eRR_sp*	*eRR_ac*	*eRR_sp*	*eRR_ac*
S01	17.900	18.281	17.863	18.281	18.140
S02	13.000	12.188	12.893	13.711	13.448
S03	12.188	12.188	12.480	12.188	12.581
S04	18.200	18.281	18.425	18.281	18.571
S05	20.955	21.328	20.526	21.328	22.609
S06	9.600	10.664	9.936	10.664	9.936
S07	7.200	7.620	7.430	7.610	7.090
S08	19.809	19.805	19.873	18.281	18.699
S09	19.158	18.281	18.828	19.805	19.141
S10	19.711	19.805	19.623	19.805	20.526
S11	13.220	13.711	13.220	13.711	13.200

In Fig. 4a some examples of thermal images captured during inspiration and expiration events are presented. In Fig. 4b examples of the filtered signal of thermal radiation changes in the nostril ROI (subject S01) and obtained results (spectrum for the filtered signal, and autocorrelation as a function of time lags with the detected peaks) are presented.

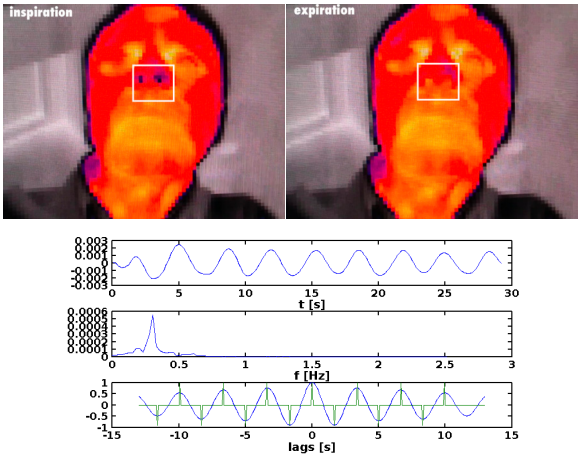


Fig. 4. Top: Thermal images for inspiration and expiration events. Bottom: Filtered signal, spectrum of the filterred signal and autocorrelation signal as a function of time lags with indicated detected peaks.

IV. Discussion and Conclusions

In this study we proposed the use of smart glasses for the estimation of respiration rate using two sensors. The first method was based on measurements performed with the accelerometer. As it could be observed in Fig. 3a respiration-related movements are mainly (for most participants) observed in *y* direction (head backward/forward). This is probably related to head movement during inspiration, when the chest cage/lungs are filled with air causing natural head movement to the back. The calculated *RMSE* value for the *eRR_ac* estimator shows that the respiration rate can be estimated with relatively good accuracy. This is usually good enough for medical screening purposes. For most subjects the accuracy of the respiratory rate was up to 2bpm. However, as it could be observed for S04, in some cases the result is not acceptable. It is important to underline, that the value of respiration rate is always calculated for a particular time window. So manually

calculated RR values for the belt-based signals are a little bit different that values automatically calculated by two estimators. This is also caused by analyzing different time windows. The manually calculate RR values considered only the whole inspiration-to-inspiration periods. Other estimators used all samples. Additionally, the estimator based on the frequency domain has limited accuracy due to the finite frequency resolution. For example, taking *N=330* samples and *fs=11Hz*, then spectral resolution is (11/330)*60=2bpm. For experiments with the use of the accelerometer the *eRR_sp* gave worse results than the *eRR_ac* estimator performing operations in time domain. The experiments were performed under optimistic conditions – subjects did not move. Further research is required to verify the methodology in more natural conditions (head movements).

Experiments performed for the small thermal camera module gave very good results. The respiration rate can be easily and reliably estimated using the described method. The method has of course limits related to required gradient between values of body temperature and ambient temperature. However, when used indoors the required temperature gradient is practically usually fulfilled (e.g. air-conditioning). One of the most interesting finding is that even such a small spatial resolution is not a problem for locating nostril regions and obtaining high dynamics of signals (ambient temperature of the laboratory room when experiment took place was high, 24-27C). In this work we proposed the use of very small thermal camera with small spatial resolution that is used for smart glasses. We also compared two different estimators: one operating in frequency domain, the second one in time domain. Results proved high accuracy of the method and both estimators. However, in this study we used manually selected ROI for nostril area. This can be performed automatically using automatic face detection for thermal images and detection of nostril regions [20]. This will be a subject for further study. It is also important to investigate different signal quality measures that can be used to evaluate if the measured signal (or RR value) is more or less reliable. Such possible measures can analyze the spectral purity of the periodical components (e.g. Hijorth parameters [21], etc.) or other similar parameters (goodness of fit for parametric models [22], periodicity of the autocorrelation function, etc.).

Using smart glasses senses of a healthcare professional could be extended providing additional knowledge about a patient acquired during routine interviews. This is more natural method of observation because it does not influence the patient by using (wearing) additional equipment. Additionally, smart glasses can provide very good source of medical data for self-diagnostics of the smart glasses user. However, practical application of such methods requires comfortable design of smart glasses, which is a task for further research.

REFERENCES

[1] Smart Glasses Portal, 2016, Available: http://www.smartglassesnews.org.

[2] Evena Medical, 2014, Available: http://evenamed.com/~even5672/products/glasses.

[3] C. Borchers, “Google Glass moves into the hospital at Beth Israel”, the Boston Globe, 14.06.2014, Available: http://www.bostonglobe.com

[6] J. Ruminski, A. Bujnowski, T. Kocejko, A. Andrushevich, M. Biallas, R. Kistler, “The data exchange between smart glasses and healthcare information systems using the HL7 FHIR standard”, 9th International Conference on Human System Interaction, IEEE, eXplore, 2016.

[6] J. Ruminski, K. Czuszynski, “Application of smart glasses for fast and automatic color correction in health care”, Proc. of the 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Milan, Italy, 2015.

[6] Philips, “Delivering vital patient data via Google Glass”, 03.10.2013, Available: http://www.healthcare.philips.com/main/about/future-of-healthcare/

[7] J. Hernandez, Y. Li, J. M. Rehg and R. W. Picard, "BioGlass: Physiological parameter estimation using a head-mounted wearable device," Wireless Mobile Communication and Healthcare (Mobihealth), 2014 EAI 4th International Conference on, Athens, 2014, pp. 55-58.

[8] J. Ruminski, “The accuracy of pulse rate estimation from the sequence of face images”, 9th International Conference on Human System Interaction, IEEE, eXplore, 2016.

[9] M. Z. Poh, D. J. McDuff, R. W. Picard "Non-contact, automated cardiac pulse measurements using video imaging and blind source separation", Opt. Expr., vol. 18, pp.10762 -10774, 2010.

[10] J. Ruminski, M. Smiatacz, A. Bujnowski, A. Andrushevich, M. Biallas, R. Kistler, "Interactions with recognized patients using smart glasses," in Human System Interactions (HSI), 2015 8th International Conference on, pp.187-194, 25-27 June 2015

[11] K. Czuszynski, J. Ruminski, “Interaction with medical data using QR-codes”, in Human System Interactions (HSI), 2014 7th International Conference on, pp.182-187, 16-18 June 2014.

[12] A. Kwasniewska, J. Klimiuk-Myszk, J. Ruminski, J. Forrier, B. Martin, I. Pecci, "Quality of graphical markers for the needs of eyewear devices," in Human System Interactions (HSI), 2015 8th International Conference on , vol., no., pp.388-395, 25-27 June 2015.

[13] P. Przystup, A. Bujnowski, J. Ruminski, J. Wtorek, “A multisensor detector of a sleep apnea for using at home”, The 6th HSI Conference, IEEE Xplore, pp. 513-517, 2013.

[14] F. Zhao, M. Li, Y. Qian, J. Z. Tsien, “Remote Measurements of Heart and Respiration Rates for Telemedicine”, PLoS One. 2013; 8(10).

[15] M. Z. Poh, D. J. McDuff, P. W. Picard, “Advancements in noncontact, multiparameter physiological measurements using a webcam”. IEEE Trans Biomed Eng 58: 7–11, 2011.

[16] A. K. Abbas, K. Heimann, K. Jergus, T. Orlikowsky, S. Leonhardt, “Neonatal non-contact respiratory monitoring based on real-time infrared thermography”, BioMedical Engineering OnLine, vol. 10:93, BioMed Central, 2011.

[17] E. A. Bernal, L.K. Mestha, E. Shilla, "Non contact monitoring of respiratory function via depth sensing," in Biomedical and Health Informatics (BHI), 2014 IEEE-EMBS International Conference on, pp.101-104, 1-4 June 2014.

[18] R. Murthy, I. Pavlidis , “Non-contact monitoring of respiratory function using infrared imaging,” IEEE Engineering in Medicine and Biology Magazine. vol.25, pp.57-57, 2006.

[19] J. Ruminski, “Evaluation of the respiration rate and pattern using a portable thermal camera”, Proc. of the 13th Quantitative Infrared Thermography Conference, Gdansk 2016.

[20] A. Kwasniewska, J. Ruminski, „Real-time facial feature tracking in poor quality thermal imagery”, Proc. of the 9th International Conference on Human System Interaction, IEEE, eXplore, 2016.

[21] L. Sörnmo, P. Laguna, “Bioelectrical Signal Processing in Cardiac and Neurological Applications”, Academic Press, 2005.

[22] J. Ruminski, B. Bobek-Billewicz, “Parametric imaging in dynamic susceptibility contrast MRI-phantom and in vivo studies,” IEMBS'04. 26th Annual International Conference of the IEEE, pp. 1104-1107, 2004.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1435–1443
DOI: 10.15439/2016F132
ACSIS, Vol. 8. ISSN 2300-5963

SARF: Smart Activity Recognition Framework in Ambient Assisted Living

Samaneh Zolfaghari, Mohammad Reza Keyvanpour
Alzahra University
Tehran, Iran
Email: s.zolfaghari.ir@ieee.org

***Abstract*—Human activity recognition in Ambient Assisted Living (AAL) is an important application in health care systems and allows us to track regular activities or even predict these activities in order to monitor healthcare and find changes in patterns and lifestyles. A review of the literature reveals various approaches to discovering and recognizing human activities. The presence of a vast number of activity recognition issues and approaches has made it difficult to make adequate comparisons and accurate assessment. Introducing the five basic components of activity recognition in the smart homes as a famous environment to remote monitoring of patients and independent living for elderly, the present paper proposes SARF framework to classify each of activity recognition approaches and then it is evaluated based on the proposed classification by some proposed measures. Using SARF proposed framework can play an effective role in selecting the appropriate method for human activity recognition in smart homes and beneficial in analysis and evaluation of different methods for various challenges in this field.**

I. Introduction

IN RECENT years automatic human activity recognition has received considerable attention due to the growing demand in many applications such as healthcare systems for monitoring the Activities of Daily Living (ADL) in smart homes, especially due to the rapid growth of elderly population, in surveillance and security environments to automatic detection of abnormal activities to alert the relevant authorities about the potential criminal or terrorist behavior, in activity-aware services to convert ideas like smart meeting rooms, home automation, personal digital assistants from science fiction to everyday fact and in entertainment environments to improve human interaction with computers [1][2][3].

Due to the many uses of activity recognition in smart homes and the availability of various approaches in this field, comparison and accurate evaluation of existing methods is difficult. Therefore, providing an account of these activity recognition approaches seems to be essential. The main contribution of this paper, after briefly introducing five basic components of human activity recognition in smart homes, is proposing SARF framework to classify different methods in this field. Then, this framework is analyzed in terms of approaches, their characteristics, challenges and also proposed measures.

The remainder of this paper is organized as follows: In Section II, basic definition for human activity recognition and its capabilities in healthcare systems will be introduced. In Section III, the overall structure of human activity recognition process in smart homes will be described in form of five basic components. In Section IV is represented the proposed SARF framework according to various activity recognition approaches and in Section V the proposed classification based on proposed measures will be evaluated.

II. Human Activity Recognition in Ambient Assisted Living

Nowadays learning and understanding the observed activity [2][4] and event mining [5][6] are central to many fields of studies. The activities of an individual affect him/her, the people around him/her, society and environment [1]. Activities refer to complex behaviors consisting of a sequence of actions and/or overlapped and interwoven actions that can be performed by a single individual or several individuals interacting with each other [1][4]. Activity recognition in healthcare systems considered as a way to facilitate the work of healthcare in order to treat and care for patients, reduce the workload of medical staff, decrease hospital stays for patients, reduce costs and improve the quality of life for people who need care [1][2]. Medical experts believe one of the best ways to identify and explore emerging medical conditions is to monitor changes in daily activities, before these conditions become serious [7].

Recently human-activity discovery [8], recognition [9], prediction [10], and abnormalities detection [11], have attracted great interest because of their high potential in context-aware computing systems such as smart environments. Activity recognition in smart homes has made it possible to track occurrences of regular activities in order to monitor healthcare and find changes in activity patterns and lifestyles, so can be a great help in providing automation, security and most importantly remote health monitoring for elderly or people with disabilities [7][8].

Thus, in recent years activity recognition has become one of the application areas in healthcare systems such as AAL and is leading important research activities including CareLab, CASAS, Gator-Tech, HIS, Aware Home, SELF, iDorm, MavHom [12].

In this study, a comprehensive classification and evaluation of human activity recognition techniques in smart homes as an AAL system is introduced which tries to cover all existing approaches.

978-83-60810-90-3/$25.00

III. Basic Components in Human Activity Recognition Process

The process of human activity recognition follows five steps including Sensing, Preprocessing, Feature Extraction, Feature Selection and Activity Learning Techniques [1][13]. Fig. 1 represents basic components of human activity recognition. Note that, depending on environmental conditions, the types of sensors used and the type of data collected, some of these steps may not be needed. Each of these steps will investigate in the following sections.

A. Sensing

In the first step sensing is performed by the sensors and the data are collected in a database [4]. In fact, this step is responsible for collecting sensor data from smart home environment [13]. The data is sent as a signal to perform preprocessing. Signals contain information about the object which is observed and measured [1] and can be numeric, time, multimedia or even quality signals.

In order to monitor human activities in smart homes wide variety of sensors have been used and there are different perspectives to sensors classification. The sensor classification from two general perspectives is also shown in Fig. 1.

The discrete sensors including Passive Infra-Red (PIR), Contact Switch Sensors (CSS) and Radio-Frequency Identification (RFID) have binary output. Due to simplicity and unobtrusiveness nature of captured data from detected objects or residents states, they are very popular. Opposite side of discrete sensors are continuous sensors including Physiological, Ambient and Multimedia sensors with simple or complex data streams such as real numbers, images or voices [1][3][14].

In one point of view, sensors are wearable or environmental. The wearable sensors including Inertial (e.g. Accelerometers and Gyroscopes) and Vital Signs sensors (e.g. Bio-sensors) [3]. Individuals use wearable sensors to generate more information about posture, motion, location and people interaction [15]. Environmental sensors are used to capture data about smart home environment such as temperature, humidity, light, pressure, noise, and etc. [14]. They are not customized for a single resident; therefore, they can be used to group activity monitoring but they cannot discriminate between residents motions or actions [1]. The example of gathered sensor data which has a binary output shown in Fig. 2 generated by the CASAS data collection system automatically.

B. Preprocessing

The aim of preprocessing is to reveal information on signal, noise reduction and to remove excess information [3]. Cleaning, completing and normalizing data are the basic tasks in preprocessing including particle filters, median filters, kalman filter, low-pass filter and discrete wavelet package shrinkage and etc. to noise reduction. Also, linear and nearest neighbour and cubic interpolation using to fill in the missing values [3][16]. Because of the continuous flow of sensor-based information, it should be divided into segments to be easily recognizable by a trained classifier [3][17].

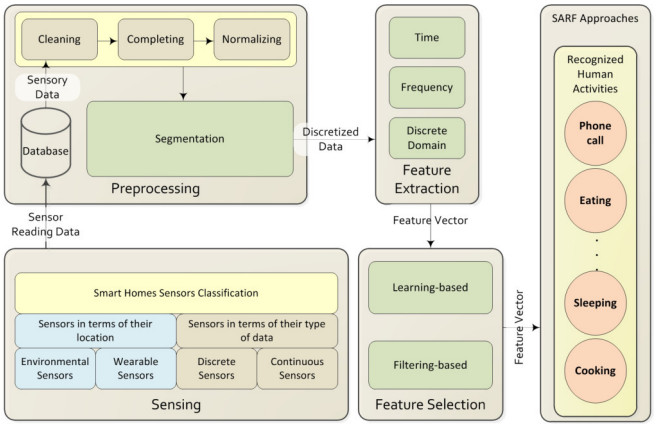


Fig. 1. Basic components of activity recognition process

Various approaches can be used to address segmentation of sensor events for activity recognition such as Change Point Detection (CPD), Time Slice based Windowing (TSW) and Sensor Event based Windowing (SEW) [1][17]. The CPD is an unsupervised segmentation and the idea is to find sudden changes in time series and recognize similar activity borders in real time [1]. The TSW is segment readings provided by inertial sensors and widely used in physical activity recognition. The SEW contains the same number of sensor events and segments the streaming data into sub-sequences [17]. Fig. 3 represents the schema of TSWs and SEWs segmentation.

In some cases (i.e. using supervised learning) at this step data annotation is done [13]. Accurate annotation of activities is important for performance evaluation of recognition models [9]. Annotation methods are divided in to two categories: Offline and Online methods. In Table I characteristics of different approaches in data annotation are represented. The output of this step as discretized data will be sent to Feature extraction step.

C. Feature Extraction

At this step the discretized data is considered as input and the feature vector as output. The purpose of this step is to select and maintain features that contribute to activity recognition. Depending on the kind of data, this step can vary [11]. The most commonly used approaches in this area

2009-10-16	08:43:59.000024	M008	ON	Watch TV begin
2009-10-16	08:44:00.000043	M026	ON	
2009-10-16	08:44:01.000095	M026	OFF	
2009-10-16	08:44:02.000079	M008	OFF	
2009-10-16	08:44:13.000093	M026	ON	
2009-10-16	08:44:17.000043	M026	OFF	
2009-10-16	08:44:24	M026	ON	
2009-10-16	08:44:26.000088	M008	ON	
2009-10-16	08:44:28.000077	M026	OFF	
2009-10-16	08:44:29.000026	M008	OFF	Watch TV end

Fig. 2. Raw data from discrete sensors

TABLE I
COMPARISON OF DIFFERENT ANNOTATION APPROACHES

Annotation Approach		Description	Advantages	Disadvantages
Off-line	Minimum Intervention	Inferences are done by using cameras, video data or recorded voices.	High Accuracy No need to user annotation	Time consuming and computationally expensive Based on resident tracking before data analysis Lack of scalability in resident and activity increasing Lack of privacy preserving
	Indirect Observation	Utilizing self-inference and sensor activation visualization by location, time and sensor location. Annotation has been done by residents and supervisors or just residents. Then these annotated data will store in a database.	High Accuracy No need to user annotation	Time consuming and computationally expensive Based on resident tracking before data analysis Lack of scalability in resident and activity increasing Lack of privacy preserving
Online	Experience Sampling	Utilizing self-report such as record activity information on paper or PDAs. This method is based on periodic alarm in resident environment to do annotation.	Reduce errors Fast Easy to use Better in convergence	Make one-sided or unrealistic data Make interruptions in residents activities Useless in a smart homes with elderly residents with dementia disease
	Direct Observation	In this method supervisor determine specific activities which have to be done by residents so the right activity label even before performing activities are clear.	Accurate annotation	Time consuming
	Time Diary	Use topic models such as LDA in order to provide brief description from activities in data, automatically.	Specify brief description of the activities in data, automatically No need to user annotation	Need to large volume of data Word order does not matter

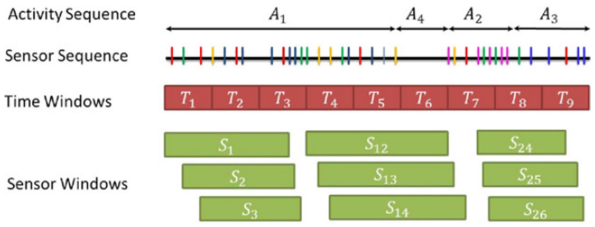


Fig. 3. Illustration of TSW and SEW approaches in Preprocessing step [18]

operate in three fields: time (e.g. Mean, Median, and Standard Deviation etc.), frequency (e.g. Wavelet Transformation and Fourier Transform) and discrete domain (e.g. Euclidean-based Distances and Dynamic Time Warping etc.) [3][15].

Actually, there is no general rule for feature extraction and it depends on the type of problem, our understanding of the problem etc. Thus, it can be done in different ways by different characteristics consideration.

Generally, sensor data features can classify into four groups: Features describing characteristics of the sensor event sequence, Features describing characteristics of discrete sensor values, Features describing characteristics of continuous sensor values, and Activity context [1].

D. Feature Selection

The purpose of this phase is to increase the accuracy of the resulting model by selecting more discriminative features. Also, to provide more robust model, reducing the dimensionality of feature vector and removing features with noise or features with irrelevant information are effective.

It should be noted, additional features will increase computational complexity and classification errors [3][13][19]. There are different approaches to feature selection in human activity recognition approaches including Learning-based and Filtering-based methods.

The Learning-based methods such as Simulated Annealing, Best First Search [1], or Genetic Algorithms [19]interact with the classifier to optimize the feature subset but makes classifier selection become an important process [19]. The idea behind the Learning-based methods is shown in Fig. 4. In the Filtering-based methods such as Minimum Redundancy-Maximum Relevance, the basic idea is not using features which are highly correlated among themselves [13]. Information Gain based on entropy ranks and weights each feature based on its ability to separate the activity instances of different classes [20]. Also Principle Component Analysis [21]

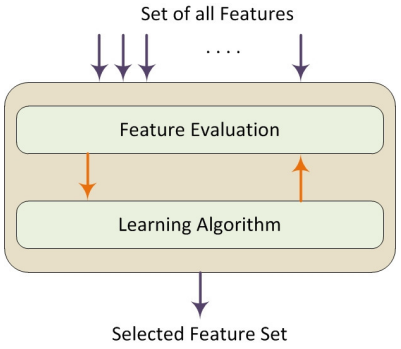


Fig. 4. The Learning-based approach to feature subset selection

TABLE II
COMPARISON OF DIFFERENT FEATURE SELECTION APPROACHES

Feature Selection Approaches	**Method Example**	**Advantages**	**Disadvantages**
Filtering-based Methods	Minimum Redundancy-Maximum Relevance, Information Gain based on Entropy, Principle Component Analysis	Fast Scalable Acceptable computational complexity Independent from classifier	No interaction with classifier Ignore effect of selected feature on classifier Ignore correlation between features Lack of appropriate criteria to specify number of required features
Learning-based Methods	Simulated Annealing, Best First Search, Genetic Algorithms	Choose simple features with low computation Interaction with classifier Consider correlation between features	Dependent to type of classifier Time-consuming in high dimension Suffer from over-fitting

is a linear technique and depends on data scaling. In this method principal components are not always easy to interpret [22]. In fact filter methods are fast, scalable and provide good computational complexity but they ignore interaction with the classifiers [19]. Table II is represented properties of different feature selections approaches in human activity recognition in smart homes.

E. Activity Learning Techniques

In this step machine learning methods are applied for learning activity using selected features [1]. Most smart homes activity recognition studies focus on the Katz index which is usually used in healthcare to evaluate the dependence level, physical and cognitive abilities of elderly people [9]. Generally, new algorithms that correlate the sensor firings, activity labels and predict activities from new sensor firings are required to identify activities from sensor activations alone [23].

A proposed general classification of different methods will address in the following section which tries to cover all existing approaches in human activity recognition in smart homes.

IV. SARF: SMART ACTIVITY RECOGNITION FRAMEWORK IN SMART HOMES

As mentioned before, when the problem of activity recognition in smart home arises, we track occurrences of regular activities in order to monitor health care and find changes in patterns and individuals lifestyle [8]. Since there are different approaches to activity recognition in related areas, presenting a general classification and examining each approach according to the applications and existing challenges seems necessary. Several categories have been presented to classify these approaches and a well-known classification is presented in [4]. This classification must be updated with new concepts and represent new challenges and future work which should be taken into consideration. This work is done by SARF framework.

In our viewpoint, human activity recognition methods can be categorized into three approaches including Bottom-Up, Top-Down and Hybrid approaches which are summarized in Fig. 5. Each of these approaches considers activity recognition intelligible from different perspectives. In this section, the SARF proposed framework will be analyzed.

A. Bottom-Up Approaches

In Bottom-Up activity recognition methods, a learning activity model uses a large collection of user behavior data obtained by the sensor through data mining and machine learning techniques and try to recognize performed activities [24]. These methods can be divided into three categories: Probability-based, Similarity-based and Integration-based methods.

1) Probability-based Methods: These methods improve the generalization ability by modeling the underlying distribution of classes from the obtained feature space [25]. These methods are flexible, since they learn the structure and relationship between the classes by exploiting prior knowledge for a given task such as Markov assumptions, prior distributions and probabilistic reasoning, although the parameters are not optimized [4][26].

An example of a Probability-based approach is to use Nave Bayes [23] classifier that estimates the parameters distribution based on the independence assumption. Let I_{js} which is an activity instances is assigned to the class A_s for which it has maximum posterior probability given by (1) in accordance Bayes Theorem. Each I_{js} observed by R sensors and represented by feature set $F_{js} = \{f_{js}^r\}_{r=1}^R$

$$p(A_s|I_{js}) > p(A_m|I_{js}) \; \forall m.s.t.1 \geq m \leq S, \quad s \neq j \qquad (1)$$

The classifier resulting from the assumption mentioned before is known as the Nave Bayes classifier given by (2).

$$p(A_s|I_{js}) = \prod_{r=1}^{R} p\left(f_{js}^r|A_s\right) \qquad (2)$$

Where $p(A_s|I_{js})$ is the product of the values of features $\{f_{js}^r\}_{r=1}^R$ of an activity instance I_{js} for a given class A_s [27].

2) Similarity-based Methods: The Similarity-based approaches when training data size is large enough, lead to higher efficiency in generalization [25]. However, these methods may suffer from over-fitting, thus making recognition models inconsistent [26]. In these methods, it is important to define the similarity measurement in order to perform patterns selection. Many approaches have been proposed to calculate the distance between different sequences, and one of the most commonly used methods is the edit distance [17].

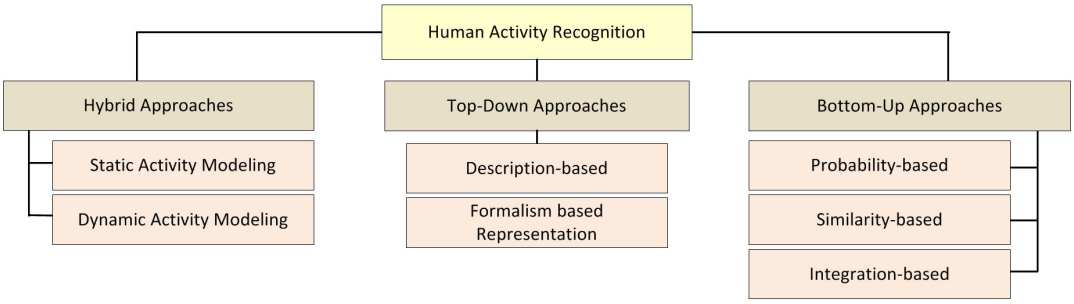


Fig. 5. The SARF proposed framework to analyze various approaches to human activity recognition

Accordingly, the similarity function between two patterns (X,Y) is defined as in (3).

$$Similarity(X,Y) = 1 - \left(\frac{e(X,Y)}{\max(|X|,|Y|)}\right) \tag{3}$$

Where $e(X,Y)$ is the number of edits required to transform an event sequence X into event sequence Y [8].

3) Integration-based Methods: Classification performance and accuracy can often be improved by combining multiple models together, instead of using a single model [28]. This is the basic idea of introducing integration based methods.

In some studies ensembles models are used in human activity recognition in smart homes such as what is done in [29]. Hence, a combination of the models, such as a voting strategy, a simple average among the models [1] and Genetic Algorithm [30] used to combine fusion weight selection of classifier within ensembles, which will decide the winning label for a particular data point and optimizing the output of multiple classifiers.

On the other hand, some studies proposed an activity recognition approach that integrates Probability-based with Similarity-based methods. For example, Fahad and Rajarajan in [28] to improve the reliability of recognitions, integrates the distance minimization and probability estimation approaches. Fig. 6 represents the Block diagram of the proposed activity recognition approach in [28].

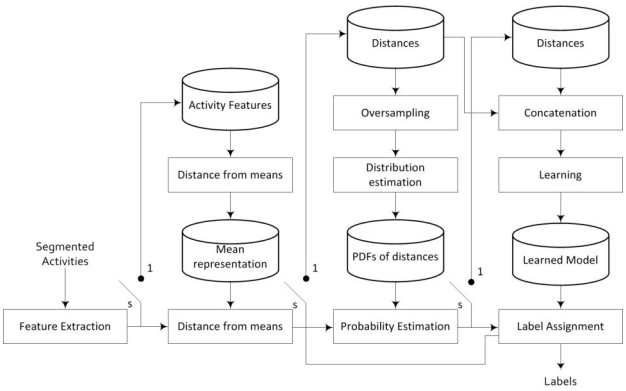


Fig. 6. Block diagram of the proposed activity recognition approach in [28]. Switch s = 1 is training.

B. Top-Down Approaches

In Top-Down activity recognition approaches activity models exploit rich prior knowledge to construct activity models directly using knowledge engineering and management technologies. This usually involves knowledge acquisition, formal modeling, and representation [4]. These methods can be divided into two categories: Description-based Activity Modeling and Formalism-based Representation Methods.

1) Description-based Activity Modeling: The Description-based activity modeling represented activity as an object and models activities as a hierarchy of classes where each class can be described by a number of properties so these approaches including a set of representational concepts [4][31]. The generated activity models are able to capture built-in interrelations between objects and activities such as proposed method in [32]. For example, to detect activity "clean up" which is a complex activity, recognition in the form of the simpler components carried out and the following axioms represented in (4) and (5) are added to the knowledge base [31].

$$CLEANUP \sqsubseteq COMPLEXACTIVITY \sqcap \forall HASACTOR.(PERSON \sqcap \exists HASSIMPLEACTIVITY.PUTINDISHWASHER) \tag{4}$$

$$CLEANUP \sqsubseteq COMPLEXACTIVITY \sqcap \forall HASACTOR.(PERSON \sqcap \exists HASSIMPLEACTIVITY.CLEANTABLE) \tag{5}$$

2) Formalism-based Representation Methods: These methods views an activity as a knowledge model that can be formally specified using various logical formalisms. Activity models generated in these methods are normally used for activity recognition or prediction through formal logical reasoning, e.g., deduction, induction, or abduction [4].

Bouchard, Giroux, and Bouzouane in [33] proposed a formal framework for the recognition process based on lattice theory and action Description Logic (DL). This framework minimizes the uncertainty about observed actors activity by bounding the plausible plans set.

C. Hybrid Approaches

The objective of these kinds of approaches is taking advantage of the features of both Bottom-Up and Top-Down modeling and fusing them in a single modeling approach [24]. Modeling ADLs is a challenging task due to their unique characteristics. For example, there are a large number of ADLs in a variety of categories which can all be modeled at multiple levels of granularity [3]. In addition, most ADLs involve performing a number of actions. The sequence of the actions to be performed is usually dependent on an individuals own preferences [34]. As mentioned before, some actions for different activities may occur together and make overlapped or interleave activities [1][4]. Thus the ideas of using Hybrid approaches have been introduced, which can be divided into two categories: Static Activity Modeling and Dynamic Activity Modeling.

1) Static Activity Modeling: The static activity modeling systems cannot automatically be adapted to accommodate new features in activities performed by the user [35]. Also Top-Down approaches are static and they cannot automatically evolve [24] such as the proposed method in [32]. Some Integration-based Bottom-Up approaches only used to model static characteristics of activities. Dynamic Activity Modeling exposed to discussion due to the modeling dynamic nature of human activities.

2) Dynamic Activity Modeling: The idea of using dynamic modeling is based on the dense sensing paradigm, which establishes the idea of inferring activities by monitoring Human-Object Interactions (HOI) through the usage of multiple multimodal miniaturized sensors [4][24]. Actually in these kinds of modeling want to model high-level activities usually share common sets of physical actions, and are difficult to differentiate based solely on physical signals [36]. To make Top-Down activity recognition systems work in real world applications, activity models have to evolve automatically to adapt to users varying behaviors. The Bottom-Up approaches can be properly addressed to model adaptability and evolution [24]. The goal of this kind of modeling is represented in Fig. 7 as an example.

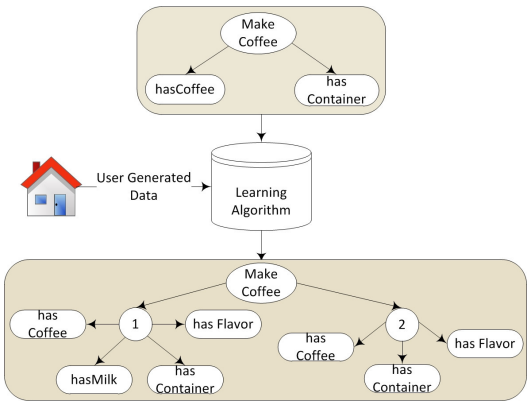


Fig. 7. Dynamic Activity Modeling objective [24]

V. Evaluation of SARF Framework

Due to a wide variety of approaches in human activity recognition, these approaches are classified as SARF framework. Table III represents each of the approaches in this proposed framework according to their characteristics and challenges as a general classification.

Particularly, it is essential to introduce specific measures to evaluate and compare these approaches accurately. The goal of evaluation is analyzing the effects of proposed approaches in human activity recognition and ensure of algorithm performance. Utilizing appropriate measures can lead to well understanding of different approaches for activity recognition in smart homes and also take advantages of them in a systematic and correct way based on the requirements.

A. Proposed Measures

There are different ways to evaluate activity recognition algorithm but generally authors use classifier-based criterias such as F-measure, Precision, Recall and most importantly Accuracy [9][37], and also Sensitivity and Specificity to ignore detailed information about the errors [25] or frameworks such as N-Fold cross validation [37] and Leave-one-day-out [25].

Basically, human activity recognition process has two overall phases: Training and Test. In N-Fold cross validation, the set of data points is split into N non-overlapping subsets. The model is trained and tested N times, on each iteration, one of the N partitions is held out for model testing and the other N-1 partitions are used to train [37]. The performance is averaged over the N iterations. In Leave-one-day-out technique the sensor readings of a whole day are used for testing and the remaining days used for training [38].

In our viewpoint along with other mentioned evaluation measures, there are some important criteria which should be taken into consideration by researchers. Thus, in this section along with Accuracy, as an important measure, these evaluation measures have been proposed.

Data Requirements: In some approaches, due to the needs of large volume of data to support training for each ADL, there is a possibility to face data scarcity which may lead to accuracy and performance reduction [4]. This issue will be increased in the assisted living context which residents are reluctant to reveal their behavioral data due to privacy and ethical considerations [34]. In Top-Down methods there is no data scarcity problem unlike Bottom-Up approaches. Therefore, volume of required data and its importance for human activity recognition in AAL systems such as smart homes must be considered by researchers.

Noise Effect: In general, sensory data are inherently noisy and has untrustworthy nature which leads to lack of reliability in the Bottom-Up, Top-Down and Hybrid approaches [29][31]. As mentioned before, there is possibility of noise existence in annotation process too and lead to accuracy reduction, increase computational complexity and classification error in activity recognition unless some actions such as what is done in [24] using hybrid approaches have been considered.

TABLE III
COMPARISON OF SARF APPROACHES

Activity Learning Approaches		Learning Examples	Main Idea	Characteristics	Challenges
Bottom-UP	Probability-based Methods	Hidden Markov Model[9], Nave Bayes[23]	Probabilistic Classification	Modeling uncertainty and temporal information Generalization Flexibility Dynamic activity modeling	Data scarcity problem Reusability Handling temporal information Dataset annotation
	Similarity-based Methods	Rashidi[8], Conditional Random Field[9], Support Vector Machine[21]	Define the similarity measurement in order to perform patterns selection	Simple and dynamic activity modeling Modeling uncertainty and temporal information Heuristic	Data scarcity problem Reusability Dataset annotation Over-fitting
	Integration-based Methods	Fahad[28], Fahim[29], Chernbumroong[30]	Integration of Similarity or Probability-based methods, or combination of both of these methods	Accuracy Reliability Generalization Efficient Reduce uncertainty in decision making Allow to recognize complex activity	The data scarcity problem Dataset annotation Number and types of classifiers Combination techniques
Top-Down	Description-based Activity Modeling Methods	Zolfaghari[31], Chen[32], Chen[34]	Using semantic and context reasoning to describe concepts and relationships, in a high-level and formal expressiveness	Lack of the data scarcity problem Clear semantic on modeling and inference Interoperability and reusability Preserve decidability Allow to recognize complex activity	Handling uncertainty and ambiguity information Handling temporal information Adaptability Scalability Static activity modeling
	Formalism-based Representation Methods	Bouchard[33]	Logical formalisms inference e.g. deduction, induction, abduction	Lack of the data scarcity problem Clear semantic on modeling and inference	Handling fuzziness and uncertainty information Adaptability Scalability Static activity modeling Flexibility
Hybrid	Static Activity Modeling	Chen[32], Bouchard[33]	Using Probability-based or Similarity-based methods and fusion them with one of Top-Down approaches	Using multiple data sources Accuracy Reliability Allow to recognize complex activity	Limited to initially defined activities Adaptability Performance
	Dynamic Activity Modeling	Azkune[24], Okeyo[35], Wen[36]	Using Probability-based or Similarity-based methods and fusion them with one of Top-Down approaches	Using multiple data sources Adaptability Reusability Allow to recognize complex activity	Common terminology Interoperability Limited to descriptive characteristics Limited to user preferences and implementation tools

Accuracy: Accuracy is the most common criteria in classifier performance analysis and human activity recognition. It should be noted, noise, class-imbalanced datasets and datasets with inappropriate features lead to accuracy reduction [1][7][13]. Higher accuracy of methods leads to error reduction and increase efficiency [16].

Scalability: In general, human activity recognition systems are performing on a particular or public datasets or considering limitation conditions. In fact, the main problem is the needs to real world data which make them inapplicable in other environments with different settings [14]. Furthermore, most of the built models are used for a specific ADL and do not change over time. Also, they do not consider ADL patterns may change due to the dynamic nature of human activities which lead to inconsistency and scalability reduction in built model. In fact, scalability in activity models is an important factor in presence of new activities and new residents in order to constructing a general model for all activities [14], new residents or transfer learning to environment with different layouts [39].

B. Evaluation of Methods According to Proposed Measures

In this section efficiency of human activity recognition approaches classified as proposed SARF framework shown in Fig. 5 is evaluated by proposed measures formerly. Table IV shows the results of this evaluation. It should be noted the values of proposed measures are relative and they are based on research investigation in this field.

As represented in Table IV, due to the Data-Driven nature of Bottom-Up approaches, they require large volume of data to make recognition unlike Top-Down approaches which utilizing prior knowledge and knowledge engineering to human activity recognition in smart homes; therefore, they need to sensory data as lower as other approaches as well as effects of noise on them. On the other hand, there are Hybrid approaches which using Bottom-Up and Top-Down methods all together

TABLE IV
EVALUATION OF PROPOSED SARF FRAMEWORK BASED ON PROPOSED MEASURES

The Proposed SARF Framework		Proposed Evaluation Measures			
		Data Requirement	**Noise Effects**	**Accuracy**	**Scalability**
Bottom-Up	Probability-based Methods	High	High	Medium	Almost Medium
	Similarity-based Methods	High	High	Medium	Almost Medium
	Integration-based Methods	High	Medium	Almost High	Almost Medium
Top-Down	Description-based Activity Modeling	Low	Low	Medium	Almost Medium
	Formalism-based Representation Methods	Low	Low	Medium	Low
Hybrid	Static Activity Modeling	Medium	Medium	Almost High	Medium
	Dynamic Activity Modeling	Medium	Medium	Medium	High

to achieve acceptable scalability along with adaptability to dynamic nature of human behavior especially in dynamic activity modeling. Therefore, in these kinds of approaches we face to sensor data requirement as well as noise effect but not as much as Bottom-Up approaches.

As mentioned in proposed SARF framework, combining multiple methods together can improve accuracy of human activity recognition in smart homes as well as using Hybrid approaches especially static activity modeling due to its static assumption. Furthermore, there is data requirement in Integration-based methods due to its Bottom-Up nature. Also, inherently noisy sensory data can lead to accuracy reduction in these methods. However, the most effective way to reduce noise impacts, as mentioned in preprocessing phase, is cleaning, completing and normalizing. As represented in Table IV, the other approaches can achieve medium and almost acceptable accuracy in human activity recognition in smart homes.

VI. CONCLUSION

In this paper different approaches to human activity recognition in smart homes investigated and described how to evaluate these approaches were classified and presented in the proposed framework, i.e. SARF, using the obtained results. In order to provide a convenient tool for selecting appropriate approaches, results presented in the form of diagrams and characteristics of each group were investigated and evaluate based on proposed measures represented in form of tables.

The results of this study show that there is no unique way to introduce a single approach, as an optimal approach, to human activity recognition in AAL systems. Since each approach is used for a specific purpose comparing the approaches does not make any sense. One of the most important issues in human activity recognition is to remove the challenges and improve the efficiency of algorithms which is a dynamic research domain warranting further investigation. Using the SARF proposed framework in this paper can play an important role in development of our knowledge in this area and a starting point to resolve some of the challenges which were outlined in this paper.

REFERENCES

[1] D. J. Cook, N. C. Krishnan,*Activity learning: discovering, recognizing, and predicting human behavior from sensor data.* John Wiley and Sons, 2015.

[2] S. R. Ke, H.L. U. Thuc, Y.J. Lee, J.N. Hwang, J.H. Yoo, and K.H. Choi, "A review on video-based human activity recognition," *Comput.,* vol. 2, no. 2, pp. 88–131, 2013.

[3] Q. Ni, A.B. Garca Hernando, and I. Pau de la Cruz, "The elderlys independent living in smart homes: A characterization of activities and sensing infrastructure survey to facilitate services development," *Sensors,* vol. 15, no. 5, pp. 11312–11362, 2015.

[4] L. Chen, J. Hoey, C. Nugent, D. Cook, and Z. Yu, "Sensor-based activity recognition," *IEEE Trans. Syst. Man Cybern. Part C: Appl. Rev.,* vol. 42, no. 6, pp. 790–808, 2012.

[5] M. Koohzadi, M.R. Keyvanpour, "An analytical framework for event mining in video data," *Art. Intell. Rev.,* vol. 41, no. 3, pp. 401–413, 2014.

[6] M. Koohzadi, M.R. Keyvanpour, "OTWC: an efficient object-tracking method," *Signal, Image and Video Proc.,* vol. 9, no. 6, pp. 1235–1247, 2015.

[7] H. Fang, L. He, H. Si, P. Liu, and X. Xie, "Human activity recognition based on feature selection in smart home using back-propagation algorithm," *ISA Trans.,* vol. 53, no. 5, pp. 1629-1638, 2014.

[8] P. Rashidi, D. Cook, L. Holder, and M. S. Edgecombe, "Discovering activities to recognize and track in a smart environment," *IEEE Trans. Knowl. Data Eng.,* vol. 23, no. 4, pp. 527-539, 2011.

[9] T. V. Kasteren., A. Noulas, G. Englebienne, and B. Krse, "Accurate activity recognition in a home setting," in Proc. Of the Ubicomp ACM., pp. 19, 2008.

[10] B. Minor, J. R. Doppa, and D. J. Cook, "Data-Driven activity prediction: algorithms, evaluation methodology, and applications," In Proc. Of the 21th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining, New York, NY, USA, 2015, pp. 805–814.

[11] P. Rashidi, N. Krishnan, and D. J. Cook, "Discovering and tracking patterns of interest in security sensor streams," *Securing Cyber-Physical Critical Infrastructure, chapitre,* pp. 481–504, 2012.

[12] P. Rashidi, A. Mihailidis, "A survey on ambient-assisted living tools for older adults," *IEEE J. of Biomedical and Health Informatics*, vol. 17, no. 3, pp. 579–90, 2013.

[13] C. Chen, B. Das, and D.J. Cook, "A data mining framework for activity recognition in smart environments," In Proc. Of IEEE Int. Conf. on Intell. Environments, Kuala Lumpur, Malaysia, 2010, pp. 80-83.

[14] M. Amiribesheli, A. Benmansour, and A. Bouchachia, "A review of smart homes in healthcare," *Ambient Intell. and Humanized Comput.,* pp. 1-23, 2015.

[15] O.D. Lara, M.A. Labrador, "A survey on human activity recognition using wearable sensors," *Communications Surveys and Tutorials,* vol. 15, no. 3, pp. 1192–1209, 2013.

[16] S. Mehr Molaei, M.R. Keyvanpour, "An analytical review for event prediction system on time series," In 2nd Inter. Conf. on Pattern Recognition and Image Analysis (IPRIA), 2015, pp. 1–6.

[17] J. Wen, M. Zhong, "Activity discovering and modeling with labeled and unlabeled data in smart environments," *Expert Syst. Appl.,* vol. 42, no. 14, pp. 5800-5810, 2015.

[18] N.C. Krishnan, D.J. Cook, "Activity recognition on streaming sensor data," *Perv. and Mob. comput.,* vol. 10, pp. 54–138,2014.

[19] T. R. D. Saputri, A. M. Khan, and S. W. Lee, "User-Independent activity recognition via three-stage GA-based feature selection," *Int. J. of Distributed Sensor Net.,* 2014.

[20] L.G. Fahad, S.F. Tahir, and M. Rajarajan, "Feature selection and data balancing for activity recognition in smart homes," In Proc. Of the IEEE Int. Conf. on Communications (ICC), London, UK, 2015, pp. 512-517.

[21] A. Fleury, M. Vacher, and N. Noury, "SVM-based multimodal classification of activities of daily living in health smart homes: sensors, algorithms, and first experimental results," *IEEE Trans. on Information Technology in Biomedicine,* vol. 14, no. 2, pp. 274–283, 2010.
[22] A. Janecek, WN. Gansterer, M. Demel, and G. Ecker, "On the Relationship Between Feature Selection and Classification Accuracy," In FSDM 2008, pp. 90–105.
[23] E.M. Tapia, S. I. Stephen, and K. Larson, "Activity recognition in the home using simple and ubiquitous sensors," Springer Berlin Heidelberg, pp. 158–175, 2004.
[24] G. Azkune, A. Almeida, D.L. de Ipia, and L. Chen, "Extending knowledge-driven activity models through data-driven learning techniques," *Expert Syst. Appl.,* vol. 42. no. 6, pp. 3115-3128, 2015.
[25] H. Alemdar, C. Tunca, and C. Ersoy, "Daily life behavior monitoring for health assessment using machine learning: bridging the gap between domains," *Personal and Ubiquitous Comput.,* vol. 19, no. 2, pp. 1–13, 2014.
[26] T. Jebara, "Discriminative, Generative and Imitative Learning, " Ph.D. thesis, MIT, 2001.
[27] L. G. Fahad, A. Ali, and M. Rajarajan, "Learning models for activity recognition in smart homes," In Proc. Of Int. Conf. on Inf. Sci. Appl., Pattaya, Thailand, 2015, pp. 819-826.
[28] L.G. Fahad, M. Rajarajan, "Integration of discriminative and generative models for activity recognition in smart homes," *Applied Soft Comput.,* 2015.
[29] M. Fahim, I. Fatima , S. Lee, and Y.K. Lee, "EEM: evolutionary ensembles model for activity recognition in Smart Homes," *Applied intell.,* vol. 38, no. 1, pp. 88–98, 2013.
[30] S. Chernbumroong, S. Cang, and H. Yu, "Genetic algorithm-based classifiers fusion for multi-sensor activity recognition of elderly people," *IEEE J. of Biomed. and Health Informatics,* vol. 19, no. 1, pp. 282-289, 2015.
[31] S. Zolfaghari, R. Zall, M.R. Keyvanpour, "SOnAr: Smart Ontology Activity recognition framework to fulfill Semantic Web in smart homes," In Second International Conference on Web Research (ICWR), 2016, pp. 139–144.
[32] L. Chen, C. Nugent, and H. Wang, "A knowledge-driven approach to activity recognition in smart homes," *IEEE Trans. Knowl. Data Eng.,* vol. 24, no. 6, pp. 961-974, 2012.
[33] B. Bouchard, S. Giroux, and A. Bouzouane, "A smart home agent for plan recognition of cognitively-impaired patients," *J. of Computers,* vol. 1, no. 5, pp. 53–62, 2006.
[34] L. Chen, C. Nugent, and G. Okeyo, "An ontology-based hybrid approach to activity modeling for smart homes," *IEEE Trans. on Human-Machine Syst.,* vol. 44, no. 1, pp. 92-105, 2014.
[35] G. Okeyo, L. Chen, H. Wang, "Combining ontological and temporal formalisms for composite activity modeling and recognition in smart homes," *Future Gener.Comput. Syst.,* vol. 39, pp. 2943, 2014.
[36] J. Wen, M. Zhong, and Z. Wang, "Activity recognition with weighted frequent patterns mining in smart environments," *Expert Syst. Appl.,* vol. 42, no. 17, pp. 6423-6432, 2015.
[37] L.G. Fahad, A. Khan, and M. Rajarajan, "Activity recognition in smart homes with self-verification of assignments," *Neurocomputing,* vol. 149, pp. 1286-1298, 2015.
[38] I. Fatima, M. Fahim., Y. K. Lee, and S. Lee, "A Genetic Algorithm-based classifier ensemble optimization for activity recognition in smart Homes," *KSII Trans. on Internet and Inf. Syst. (TIIS),* vol. 7, no. 11, pp. 2853–2873, 2013.
[39] D.J. Cook, K. Feuz, and N.C. Krishnan, "Transfer learning for activity recognition: A survey," *Knowledge and Info. Sys.,* vol. 36, no. 3, pp. 537-556, 2013.

Joint Agent-oriented Workshops in Synergy

JOINT Agent-oriented Workshops in Synergy is a coalition of agent-oriented workshops that come together to build upon synergies of interests and aim at bringing together researchers from the agent community for lively discussions and exchange of ideas. For the first time JAWS was organized during the 2011 FedCSIS Conference. Workshops that constitute JAWS in 2016 are:

- MAS&S'16 - 10^{th} International Workshop on Multi-Agent Systems and Simulations
- SEN-MAS'16 - 4^{th} International Workshop on Smart Energy Networks & Multi-Agent Systems

10^{th} International Workshop on Multi-Agent Systems and Simulations

MULTI-AGENT systems (MASs) provide powerful models for representing both real-world systems and applications with an appropriate degree of complexity and dynamics. Several research and industrial experiences have already shown that the use of MASs offers advantages in a wide range of application domains (e.g. financial, economic, social, logistic, chemical, engineering). When MASs represent software applications to be effectively delivered, they need to be validated and evaluated before their deployment and execution, thus methodologies that support validation and evaluation through simulation of the MAS under development are highly required. In other emerging areas (e.g. ACE, ACF), MASs are designed for representing systems at different levels of complexity through the use of autonomous, goal-driven and interacting entities organized into societies which exhibit emergent properties The agent-based model of a system can then be executed to simulate the behavior of the complete system so that knowledge of the behaviors of the entities (micro-level) produce an understanding of the overall outcome at the system-level (macro-level). In both cases (MASs as software applications and MASs as models for the analysis of complex systems), simulation plays a crucial role that needs to be further investigated.

TOPICS

MAS&S'16 aims at providing a forum for discussing recent advances in Engineering Complex Systems by exploiting Agent-Based Modeling and Simulation. In particular, the areas of interest are the following (although this list should not be considered as exclusive):

- Agent-based simulation techniques and methodologies
- Discrete-event simulation of Multi-Agent Systems
- Simulation as validation tool for the development process of MAS
- Agent-oriented methodologies incorporating simulation tools
- MAS simulation driven by formal models
- MAS simulation toolkits and frameworks
- Testing vs. simulation of MAS
- Industrial case studies based on MAS and simulation/testing
- Agent-based Modeling and Simulation (ABMS)
- Agent Computational Economics (ACE)
- Agent Computational Finance (ACF)
- Agent-based simulation of networked systems
- Scalability in agent-based simulation

STEERING COMMITTEE

- **Cossentino, Massimo,** ICAR-CNR, Italy
- **Fortino, Giancarlo,** Universita della Calabria, Italy
- **Gleizes, Marie-Pierre,** Universite Paul Sabatier, France
- **Pavon, Juan,** Universidad Complutense de Madrid, Spain
- **Russo, Wilma,** Universita della Calabria, Italy

EVENT CHAIRS

- **Fortino, Giancarlo,** Universita della Calabria, Italy
- **Fuentes-Fernández, Rubén,** Research Group on Agent-based, Social & Interdisciplinary Applications (GRASIA), University Complutense of Madrid (UCM), Spain
- **Niazi, Muaz,** COMSATS Institute of IT, Pakistan
- **Seidita, Valeria,** Università degli Studi di Palermo, Italy

PROGRAM COMMITTEE

- **Alam, Shah Jamal**
- **Antunes, Luis**
- **Arcangeli, Jean-Paul,** Université Paul Sabatier, France
- **Bernon, Carole,** Universite Paul Sabatier, France
- **Cipresso, Pietro**
- **Cossentino, Massimo,** ICAR-CNR, Italy
- **Davidsson, Paul,** Malmö University, Sweden
- **Garro, Alfredo,** University of Calabria, Italy
- **Gomez-Sanz, Jorge J.,** Universidad Complutense de Madrid, Spain
- **Gravina, Raffaele,** University of Calabria, Italy
- **Guerrieri, Antonio,** University of Calabria, Italy
- **Klügl, Franziska,** Örebro Universitet, Sweden
- **Molesini, Ambra,** Università di Bologna, Italy
- **Petta, Paolo,** OFAI, Austria
- **Ribino, Patrizia,** Istituto di Reti e Calcolo ad Alte Prestazioni - Consiglio Nazionale delle Ricerche, Italy
- **Savaglio, Claudio,** Universita della Calabria
- **Vizzari, Giuseppe,** Università di Milano Bicocca, Italy

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1449–1452
DOI: 10.15439/2016F359
ACSIS, Vol. 8. ISSN 2300-5963

Agent-oriented Modeling and Simulation of IoT Networks

Giancarlo Fortino
DIMES - University of Calabria
Via P. Bucci, cubo 41C, 87036
Rende (CS), Italy
g.fortino@unical.it

Wilma Russo
DIMES - University of Calabria
Via P. Bucci, cubo 41C, 87036
Rende (CS), Italy
w.russo@unical.it

Claudio Savaglio
DIMES - University of Calabria
Via P. Bucci, cubo 41C, 87036
Rende (CS), Italy
csavaglio@dimes.unical.it.

***Abstract*—Internet of Things (IoT) networks are being continually developed in several domains, however no systematic processes for their modeling and simulation exist so far. In this paper, an agent-oriented approach to IoT networks modeling is proposed by exploiting the ACOSO model. Then, agent-modelled IoT networks of different scales are simulated through the Omnet++ simulation platform, with the goal of analyzing issues and bottlenecks at communication level.**

I. Introduction

SMART Objects (SOs) constitute a new generation of enhanced everyday things (able to perceive the surrounding physical environment, elaborate and communicate the acquired information, and hence provide cyber-physical services) that are globally networked and mutually interacting, even without a steady human orchestration [1]. SOs are technologically and functionally heterogeneous, thus their clustering constitutes "Internet of Things" (IoT) networks [2] that look like loose collections of heterogeneous devices and sub-networks requiring distributed mechanisms of communication and management. Whereas defining methods to model, design and simulate IoT networks before their final implementation is an open challenge, no research efforts have currently been devoted to systematically address such issue.

As IoT networks share many substantial features/issues with multi-agent systems (MAS) [3] and SOs may be effectively modeled as agents, in this paper an agent-oriented approach to model IoT networks is presented. In particular, the ACOSO (Agent-based Cooperative Smart Object)-based SO model [4] has been exploited to describe, from the agent perspective, SO main features, relationships and interactions. Moreover, in order to analyze IoT networks at SO communication level, the Omnet++ [5] network simulation platform has been used to evaluate bottlenecks and issues in specifically defined scenarios. The rest of the paper is organized as follows: in Section II, the ACOSO-based SO model and other concrete examples of the Agent-oriented Modeling (AoM) applied to the SO-based IoT context are introduced and compared. In Section III, the designed simulation scenarios, the selected metrics, and the obtained results are presented and discussed. Finally, conclusions are drawn and future work is briefly delineated.

This work has been partially carried out under the framework of INTER-IoT, Research and Innovation action - Horizon 2020 European Project, Grant Agreement #687283, financed by the European Union.

II. Agent-oriented Modeling

Agent-based Computing paradigm has been successfully used over the years for the analysis, modeling and implementation of complex, cooperative and adaptive distributed systems [3]. The *agent* abstraction, indeed, is suitable to model and implement autonomous, intelligent and interacting entities, characterized by different behaviors and goals. Nevertheless is well-established that the AoM [6], differently to other modeling paradigms (e.g. object-oriented, service-oriented, etc.) is able to fully support proactiveness and situatedness, key features in SO-based IoT networks, to date few agent-oriented models are available in the literature in this context.

With reference to the agent-oriented SO modeling, [7] and [8] present coarse-grained models, characterized by a high degree of abstraction and therefore mostly suitable to support the SO analysis phase; the ACOSO-based SO model [4], instead, specializes the SO ecosystem in a deeper degree of detail, thus supporting also the SO design and implementation phases. In detail, in [7] the authors envision an IoT system architecture where each resource (e.g. computer, SO, human user) is represented by an agent and interconnected to the rest of the cyber-physical world by means of specific adapters. Each agent has a role (and not an identifier) that determines its own behaviors, tasks and communication paradigms. Both roles and behaviors are taken from two repositories which are assumed to be managed depending on the application scenario. In [8] the SO model comprises five elements: the execution environment (to run the actual agent task and manage its lifecycle), a repository (which contains both database and knowledge base), a set of physical components (such as sensors and actuators), the agent interface (to enable the intra-agent information exchanges among the aforementioned SO-model elements) and the object interface (for communication with other SOs and with the system).

The ACOSO-based SO model allows the modeling of high-level SO main features (basic information of an SO, its augmentation devices like sensors and actuators, its services, etc.) and it also extensively describes the functional components of the system, their relationships and interactions. Due to such reasons, the ACOSO-based SO model represents one of the cornerstones of ACOSO (Agent-based Cooperative Smart Object) [9], a middleware specifically conceived for the full management and development of agent-oriented cooperating SOs. In detail, the ACOSO-based SO model abstracts SOs in event-driven cooperating agents whose specific objectives are encapsulated in their behavior and modeled as *Tasks*. A Task is an event-driven and state-based component that can refer to the common operations required for the agent lifecycle management (SystemTask) or to specific-purpose operations defining the specific behaviors of the SO (UserDefinedTask). Indeed, SO services are mapped on UserDefinedTask. By means of different tasks the SO exploits different subsystems in order to react to external stimulus, to fulfill specific goals and to exploit inference rules on local/remote knowledge bases. In particular:

- The *DeviceManagementSystem*, through multiple DeviceAdapters, handles the SO augmentation devices that enables SO to interact with the physical world generating Device Events. In particular, the BMFAdapter [10] and the SPINEAdapter [11-13] enable the management of environmental and wearable sensor networks.
- The *CommunicationManagementSystem* provides communication services between agents and external entities. Different kinds of interactions (e.g. intra-agent FIPA-ACL based interactions or inter-entities UDP/TCP-based interactions) are enabled by means of different CommunicationAdapters. Both events generated inside (InternalEvent) or outside (ExternalEvent) the SO are handled by the CommunicationManagementSystem.
- The *KBManagementSystem* exploits local or remote knowledge bases to handle information pertaining the SO, its current status, its inference rules and other useful data that can be shared among the agent tasks.

TABLE I
AGENT-BASED SO MODELS COMPARISON

Agent-based SO Modelling	**[7]**	**[8]**	**ACOSO-based SO model [4]**
SO-Modeling Phase / *SO Characteristic*	Analysis	Analysis	Analysis & Design & Implementation
Augmentation Devices	Adapter	Physical Components	DeviceManagementSystem (DeviceAdapter)
Communication	Role	Int./Ext. Agent Interface	CommunicationManagementSystem (CommunicationAdapter)
Decision-Making	Role	Execution Environment	Set of Task (Agent-Behavior)
Service Provisioning	N/A	N/A	UserDefinedTask, Event
Knowledge	Role	Repository	KBManagementSystem, Behavior

A. Comparison

Table I shows a comparison of the three agent-oriented SO models introduced in this Section. First of all, it has been highlighted that they are suitable in different modeling phases. Indeed, [7] and [8] are mostly suitable to approach the SO analysis phase while the ACOSO-based model support also design and implementation phases. This implies that the three models describe the main SO characteristics (augmentation, communication, decision-making, service provisioning and knowledge) with a different degree of detail. In particular:

- *Augmentation Devices*: in SO models of [7] and [8] sensors and actuators are not explicitly modeled as they are considered minor SO components and their interactions are not elicited. In the ACOSO-based SO model augmentation devices are managed by the DeviceManagementSystem and handled by multiple DeviceAdapters.
- *Communication*: in [7] the SO role determines its communication paradigm and structure, while in [8] internal/external SO communication interfaces are defined. In the ACOSO-based SO model communication is managed through the CommunicationManagementSystem and a set of CommunicationAdapters.
- *Decision-Making*: in [7] an SO/agent autonomously behaves to achieve certain goals depending on its role in the domain (e.g. smart car, smart driver-support or smart road within the smart transportation domain), while in [8] task execution concerns the Execution Environment. Decision-making relies on the agent behaviors in ACOSO-based SO model, defined through tasks.
- *Service Provisioning*: in [7] and [8] the concept of SO cyber-physical service is not completely declined and is mostly reduced to a simple Web Service. In the ACOSO-based SO model services are mapped on UserDefinedTasks (enabled by specific events).
- *Knowledge Base*: in [7] the SO knowledge is stored in repository organized in function of the SO role, while in [8] relational databases record SO-related information. In the ACOSO-based SO model the information is spread between the KBManagementSystem and the agent behavior.

III. Simulation-based Analysis

As highlighted in the previous Section, the ACOSO-based SO model, leveraging on both the agent and MAS concepts, enables the effective modeling of SO-based IoT networks at different development phases. In this Section, the simulation of IoT networks allowing the validation of models, protocols and algorithms before the actual deployment of the network infrastructure, is addressed. Indeed, IoT networks simulation is an important but complex task because, depending of different scales, the number of the SOs may vary from dozens (e.g. home automation or body sensor networks) to thousands (e.g. Smart City scenario), with a different degree of density and different communication paradigms depending on the specific service. In addition, factors unrelated to the applications but specifically associated to the networking (e.g. traffic congestion, wireless signal attenuation and coverage, etc.) influence the SOs interactions and the service provision/fruition. Taking into account these issues and particularly focusing on communication among SOs, the IoT networks previously described through the agent-oriented approach are simulated by means of Omnet++ [5]. The modeling of an agent through an Omnet++ network node is straightforward. In fact, each network node/SO can be considered as an autonomous agent whose behaviors and tasks, which realize SO decision making and service provisioning (see Table I), are implemented at the application layer. All the other tasks related to transport-network-link protocol implementations, wireless connectivity issues, physical environment modeling are carried out by Omnet++.

In particular, Omnet++ makes it possible to design nodes with different communication boards or interface ports (to simulate the connection with external physical devices); in the following simulations, due to the intrinsic wireless nature of the IoT networks, nodes with 802.11 wireless boards have been considered.

Simulations have inspected IoT networks in the Information Exchange phase (IE) by exploiting TCP-based reliable and UDP-based unreliable transport protocols. The round trip time (RTT) and the packet delivery ratio (PDR) have been hence measured considering nodes that exchange empty messages and exploiting either a Client/Server (C/S) or a Peer-to-Peer (P2P) paradigm. Deterministic (1 pk/s) and stochastic Normal (with 0.5 mean and 0.2 variance) data generation models (DGM) have been used. In Table II the communication settings exploited in the simulations for the RTT/PDR measuring are shown.

Table II
Information Exchange (IE) Settings

Parameters	DGM (Data Generation Model)
Patterns	P2P (Peer-to-Peer), C/S (Client/Server)
Protocols	R (Reliable), U (Unreliable)

Both RTT and PDR have been evaluated in the context of small-, medium-, large-scale IoT networks with different SOs density. In particular, simulations took into account:

- The number of involved SOs (#SOs), since network congestion may increase depending of the SOs population. In the following SOs population is considered limited to 100 nodes for small-scale networks, 500 nodes for medium-scale networks and 1000 nodes for large-scale networks;
- SO distribution in a different number of subnetworks (#subnetworks). In the following it is assumed that small-scale networks are constituted by a single network, medium-scale networks comprise two or more subnetworks deployed in the same area so that their coverages overlap, and finally large-scale networks comprise two or more subnetworks deployed in not adjacent area;
- The deployment area in which SOs are located (supposing that they have no mobility patterns) since the proximity of multiple SOs may cause signal interferences in the wireless communications. For sake of simplicity, square grid areas with different side dimensions have been considered.

Table III summarizes the scenarios tested during the simulation phase.

Table III
Simulation Scenarios

	#SOs	**# subnetworks**	**Grid side [m]**
Small Scale (S)	5..100	1	5..100
Medium Scale (M)	20..500	1..10	From 100
Large Scale (L)	20..1000	1..20	From 100

In the following, for the sake of space, only some results of the simulations are shown, and in particular the PDR in the case of small-scale networks with #SOs increasing (see Fig. 1) and the RTT in the other two scales with #subnetworks increasing (see Fig. 2 and 3).

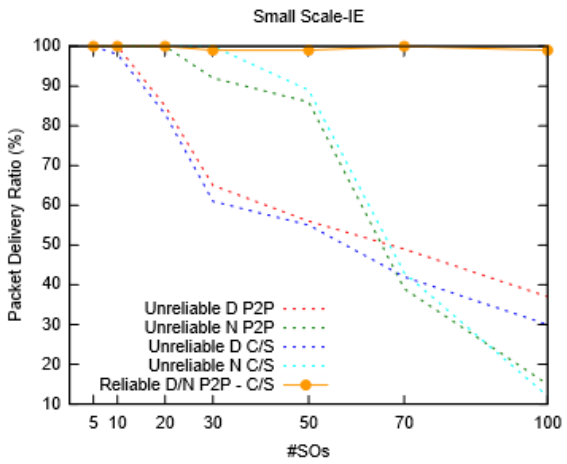


Figure 1: PDR when the #SOs increases in a square grid of side 100 m

Fig. 1, in the case of unreliable protocol shows that (i) with increasing of SOs, PDR decrease due to communication fails caused by interferences; (ii) with the same DGM, C/S and P2P patterns are equivalent; (iii) with a small number of SOs (less than 70) non-deterministic data generation models

outperform the deterministic ones, while with the increase of SOs the trend inverts. In the case of reliable protocol, as might be expected, the PDR keeps the maximum value regardless of the DGM, patterns, and #SOs.

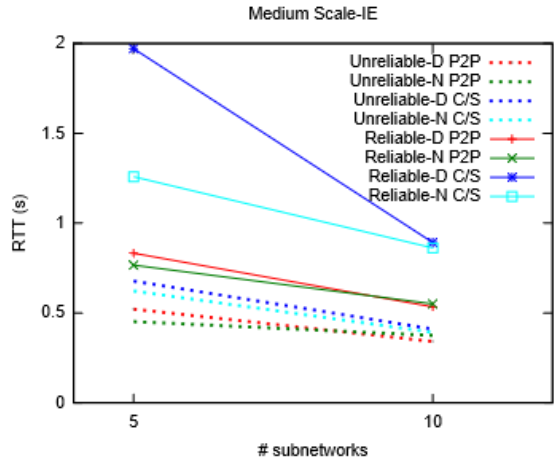


Fig. 2: RTT of 500 SOs equally distributed into 5 and 10 subnetworks

Fig. 2 shows that the RTT decreases if the same number of SOs (500) is deployed on the same area but distributed on more subnetworks (5 and 10 square grids of side 100 m), since the traffic is well-balanced. As might be expected, the unreliable protocols outperform the reliable ones.

Fig. 3 shows that in the large-scale scenario (5, 10 and 20 subnetworks deployed in a squared area of side 100m with 50 SOs each) the absence of interferences among the subnetworks generates RTT values quite stable and lower than the correspondent ones in the medium-scale scenario.

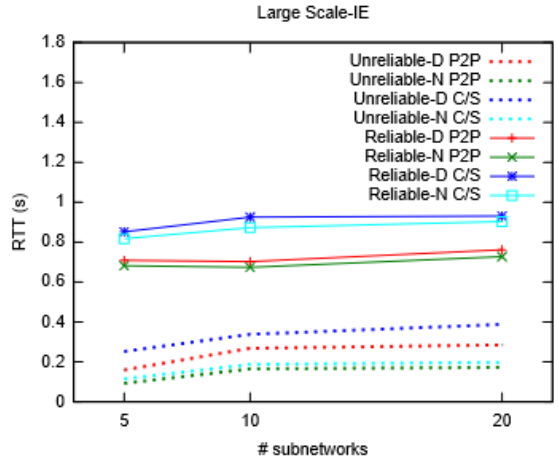


Fig. 3: RTT of 5,10 and 20 subnetworks of 50 SOs each

IV. Conclusion

In this paper the agent-oriented modeling of SO systems through the ACOSO-based SO model has been presented, by considering that (i) SOs and software agents share multiple features, and (ii) agent-based modeling flexibly assists the conceptualizing of dynamic and autonomous distributed systems in different contexts. Beside the agent-oriented SO modeling, IoT networks of different scales have been simulated through the Omnet++ simulator, with a particular attention to the SO communication in order to evaluate performances (focusing on bottlenecks, issues and networks dynamics) and validate network design choices. Simulation results highlight that multiple factors influence the network performance (as the SO number, their more or less dense distribution and the selected communication settings) and thus their combination should be made by following an application-driven approach.

Further research efforts will be devoted to (i) provide a full mapping between ACOSO-based SO model and Omnet++ simulation, in order to translate on such simulator the behavioral/application side of the SO systems, including also the simulation of the device layer; (ii) develop a full-fledged methodology for the analysis, design, simulation, implementation and validation of SO-based IoT systems [14, 15].

References

[1] G. Kortuem. et al. “Smart objects as building blocks for the internet of things,” *Internet Computing*, IEEE, vol. 14, no. 1, 2010, pp. 44-51.

[2] L. Atzori, A. Iera, and G. Morabito, “The internet of things: A survey,” *Computer networks,* vol. 54, no. 15, 2010, pp.2787-2805.

[3] W. Van der Hoek, and M. Wooldridge, “Multi-agent systems,” *Handbook of Knowledge Representation,* 2008, pp. 887-928.

[4] G. Fortino, A. Guerrieri, W. Russo, and C. Savaglio, “Towards a Development Methodology for Smart Object-Oriented IoT Systems: A Metamodel Approach,” in *Systems, Man, and Cybernetics (SMC), 2015 IEEE International Conference on*, 2015, pp. 1297-1302.

[5] A. Varga, *OMNeT++. Modeling and Tools for Network Simulation*, Springer Berlin Heidelberg, 2010, pp. 35-59.

[6] C. Macal, and M. J. North, “Tutorial on agent-based modeling and simulation,” in *Proceedings of the 37th conference on Winter simulation*, 2005, pp. 2-15.

[7] A. Katasonov, et al. “Smart Semantic Middleware for the Internet of Things,” *ICINCO-ICSO*, vol. 8, 2008, pp. 169-178.

[8] T. Leppänen, J. Riekki, M. Liu, E. Harjula, T. Ojala, “Mobile agents based smart objects for the internet of things,” *Internet of Things Based on Smart Objects, Springer Int. Publishing*, 2014, pp. 29-48.

[9] G. Fortino, A. Guerrieri, M. Lacopo, M. Lucia, and W. Russo, “An agent-based middleware for cooperating smart objects,” *Highlights on Practical Applications of Agents and Multi-Agent Systems*, Springer Berlin Heidelberg, 2013, pp. 387-398.

[10] G. Fortino, A. Guerrieri, G. M. P. O'Hare, and A. G. Ruzzelli, “A flexible building management framework based on wireless sensor and actuator networks,” *Journal of Network and Computer Applications,* vol. 35, no. 6, 2012, pp. 1934-1952.

[11] G. Fortino, A. Guerrieri, F. Bellifemine, and R. Giannantonio, “SPINE2: developing BSN applications on heterogeneous sensor nodes,” *IEEE International Symposium on Industrial Embedded Systems*, 2009, pp. 128-131.

[12] F. Bellifemine, G. Fortino, A. Guerrieri, and R. Giannantonio, “Platform-independent development of collaborative Wireless Body Sensor Network applications: SPINE2,” *Systems, Man and Cybernetics (SMC 2009), IEEE International Conference on*, 2009, pp. 3144-3150.

[13] G. Fortino, S. Galzarano, R. Gravina, and W. Li, “A framework for collaborative computing and multi-sensor data fusion in body sensor networks,” *Information Fusion,* vol. 22, 2015, pp. 50-70.

[14] G. Fortino, A. Garro, and W. Russo, “Achieving Mobile Agent Systems interoperability through software layering,” *Information & Software Technology*, vol. 50, no. 4, 2008, pp. 322-341.

[15] G. Fortino, A. Garro, and W. Russo, “An integrated approach for the development and validation of multi-agent systems,” *International Journal of Computer Systems Science & Engineering*, vol. 20, no. 4, 2005, pp. 259-271.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1453–1460
DOI: 10.15439/2016F547
ACSIS, Vol. 8. ISSN 2300-5963

Modelling Group Constructions for Social Analysis

Daniela Xavier
GARP (Genomic and RNA Profiling Core),
Department of Molecular and Human Genetics,
Baylor College of Medicine
in Houston, United States
Email: xavier@bcm.edu

Rubén Fuentes-Fernández
GRASIA (Research Group on
Agent-based, Social & Interdisciplinary Applications),
Universidad Complutense de Madrid
in Madrid, Spain
Email: ruben@fdi.ucm.es

***Abstract*—Agent-based modelling is becoming widely used for studies in Social Sciences. However, its application faces limitations coming from its bias to software development, which precludes a more active involvement of social researchers. In order to deal with this problem, this work proposes using domain-specific modelling languages based on the socio-psychological Activity Theory. These languages apply agent research to crystallize that theoretical framework in a formal definition suitable for automated processing but close to Social Sciences. The paper focuses on the language for the specification of group constructions such as organizations, norms and shared knowledge. A case study about contradictory decisions in the space shuttle program illustrates the discussion.**

I. Introduction

AGENT-BASED Modelling (ABM) [1] has become a mainstream technique for research in Social Sciences. Traditionally, this research requires the gathering of data over potentially long periods of time and about large populations that are not fully controlled, which largely increases its costs. The formal description of social systems with models [2] allows applying analysis techniques based on simulation and verification, reducing the needs of eliciting data for the initial testing of hypotheses. ABM facilitates modelling by providing social and intentional computational abstractions that are closer to the concepts used in this research field than those of other approaches.

Despite of its advantages, ABM is limited by its inherent development complexity. The design, implementation and use of an archetypical agent-based model involve different subtasks and roles, which need diverse backgrounds and competences [3]. This situation may lead to misunderstandings between the different stakeholders, which make it hard to guarantee that the model really corresponds to the initial requirements of social science researchers [4]. Available methodologies to develop such ABM models [3], [5], [6] offer little help to address this problem. They focus on the researchers' conceptual models and describe very general tasks for the development. Agent-Oriented Software Engineering (AOSE) also seems not to be a suitable solution, since it mostly deals with the development of standard Multi-Agent Systems (MAS) [7]. On the contrary, ABM focuses more on the translation of the conceptual models of social science researchers to computational models using agent abstractions [5]. Besides, MAS are usually tied to certain architectural patterns, while ABM deals with an enormous heterogeneity of structures [1]. There are also differences about implementation. ABM applications usually require centralized monitoring components that can access the internals of agents in order to gather analysis data [8], and sacrifice the individual agent complexity in favour of huge populations [4]. These aspects are often not considered by AOSE methodologies [7].

Model-Driven Engineering (MDE) [9] with Domain-Specific Languages (DSL) [10] has been proposed as a way to overcome some of these limitations [11], [12]. A DSL for ABM uses a vocabulary grounded on conceptual frameworks from Social Sciences. The DSL has a formal definition which enables the use of automated MDE techniques to process its models, for instance to generate the code for simulations. Such approach reduces the impact of misunderstandings in ABM in two ways. First, social science researchers are able to perform the modelling themselves using a language they are familiar with. Second, transformations from conceptual models to computational ones can be generalized and reused in different projects. This offers improved opportunities for verification and validation. There are some preliminary examples of this trend, but they are still too biased to their foundations in software development [12] or only present partial solutions for some modelling aspects [11]. Especially issues like the management of large populations or centralized supervision are still open.

This paper introduces a DSL for modelling social constructs and organizational interactions. It is part of the ATCAS (Activity Theory for the Computational Analysis of Societies) framework. The Activity Theory (AT) [13] is a well-known paradigm for the analysis of human groups. It focuses on the study of *activities*, which are interactive acts between people and their physical and socio-cultural environment, where both of them act on and mould the other at the same time. ATCAS is intended to provide a general and extensible basis for ABM in social research supporting different applications. The actual representation of AT concepts in ATCAS depends on agent concepts, for instance about inconsistency management [14], code generation [12] and organizational interactions [15]. This last aspect is the focus of the current paper.

From the AT perspective [16], the social aspect of groups regards **communities** of subjects engaged in shared activities. The norms of the *division of labour* organize these activities and *rules* emerging from the socio-cultural environment influ-

978-83-60810-90-3/$25.00

ence and constraint both communities and activities. AT works at the level of abstraction of human societies but the automated analysis of models in MDE requires formalizing them as computational abstractions. ATCAS-IL adopts for this purpose the OperA [15] framework for the specification and analysis of agent organizations. OperA provides predefined modelling primitives for the high-level specification of organizations and the Logic for Contract Representation (LCR), which is an extension of deontic temporal logic, for the fine-grained details. The choice of OperA is motivated by the fact that AT explicitly considers social features that are not embedded in individual agents, but apply to the society as a whole. ATCAS-IL extends OperA primitives with AT concepts and introduces a macro mechanism intended to tailor the language for the specific needs of social science researchers. This gradually increases their autonomy in modelling, as they can define their own abstractions.

The formal description of ATCAS-IL uses metamodels. This is a common technique for language definition in MDE [9] that facilitates language extension and evolution. MDE manipulates models (i.e. instances of metamodels) mainly using standard transformation languages. In this way, engineers do not need to devote effort to the low-level details of model processing, but just to define the transformations for the different purposes. In the case of ATCAS-IL, transformations consider the semantics of the DSL according to AT and agent research. These transformations generate code, for instance, for checking contradictions as [14] and simulation as [12]. This kind of approach has already been successfully tested in other domains [17].

The remainder of the paper further explains the elements in this introduction. Section II presents AT and the case study that guides the discussion in this paper. Section III provides a brief introduction to OperA. ATCAS-IL is introduced in section IV, and section V applies it to the analysis of the case study. Section VI compares the presented approach with existing ABM works. Finally, section VII discusses some conclusions about ATCAS-IL.

II. Activity Theory

The Activity Theory (AT) [13] is a socio-psychological paradigm for the study of human behaviour. It focuses on the mutual dialectics between people and their physical and social environment: the environment shapes human actions and their execution, and is also changed by these same actions. Hence, human acts cannot be analyzed independently of their context. These contextualized acts constitute the minimal meaningful unit of analysis and are called *activities*.

An *activity* [18] is a transformation process driven by people's needs. These needs are satisfied with an *outcome* produced transforming an *object*. Any element used in this process is a *tool*. The active component that carries out the activity is the *subject*. Subjects with a set of common social meanings constitute a *community* [16], which represents the socio-historical context of the activity. Two bodies of social constructions mediate the relationships of communities in the activity: *rules* with the subject and the *division of labour* with the object. Both of them contain similar elements, such as knowledge, implicit assumptions or norms. The key difference is the focus. The division of labour regards task specialization in the community through aspects such as power relationships, goal decomposition or the assignment of responsibilities. On the contrary, rules are guides and constraints not targeted specifically to the activity but affecting it, such as group beliefs, country laws or accepted scientific theories. The different elements can be both physical and mental, so AT considers both types of activities with a unifying analysis. All these elements make up the context of an activity, which is named its *activity system*. Its traditional depiction [16] appears in Fig. 1.

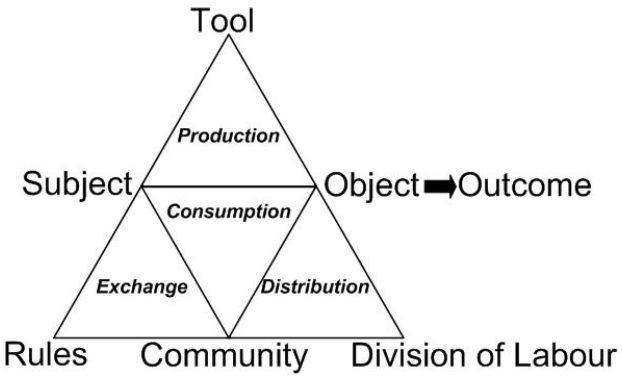


Fig. 1. AT depiction of an activity system.

Activity systems always exist in neighbourhoods of interconnected activity systems linked by shared elements. The execution of an activity produces outcomes that become the artefacts (e.g. subject, tool or rules) needed to execute other activities. Subjects carry out activities in these networks following their own rationality.

AT also considers the hierarchical decomposition of activities. *Activities* pursue high-level *objectives* that meet people needs. These activities are executed through sequences of *actions*, which try to achieve low-level *goals*. These goals do not satisfy by themselves any need, but they contribute or are part of higher-level goals. In their turn, actions are implemented through *operations* that depend on the specific state of the *environment*.

The evolution of activity systems over time depends on their inner *contradictions*. These contradictions are conflicts between the elements in the networks of activity systems, and they can appear both inter and intra systems. Subjects try to remove contradictions through the evolution of the involved activity systems, commonly generating new tensions that produce further evolution.

As an example of AT analysis, this paper considers the work in [19] about the Challenger crash. In 1986, the Challenger shuttle exploded shortly after launching. This accident has been frequently used as a case study about engineer ethics, communication and group thinking. According to Holt and

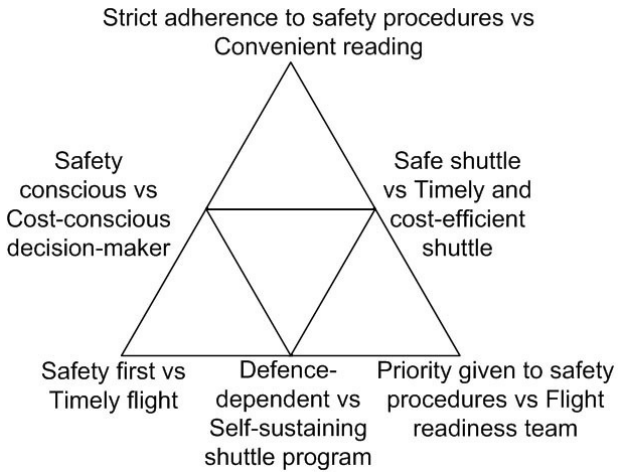


Fig. 2. Shuttle flight readiness activity system in the 1980s adapted from [19].

Morris [19], NASA began in the early 1960s as a tightly coupled set of subsystems aimed at space research and development. Over the 1980s, it became the home of several competing groups. At the same time, the lowering perceived value of the space research in public opinion put the agency under increasing pressure to cut down costs. NASA could not satisfy all its opposing goals, so a critical situation made it inevitable to violate some of them.

The analysis focuses on the readiness activity where the decision is taken to launch the shuttle or not. It discusses how the inner contradictions present in the NASA's organization culture at the end of the 1980s led to the catastrophe, how these contradictions generated the new and culturally more advanced activities in today's organization, and the still existing contradictions in the current forms of activities. Though the case is inevitably simplified, it synthesizes several complex settings and lengthy sources of information. For instance, the report of the Presidential Commission about the accident comprehends 5 volumes of documentation and analysis[1].

Fig. 2 depicts the readiness activity system. The nodes of the triangle show the opposition between the two perspectives about the NASA. The initial NASA from the time of the Cold War had priorities focused on research and safety, and almost unbounded resources. The new cost-aware agency competes with other agencies all over the world for the launching of satellites. The roots of this conflict appear in the object, which is the launching of a shuttle. It opposes the *safe shuttle* need to the *timely and cost-efficient shuttle* need. This opposition permeates all the other elements in the activity system. For the *subject*, it opposes the safety and the cost consciousness of the decision maker. The tools include information, devices and procedures supporting the activity. In this case, they are focused on the safety procedures, confronting a strict adherence to them with a convenient reading, which is only concerned about critical errors and relies on the low probability of certain failures. The social context of the activity shows this same duality. The community includes the NASA *defence-dependent* with high funding of the Cold War, and the one of decreasing funding in the 1980s aimed at getting a *self-sustaining shuttle program*. The division of labour regards work organization. In this case, it defines the acceptable tradeoffs between prioritizing security and the *flight readiness team*. These norms emerge from a wider context of rules where NASA was pushed to have successful and cost-effective *timely flights* more than to guarantee *safety first* beyond any reasonable risk.

These intra-node tensions scale up to contradictions between nodes. When the decision-maker needed to decide about launching, he/she was confronted with the dual nature of its community, tool and rules. Whatever decision this engineer made, it could lead to failure in some of the objectives. This situation is known as a double-bind contradiction [20]. In it, the system trying to satisfy opposed goals is under growing pressure that it can only overcome through its evolution.

This paper uses this case study to illustrate the application of ATCAS-IL to model complex organizations and subsequently analyze particular features of them. Specifically, it shows how its formal specification method supports the automated identification and analysis of the inherent contradictions of the NASA organization.

III. The OperA framework

AT considers that the integrated analysis of the individual and social contexts of activities is key for their understanding. Thus, its formalization as a DSL needs to include and define precisely the behavioural aspects common to the artefacts present in both scopes. For this reason, we have chosen OperA as the basis for ATCAS-IL.

OperA [15] is a framework for the specification of agent organizations. It considers the requirements regarding the structure, norms and external behaviour of these organizations and their members. Being focused on specification, OperA makes no assumption about the internals of the agents implementing the organization, and represents interactions through landmark patterns. These provide abstract representations of families of protocols. As such, OperA specifications enable different actual instantiations. Three models specify its organizations.

The Organizational Model (OM) describes the organization requirements using as main concepts, roles and scenes. The definition of a role includes its objectives and their decomposition, the norms applicable to that role, and the rights or capabilities it has. Scenes describe interactions between roles. They specify the participant roles, the ordering of actions through landmarks, the norms governing the scenes, and the results of their execution. This description makes use of information about the domain and the general normative structure applicable in the system.

The Social Model (SM) specifies the activity of agents in the society. Agents are executable entities able to enact the OM roles. OperA establishes constraints on the behaviour of agents using social contracts, but not their implementation.

[1]http://history.nasa.gov/Shuttlebib/ch7.html

These contracts indicate under which conditions an agent will play a role, and are used, for instance, to describe the benefits that the agent obtains, additional constraints or potential plans.

Finally, the Interaction Model (IM) describes how agents enacting roles participate in the scenes. It specifies the actual execution of the scene through a sequence of actions and adds additional norms for it.

OperA uses the Logic for Contract Representation (LCR) for the specification of contracts and norms in agent organizations. As stated in the introduction, it is an extension of deontic temporal logic that includes the *stit* operator (i.e. an agent *sees to it that*), the temporal operators of the branching temporal logics, and several deontic operators (i.e. obligation, permission and prohibition). Its expressions can be linked to deadlines corresponding to events or observed situations.

The OperA framework includes a graphical tool, OperettA [21], an IDE built as a plug-in of Eclipse. OperettA has facilities for the visual specification of organizations, syntax checking based on the OperA metamodel, static analysis, (e.g. dependent roles have to have a social link) and normative verification (e.g. inconsistencies between objectives and norms).

The formal logical semantics of OperA form a basis for the formal verification of the new specification language ATCAS-IL, as the paper shows in the next section.

IV. ATCAS-IL

ATCAS-IL extends the underlying OperA framework in several aspects determined by three main requirements. The first one is the foundation of ATCAS-IL in AT. It makes activities the focus of the analysis, which takes in turn the environment becoming a first class citizen of the specifications. The second one is the aim of ATCAS-IL for social analysis through automated tools. This requires specifying not only the constraints for the behaviour of the organization, but also some aspects of the actual behaviour of its members. The last requirement is the need of providing support to deal with the complexity of the description of social systems. Since these specifications account for a wide variety of issues and individuals, the language has to provide mechanisms to work with them at different levels of abstraction.

An activity is an act contextualized by its activity system [18]. The activity system includes subjects using tools to transform objects into outcomes, in the context of communities that specify the division of labour and the general rules applicable to that activity. While subjects, communities, and rules and division of labour can be roughly approximated by roles/agents, groups of these, and norms respectively, there are no suitable OperA concepts to describe activities, objects and tools.

Objects and tools are elements of the socio-physical environment. In ATCAS-IL, an *artefact* represents a general element of the environment. Following a widespread perspective in MAS [7], ATCAS-IL characterizes artefacts in terms of an internal state, the operations available to manipulate them, the events they can generate, and the knowledge and norms about their use. An *operation* is a basic act that can be executed when certain constraints are satisfied, and that generates some results. Both the artefact state, and the constraints and results of operations are defined with OperA formulae. *Environments* group artefacts adding specific norms.

Algorithm 1 Definition of the basic artefact. Bold words are ATCAS-IL (or OperA) keywords.

```
Artefact ( basic-id,
    Capabilities:,
    Knowledge:
        state(basic-id, non-available) or
        state(basic-id, available) or
        state(basic-id, blocked)
    Rules:
        FORBIDDEN (rule-0,
            Environmental-contract(
                Instance: _, Artefact: basic-id, Clauses: _ )
        )
        OBLIGED (rule-1,
            state(basic-id, available) BEFORE
            state(basic-id, blocked)
        )
        OBLIGED(rule-2,
            ( blockedBy(Ac1, basic-id) and
            blockedBy(Ac2, basic-id) ) ->
            ( Ac1 = Ac2)
        )
)
```

Algorithm 1 shows an example of the specification of an artefact. It introduces as knowledge the different states in which an artefact can be: *non-available* (i.e. non-created and therefore non-usable in activities), *available* (i.e. usable by activities) and *blocked* (i.e. used exclusively by an activity). Several specific rules limit the behaviour of this basic artefact: *rule-0* states that this is an abstract artefact and therefore no instance of it can be created; *rule-1* indicates that the artefact must be created (i.e. available) before it can be locked for exclusive use (i.e. blocked); *rule-2* states that an artefact cannot be blocked at the same time by different activities. All the artefacts in the case study satisfy this basic specification.

Communities control the access of subjects to *environments*. A community is a group of subjects that share social meanings and artefacts. ATCAS-IL defines them as an extension of the concept of role group in OperA. It comprehends several sets of elements: roles representing the subjects; the environments it has access to; accessibility rules mediating between subjects and artefacts; and knowledge and norms applicable in it. Roles are standard OperA roles whose capabilities are represented as operations. The accessibility rules indicate that when the states of the role and its environment meet some conditions, the role can use the operations of the artefact and be aware of its events.

The previous elements are put together in *activities*, whose definition following AT [16] includes the following elements:

- *Subjects* indicate who are in charge of executing the

activity. OperA roles describe them. *Objects* and *tools* represent the elements of the environment affected by the execution of the activity. *Artefacts* represent them.
- *Communities* determine how the subjects can access the artefacts related with the activity system.
- *Rules* and *division of labour* establish the norms that provide constraints for the execution of the activity.
- *Outcomes* of the activity are modelled as OperA formulae that become true after the activity execution.
- *Patterns* establish intermediate steps in the execution of activities.

As seen in Section II, activities are hierarchically decomposed in actions, which are further decomposed into operations. ATCAS-IL only considers operations in its refinement of activities. Patterns allow specifying the preferred order of generation of the results of operations.

The previous elements (i.e. communities and activities) define general patterns of behaviour, and appear at the level of OperA OM. The elements that actually implement them are the *agents* for the subjects and the *instances* for the artefacts. SM define these elements. An agent is characterized by its objectives, capabilities (i.e. operations), applicable norms and priorities about objectives. Objectives are states of the world that agents pursue defined as OperA formulae. Priorities determine the objectives that agents prefer trying to achieve when their states and that of the accessible environments meet certain conditions. These priorities provide the means to further describe the dynamic behaviour of systems. OM indicate existing objectives and their satisfaction conditions, and the priorities establish a simple way to choose the objective to attempt when several alternatives are available. The instances of artefacts are described in terms of their capabilities, and the knowledge and rules about them.

Finally, IM define for each activity the specific agents and instances that act in it, and sequences of operations to execute it. These sequences of operations must be able to satisfy the patterns included in the definition of the activity.

The division between the conceptual definition of the model (through communities and activity systems) and its actual realization (with instances, agents, and interactions) facilitates two goals in ATCAS-IL. First, it isolates the researchers' hypotheses, which appear at the conceptual level of OM, from the variability of individual elements, which SM and IM define. This allows managing the heterogeneity of individuals in populations, which is relevant to check that the obtained results are really a consequence of the hypotheses and not of specific configurations of the population. Second, it enables the independent evolution of the definition and implementation of subjects and artefacts. This evolution is required to adapt the realization of agent-based models to the requirements of different applications, as the already mentioned centralized supervision and different levels of complexity in the implementation of the individual agents depending on the population size.

ATCAS-IL also incorporates mechanisms to manage the models complexity. ABM [5] is increasingly being accepted as a tool for social research due to the inadequacy of analytical and traditional computational models to deal with large, heterogeneous and non-linear systems. Nevertheless, ABM models can also become quite complex, in cases that: involve several types of agents; where individuals evolve in different ways; or, large bodies of knowledge and rules affect the society behaviour. The management of such complexity requires that approaches incorporate abstraction mechanisms. Besides the decomposition of specifications in several models present in OperA, ATCAS-IL includes an extension mechanism for modelling primitives and the capability of defining macros.

The extension mechanism allows indicating that a given sub-concept extends a super-concept by including all its attributes. This mechanism is available for artefacts, environments, roles, communities, activities, activity systems, agents and instances. It reduces the size of the specifications and allows building conceptual hierarchies of concepts. Such hierarchies highlight the common features of concepts, allow their incremental description, and facilitate the definition of exceptions. For instance, they allow indicating that subjects usually comply with the rules of their communities, but a minority of them are going to break the law, that is, ignoring some of those constraints.

The second mechanism is the definition of macros. Deontic logic, used as formal basis of OperA, does not belong to the standard background of social science researchers, so they need the support of experts to use it. In order to improve the researchers' autonomy when using ATCAS-IL, their experiences are gradually crystallized in a set of tailored macros suitable for a domain. That is, researchers initially determine what they want to specify and experts in logics help them to describe these operators with basic logic primitives. After some studies, this early joint effort defines a researcher-friendly specialization of the DSL for that domain. That specialization includes macros for the most commonly used operators and concepts. Then, researchers can specify models on their own, and only need experts in logics to describe new and unusual properties.

V. Case study: NASA shuttle

This section uses ATCAS-IL to analyze the conflicts existing in the **flight readiness** activity system described in Section II. The AT description of this activity system has been presented in Section II and is depicted in Fig. 2. As stated there, the conflicts emerge from the NASA duality between an agency focused on research and safety, and the modern one trying to reduce costs. It affects the shuttle launching opposing a total guarantee of safety and a reduced one but with a timely launching.

The first step of this analysis is the formalization of the activity system using the primitives presented in Section IV. Algorithm 2 shows a simple mapping for it. The *flight-readiness-activity* extends the *basic-activity*, which includes basic information about activities. Its object has been simplified to focus on the contradictions coming from the community. The two communities reflect the duality of goals in NASA between the

Algorithm 2 ATCAS-IL activity system for Fig. 2.

Activity basic-activity (flight-readiness-activity,
Subjects : decision-maker,
Objectives : ,
Objects : Shuttle,
Tools : safety-procedure(Shuttle),
Communities : public-funded-agency,
autonomous-agency,
Rules : safety-first, timely-flight(Shuttle),
DivisionOfLabour : priority-to-safety-procedures,
flight-team-readiness,
Outcomes : shuttle-flight(Shuttle),
Patterns :
)

Algorithm 3 Communities of the flight readiness activity system.

Community (basic-community,
Roles : decision-maker,
Environments :,
Knowledge :,
Rules : **OBLIGED**(rule-3, cost(W) **and** safety(X) **and** funding(Y) **and** income(Z) **and** (W + X) < (Z + Y)),
Role-dependencies :,
Artefact-accessibility :
)
Community basic-community (public-funded-agency, ...)
Community basic-community (autonomous-agency,
Roles : cost-conscious-decision-maker : decision-maker,
Environments :,
Knowledge : funding(0),
Rules : ,
Role-dependencies : ,
Artefact-accessibility :
)

public-funded-agency and the *autonomous-agency*. The knowledge related with the artefacts in this activity is described as part of their definitions.

Algorithm 3 shows a partial definition of the communities involved in the previous activity system. The *basic-community* contains the role *decision-maker*, who is the subject of the activity, and a rule that states that the agency must run balancing its expenses (i.e. functioning and safety costs) and the incoming money (i.e. funding and incomes). As there are

Algorithm 4 Definition of rules.

DEF safety-first(Shuttle) = safety(Shuttle, X) **and** limit_safety(Y) **and** X < Y

DEF timely-flight(Shuttle) = expected_launch_time(Shuttle, T1) **and** launch_time(Shuttle, T2) **and** T2 < T1

Algorithm 5 Definition of roles.

Role (decision-maker,
Objectives : take-decision(Shuttle),
Sub-objectives :,
Rights : Check-readiness(Shuttle),
Decide-launch(Shuttle),
Knowledge :,
Rules :
)
Role (safety-conscious-decision-maker,
Objectives :,
Sub-objectives :,
Rights : Check-safety(Shuttle),
Knowledge :,
Rules : safety-first
)
Role (cost-conscious-decision-maker,
Objectives :,
Sub-objectives :,
Rights : Check-costs(Shuttle),
Knowledge :,
Rules : timely-flight
)

Algorithm 6 Implementation of the activity system.

Interaction-contract (
Activity : flight-readiness-activity,
Parties : (decision-maker : engineer),
Environment : (shuttle : challenger),
(safety-procedure : frr),
(public-funded-agency : nasa),
(autonomous-agency: nasa),
Clauses :,
Protocol : Check-readiness(challenger),
Check-safety(challenger), Check-costs(challenger),
Decide-launch(challenger)
)

no constraints about artefact accessibility, all the roles have granted full access to all the artefacts of the community. Depending on the type of agency, the specification adds more rules to its description. For instance, the *autonomous-agency* should run without public funding, which is asserted as *knowledge* of that community. Note that the specification of the *autonomous-agency* indicates that it constrains the general *decision-maker* to a *cost-conscious-decision-maker*, so it does not add a new subject but specializes the existing one.

The definition of the rules is illustrated with *safety-first* and *timely-flight* in Algorithm 4, which are also examples of the use of macros introduced by keyword *DEF*. The rule *safety-first* states that the probability of failure cannot be over a given limit, and *timely-flight* indicates that launchings must adhere to the expected planning.

The last type of elements to define is the roles involved

in the activity system. Algorithm 3 introduced the *decision-maker* as the subject of the flight readiness activity system. As also seen in Fig. 2, it is specialized in the safety-conscious and the cost-conscious decision makers. The definitions of these roles can be seen in Algorithm 5. They have available several operations, such as *check-readiness* in the *decision-maker*. All of them produce pieces of information according to the available state of the shuttle. The operation *decide-launch* takes this information to approve or deny the launching of the shuttle, which satisfies the objective *take-decision*.

Algorithms 3, 4, and 5 describe the general behaviour of the activity system. Then, the SM specifies the actual individuals implementing it, and the IM how they carry out the activity.

In this case, the SM just considers one instance per artefact and one agent for the decision maker (in this case an *engineer*) playing the two possible roles, the safety-conscious and the cost-conscious decision maker. This models the kind of situation that faced NASA engineers about launching the shuttle in the presence of non-optimal conditions.

The final element of the specification is the IM in Algorithm 6. It uses the previous instances to implement the activity system. The protocol establishes the sequences of steps that the roles can perform to execute the activity. The operations have been already discussed in this section.

The verification of the previous specification is the second step of the analysis with ATCAS-IL. It finds out the double-bind contradiction pointed out in [19]. When the engineer needs to abort the launching because it does not meet the *safety-first* condition, there is a violation of the *timely-flight* condition as no launching time is scheduled before the planned time. If the engineer decides launching anyway, there is a violation of the *safety-first* condition. This inability to take an action without violating constraints corresponds to the double-bind contradiction [20]. Thus, the verification has been able to automatically find the contradiction and to point out the conflicting properties. Details on the analysis process can be found in [17].

The presentation of the case study has used a textual specification with ATCAS-IL. OperettA supports visual modelling as long as the corresponding Eclipse models with the abstract and concrete syntaxes of the language are available. This step is required to complete a DSL suitable for social science researchers. They can easily grasp the primitives related with AT concepts, as they correspond to a structured textual representation of activity systems. However, the deontic logic used to specify the low-level details do not belong to their standard background. The use of macros and a graphical notation can reduce their difficulties to use it in their models.

VI. Related work

The field of ABM shares with general agent research many of its limitations. The review of Gilbert and Terna [5] about the implementation of ABM points out the lack of conceptually well-defined blocks for modelling and implementation guidelines.

ABM applies models to a large extent as a conceptual tool with only a swallow agreement about what an agent is [3]. Just to discuss some examples, the research in [11], [22], [6], [12] can be considered. The agents in [6] are modelled as sets of simple variables and rules to modify them. There is a set of external parameters not modifiable by agents, and a set of internal parameters that agents can modify with actions. Actions are rules triggered by certain conditions. There are also information links between agents that indicate when changes in their variables are propagated to other agents. The model can consider some noise in the communications to allow non-modelled environmental effects. Works as [22] have a more complex representation of the involved agents. Their condition-action rules include symbolic representations of elements such as goals, tests or capabilities, being therefore closer to traditional MAS. However, they do not represent key elements in MAS like the society, norms or decision making. Some researchers [11], [12] advocate the use of common MAS for ABM. This approach allows for the richer and more abstract modelling of individuals, but overlooks several relevant facts. The first is that MAS abstractions come from software development, and therefore they are not well-suited for social researchers. The second is that ABM usually deals with huge populations that cannot be implemented with computationally expensive agents.

ATCAS-IL aligns with research that promotes MAS for ABM, as it allows making modelling as complex as required. However, its language foundation is in AT, and thus in Social Sciences instead of AOSE. Moreover, ATCAS-IL relies on automated transformations of models to manage the implementation and its tradeoffs. This is an approach already pointed out in [12], though it proposes the use of programming languages instead of model transformation languages [9] as ATCAS-IL does. This last approach is intended to be closer to the concepts of the domain.

The issue of the lack of implementation guidelines was already mentioned in the introduction. ABM methodologies focus on capturing the information of the social systems at a very abstract level, but disregard how to migrate from these conceptual models to computational ones [3]. From the already mentioned works, [11], [22], [6] only consider the general features of agents in their models and the results of their simulations. The absence of general implementation details suggests that they rely on specific implementations for the problem at hand. This constitutes a major problem to validate the models, as it is not easy to know whether some of the results come from unintended features of the coding [4]. Approaches emerging from MDE and promoting the automated transformation of models [11], [12] facilitate to some extent this validation, as the same transformations are applied to different models. Nevertheless, the full validation of models in these example approaches still requires a complete understanding of the program code involved in the transformation. For this reason, model transformation languages [9], which work at the level of models, seem a more suitable approach for this task. Nevertheless, complete development

processes for ABM that guide researcher in the modelling process are still an open issue.

VII. Conclusion

This paper has introduced ATCAS-IL as part of the ATCAS framework for ABM. ATCAS-IL is a DSL to represent social constructions. It is based on two main sources: the socio-psychological AT to define its conceptual framework; the OperA framework for agent organizations for its formal definition. This foundation pursues two objectives. First, it tries to increase the autonomy of social researchers in ABM by directly applying their own concepts. This reduces the misunderstandings that inevitably appear when researchers need to rely on engineers without a background in Social Sciences for the modelling. Second, the formal definition of the language enables the automated processing of its models. ATCAS-IL proposes making these transformations through standard model transformation languages. This implies that the transfer of information between models is partly specified at the level of abstraction of models, and not with code. It is also expected to improve comparability between models: the application of the same transformations to different models of the same hypotheses should generate equivalent results.

The paper has illustrated the use of ATCAS-IL with the problem of the identification of the contradictions that led to a well-known failure in the space shuttle program. While the original AT work relies on the human analysis of data, a suitable model allows the automated discovery of the problem and its potential reasons.

ATCAS-IL has currently three main limitations. First, the domain-specific primitives are not enough to model a complete system. As the case study illustrates, low level-details have to be expressed with logics, which are not suitable for social researchers. Ongoing work is intended to determine the additional primitives required in the language according to AT and agent research. AT also considers recurrent social patterns [16], [14] that can be described with reusable macros. Second, textual specifications are too verbose given the amount of details required in the models. The extensions of the DSL can help to solve this issue. Besides, the use of the visual modelling capabilities of OperettA can simplify the development of the specifications, hiding the repetitive details of modelling. Third, facilitating the development of the automated model transformations is still an open issue, as social researchers are not expected to be experts in transformation languages. Approaches based on the automated generation of transformations from model prototypes are a potential solution.

Acknowledgment

This work has been done in the context of the project "Collaborative Ambient Assisted Living Design (ColoSAAL)" (grant TIN2014-57028-R) supported by the Spanish Ministry for Economy and Competitiveness, the research programme MOSI-AGIL-CM (grant S2013/ICE-3019) supported by the Autonomous Region of Madrid and co-funded by EU Structural Funds FSE and FEDER, and the "Programa de Creación y Consolidación de Grupos de Investigación" (UCM-BSCH GR35/10-A).

References

[1] M. W. Macy and R. Willer, "From factors to actors: computational sociology and agent-based modeling," *Annual Review of Sociology*, vol. 28, pp. 143–166, 2002. [Online]. Available: http://www.jstor.org/stable/3069238

[2] R. Axelrod, "Advancing the art of simulation in the social sciences," in *Simulating social phenomena*. Springer, 1997, pp. 21–40.

[3] A. Drogoul, D. Vanbergue, , and T. Meurisse, "Multi-agent based simulation: where are the agents?" in *Multi-Agent-Based Simulation II*, vol. 2581. Springer, 2003, pp. 43–49.

[4] R. L. Axtell and E. J. M., "Agent-based modeling: understanding our creations," *The Bulletin of the Santa Fe Institute*, vol. 9, no. 2, pp. 28–32, 1994.

[5] N. Gilbert and P. Terna, "How to build and use agent-based models in social science," *Mind & Society*, vol. 1, no. 1, pp. 57–72, 2000.

[6] A. Pyka and G. Fagiolo, "Agent-based modelling: a methodology for neo-schumpeterian economics," *Volkswirtschaftliche Diskussionsreihe*, vol. 272, pp. 1–26, 2005. [Online]. Available: http://www.wiwi.uni-augsburg.de/vwl/institut/paper/272.pdf

[7] B. Henderson-Sellers and P. Giorgini, Eds., *Agent-oriented methodologies*. IGI Global, 2005.

[8] N. M. Gotts, J. G. Polhill, , and A. N. R. Law, "Agent-based simulation in the study of social dilemmas," *Artificial Intelligence Review*, vol. 19, pp. 3–92, 2003.

[9] R. France and B. Rumpe, "Model-driven development of complex software: a research roadmap," in *Proceedings of the 2007 Future of Software Engineering Conference (FOSE 2007)*. IEEE Computer Society, 2007, pp. 37–54.

[10] A. van Deursen and J. Visser, "Domain-specific languages: an annotated bibliography," *ACM Sigplan Notices*, vol. 35, no. 6, pp. 26–36, 2000.

[11] S. Hassan, R. Fuentes-Fernández, J. M. Galán, A. López-Paredes, , and J. Pavón, "Reducing the modeling gap: On the use of metamodels in agent-based simulation," in *Proceedings of the 6th Conference of the European Social Simulation Association*, 2009, pp. 1–12.

[12] C. Sansores, J. Pavón, and J. J. Gómez-Sanz, "Visual modeling for complex agent-based simulation systems," in *Multi-Agent-Based Simulation VI*, vol. 3891. Springer, 2006, pp. 174–189.

[13] L. S. Vygotsky, Ed., *Mind and Society*. Harvard University Press, 1978.

[14] G.-S. J. Fuentes-Fernández, R. and J. Pavón, "Managing contradictions in multi-agent systems," *IEICE Transactions on Information and Systems*, vol. E90-D, no. 8, pp. 1243–1250, 2007. [Online]. Available: http://search.ieice.org/bin/summary.php?id=e90-d_8_1243

[15] V. Dignum, F. Dignum, , and J. Meyer, "An agent-mediated approach to the support of knowledge sharing in organizations," *The Knowledge Engineering Review*, vol. 19, no. 2, pp. 147–174, 2005.

[16] Y. Engeström, Ed., *Learning by expanding: an activity-theoretical approach to developmental research*. Orienta-Konsultit, 1987.

[17] P. Moreno-Ger, S.-R. J. Fuentes-Fernández, R., and B. Fernández-Manjón, "Model-checking for adventure videogames," *Information and Software Technology*, vol. 51, no. 3, pp. 564–580, 2009.

[18] A. N. Leontiev, Ed., *Activity, Consciousness, and Personality*. Prentice-Hall, 1978.

[19] G. R. Holt and A. W. Morris, "Activity theory and the analysis of organizations," *Human Organization*, vol. 52, no. 1, pp. 97–109, 1993.

[20] G. Bateson, Ed., *Steps to an Ecology of Mind*. Ballantine Books, 1987.

[21] D. Okouya and V. Dignum, "Operetta: a prototype tool for the design, analysis and development of multi-agent organizations," in *AAMAS 2008 demo papers*. International Foundation for Autonomous Agents and Multiagent Systems, 2008, pp. 1677–1678.

[22] M. K.-K. T. Murakami, Y. and T. Ishida, "Multi-agent simulation for crisis management," in *Proceedings of the IEEE Workshop on Knowledge Media Networking*. IEEE Computer Society, 2002, pp. 135–139.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1461–1466
DOI: 10.15439/2016F424
ACSIS, Vol. 8. ISSN 2300-5963

Token-based Autonomous Task Allocation in Flocking Systems

Andras Kokuti, Vilmos Simon and Bernat Wiandt
Department of Networked Systems and Services
Budapest University of Technology
HU-1117 Magyar tudosok 2, Budapest
Email: {kokuti,svilmos,bwiandt}@hit.bme.hu

***Abstract*—There are serious contributions to the theoretical foundations of flocking systems, but there are only few systems which have the capability of autonomous task allocation, however, many use cases demand this functionality. The implementation of a task allocation algorithm could be a serious challenge even in a simulated environment due to the numerous problems arising from the nature of these systems.**

This paper proposes a novel algorithm to find the optimal allocation of heterogeneous agents to heterogeneous tasks by utilizing distributed auctions based on local peer-to-peer wireless communication and exploiting graph theory with a tree-based multicast protocol. The solution was tested over a number of different scenarios and compared to existing algorithms in order to measure and prove its capability in handling autonomous task allocation in different systems as well.

I. Introduction

THE FLOCKING phenomena (the notion of flocking is used as a synonym of collective motion), observed in various fields in nature (such as insect swarms, fish schools, herds of wildebeests, etc.), can be an appropriate solution in many engineering applications as well. Applications of flocking include massive mobile sensing in an environment [1]; parallel and simultaneous transportation of vehicles or delivery of payloads [2]; performing military missions, such as reconnaissance [3], surveillance [4], and combat using a cooperative group of Unmanned Aerial Vehicles (UAVs). A flocking group of robots can perform tasks like exploration of an area, autonomous navigation for deployment, surveillance, or search and rescue operations [5].

In contrast to the huge number of flocking algorithms already published in the literature, most of them handle the group as a whole, no dynamic and autonomous reconfiguration or partition is possible [6]. Several use cases require utilizing autonomous regrouping of the flock, where a subset of the flock can leave the group, move to a given destination and perform various tasks on the spot. Several important uses cases impose scenarios where the flock has a large set of nodes, but only a few are required to move towards a specific location, for example when monitoring the behavior of the crowd during mass events or when observing the condition of a dam at multiple points with cameras. In cases where there is only one operation center broadcasting commands, the selection of the subset of nodes acting on the command and routing those nodes to the event site needs to happen autonomously based on the implemented control algorithm and the peer-to-peer communication between the nodes. This is an autonomous task allocation problem, which is an NP-hard [7] combinatorial problem.

In this paper we propose an algorithm capable of choosing the optimal task allocation based on the actual requirements without any central supervision, relying only on local interactions of the nodes. The algorithm does not depend on the tasks or the requirements directly, since it can select the optimal allocation with respect to the criteria defined by the use case.

II. Related Works

Multi agent task allocation is a problem that arises where a number of agents are working together in the aim of achieving a common goal or task which is subdivided into a number of subtasks. The problem can be formulated as a Multiple Traveling Salesmen problem which is NP hard [8]. Many solutions try to approximate the optimal solution by relying on auctions.

There are fully centralized solutions, such as [9] that require a central task allocator to determine the optimal allocation of tasks for the whole system. This approach certainly has advantages, such as the optimal global cost can be calculated easily, but it is usually not feasible in a real life scenario, especially in scenarios mentioned in Section I, as a fully centralized approach requires complete and perfect communication between nodes and the central task allocator has to keep track of the state of the whole system including internal knowledge of each agent's state.

Some solutions utilize combinatorial auction [10](where bidders can bid on combinations of items) in a centralized manner in order to find the optimal task allocation such as in [11] and [12]. Combinatorial auction compared to centralized task allocation has many advantages, however, a few disadvantages as well. Due to the distributed cost computation, this method does not require a fully centralized task allocator to keep track of the internal state of each node in the system. However, it has the communication and computational disadvantages of a fully centralized approach. Combinatorial auctioning can provide globally optimal solution based only on local interactions but it is well known to be an exponential

978-83-60810-90-3/$25.00

algorithm, therefore, it does not scale well with the number of nodes which can result in extreme energy consumption [13].

Distributed auctions are used to exploit the peer-to-peer communication between nodes in [14] and [12]. This class of methods is the primary mechanism for providing intentional coordination between agents. Distributed auctions have the advantage of not requiring the auctioneer to have a model of each agent (for example a robot). They do not require full communication between the agents and the auctioneer either (which means the network graph does not have to be complete, only connected) and can be robust with respect to communication failures. However, they are inherently suboptimal when a complex global cost function is used.

Fully distributed approaches [15], [16] do not use auctions in a conventional manner, therefore, they do not require direct communication between nodes. Each robot in this case retains local control and chooses its own actions based on its observations of the environment. Cooperative actions however, can be achieved by estimating models of the other agents' strategies. Distributed approaches are robust to communication failures, but can result in worse performance than distributed auctions due to a lack of intentional coordination via an auctioneer.

After studying the listed solutions thoroughly, we have chosen the method of distributed auctions for our algorithm, as it does not require full communication between the agents and being robust to wireless communication failures. The initiator event can be triggered only on one agent, but it uses graph theory solutions to select the most appropriate local auctioneers that do the auctions in a parallel and distributed way. To convey the tasks and the bids between the agents our solution uses tokens which can be transferred via peer-to-peer communication through the whole network.

III. Proposed Distributed Task Allocation Algorithm

We are investigating the problem of allocating a subset of agents based on local events, such as when a node senses disturbances, hazards, etc. or global events: the control center orders nodes to perform tasks. The event can originate from a higher lever entity in the system but the allocation process has to be performed in a self-organized manner among the nodes. A global or local event can be associated with certain requirements, such as a minimum energy level required or various properties of the nodes: speed, agility, maximum operating altitude, presence/accuracy of sensors/actuators, etc. The goal then is to select a subset of the nodes with the best fit with respect to the requirements associated with the event.

A. Problem formalization

An event triggers a request in the system that can be translated into a logical entity called the token. In our terminology a token is a simple data structure which contains the requirements (e.g. in a key-value map) associated with the event and nodes can manipulate it during the allocation process. This is very important, since a node needs to add its own data to the token in order to indicate that it has previously seen the token. A token can be transferred between nodes via wireless communication. It follows that a token is the subject of auctions, since every node is competing for the tokens based on the requirements in the token and the node's properties. Therefore, the selection process can be viewed as a highly distributed auction, where the bidders are the agents in the system and the "goods" are the tokens. The main benefit of the token representation of tasks is that the initial abstract problem can be mapped to a more formal problem which now can be solved by using graph theory and other mathematical tools. Thus the autonomous task allocation problem can be formalized as a token translation step and based on the token a distributed auction between the nodes. In the token translation step the algorithm should translate the requirements of the event to a transferable entity called token. And with distributed auctions (starting from the triggered agent) the nodes in the system can exchange the tokens in a controlled manner. At the end of the auction a task allocation can be done based on the bids contained by the token. In order to guarantee the optimal allocation of tokens to nodes, each node must bid on each token at least once, otherwise there could be a node in the system which did not bid to a token but would be the perfect fit.

In a distributed system limiting the number of local auctions is beneficial, since an auction can consume a huge amount of resources. In this case the auctioning algorithm is coupled with a flocking algorithm that depends on wireless communication to synchronize the state of the nodes such as position, velocity and heading. A naive solution would use no auctions at all but nodes that have the token could bid on it. This solution however involves unnecessarily broadcasted messages, as all tokens have to be transferred to potentially all nodes in the system. This approach is one extremity, where the number of auctions is minimized. Thus the algorithm should find an optimal trade-off between the number of the auctions and the required time to complete the allocation process. To minimize used resources and ensure that an optimal solution is found, our algorithm can be broken down to these general steps:

1) Select the auctioneer nodes based on the network graph
2) Build up a multicast tree in which the source is the node where the event was triggered and the destinations are the auctioneer nodes
3) Use the multicast tree to make communication more efficient in the process of allocating nodes to tokens

In the next subsection we focus on the first sub problem, since the second and third can be solved by enhancing traditional tree-based multicast algorithms (will be described in a later subsection).

B. The auctioneer selection algorithm

As we are considering mobile agents, the graph representing the network connections is not stable and not known a priori, thus some assumptions have to be made. We assume that each node has a local network table, which contains all other nodes in an adjacency matrix. This table is built up using the periodically received topology information (containing

the whole network from the sender's perspective) from the neighbors (similar to the distance-vector routing method [17]). Thus, every node transmits its local network table periodically (like the heartbeat packet in distributed systems) in order to ascertain the node is still operating (since a device can be broken down at any time because of many reasons) and to create a local copy form the whole network topology in all nodes. The next assumption deals with the structure of the network, since in our case the graph is undirected and connected before the task allocation phase. However after it can change, due to for example a task which requires that few nodes from the connected flock move away for a time and do dedicated tasks such as exploration and then join back to the flock. The undirected graph can be guaranteed by allowing only pairwise connections. Thus, if only one node can communicate with the other but the other can not (because of different sizes of transmission ranges) then this asymmetric connection will not be placed in the network matrix and hence the undirected communication graph is guaranteed.

Based on these assumptions auctioneer selection can be formalized and solved by using tools from graph theory. It is a selection problem and the goal is to select the minimum number of auctioneer nodes in the system. Let G denote the graph representation of the whole network, and $G = (V, E)$, where V are the vertices (the nodes) of the graph and E is the set of edges, representing the communication links between the nodes. Thus, the problem is to select a subset (V') from V, which has the minimum number of elements (minimum cardinality set) and can cover all the nodes in the network with the edges covered by the vertexes in the subset. More formally, we are seeking a subset V' of V, such that for all $u \in G$ there exists $(u, v) \in E$, where $v \in V'$ and V' is the minimum cardinality set. This set is very similar to the minimum vertex cover, however, in this case the goal is to cover all nodes instead of all edges.

Assume that every vertex has an associated cost of $c(v) \geq 0$. Then the problem discussed earlier can be formulated as the following integer linear program (x_v is a label which means whether the node is chosen as the element of the minimal set):

$$\text{minimize} \sum_{v \in V} c(v) * x_v \tag{1a}$$

$$\text{subject to} \sum x_u + x_v \geq 1 \ \forall v \in V(G), \ \forall u : (u, v) \in E(G) \tag{1b}$$

$$x_v \in \{0, 1\} \ \forall v \in V(G) \tag{1c}$$

The formulation as an integer linear program can be done in polynomial-time (since from the adjacency matrix the equations can be formulated, processing the whole matrix in O(n^2). As the graph should be processed n times, the reduction can be done in O(n^3)). The 0-1 Integer Linear Programming (a special case of the ILP) problem is one of Karp's 21 NP-complete problems. Thus, we were able to reduce our problem in polynomial-time (i.e. Karp reduction, where $c(v) = 1 \ \forall v$) to an NP-complete problem, therefore, our problem is NP-hard. It follows straight-away from the fact that ILPs are NP hard.

Algorithm 1 Calculating the covering nodes

$V' \leftarrow \emptyset$ (the set of the selected nodes), and $V'' \leftarrow \emptyset$ (the set of covered nodes);
2: **while** $V \neq \emptyset$ **do**
$v \leftarrow$ the node with the maximum degree;
4: $V' \leftarrow V' \cup v$;
$V'' \leftarrow V'' \cup v \cup \forall u$ where $(u, v) \in E(G)$;
6: $E(G) \leftarrow E(G) \setminus (E(G(V'')) \ \cup \ \forall E(u, v), where \ u \in V \ and \ v \in V'')$;
$V \leftarrow V \setminus \forall u$, where $d(u) = 0$;
8: **end while**
return V'

This problem also contains the well-known minimum vertex cover problem as a special case, when in the subject we do not summarize all the neighbors instead one (thus subject to $x_u + x_v \geq 1 \ \forall (u, v) \in E(G)$).

Since the problem is NP-hard, we provide a constructive heuristic algorithm which finds only a suboptimal solution but does it in polynomial time. The main idea behind the algorithm is to select nodes with higher degree as they contain more previously uncovered nodes of the graph. Obviously the degrees of the nodes have to be updated after each selection since the selected node and all of its neighbors are covered. Thus, our algorithm can be described by the following pseudo-code:

Theorem: The Algorithm 1 does find a (sub optimal) solution in polynomial time.

Proof: It is easy to see that the above algorithm does find a solution, since it can stop only after Step 2, and after this step it always stands that $V'' \cup V$ equals to all the vertices in the graph, and it returns only if $V = \emptyset$, thus V'' (which is the set of the covered nodes) contains all the devices in the network. And it can be proved as well, that it stops in polynomial time as the algorithm steps in the loop (Step 1-2) at most n times, where n is the number of vertices in the graph ($n = |V(G)|$). To find the node with the highest degree from the adjacency matrix is O(n) and the update process of the matrix (after removing the vertices and the edges) is O(n^2). Thus, the running time is $O(n*(n+n^2))$ = $O(n^3)$. The $O(n^3)$ time relates to only the auctioneer candidate selection which runs locally on the triggered node and therefore, can be really fast when the adjacency matrices are stored in-memory.

It would be beneficial as well if an upper bound could be determined for the number of selected nodes, since it is a good indicator for the speed of the whole process (includes the distributed auction as well).

Theorem: An upper bound for the number of the covering nodes is $\lfloor |V(G)/2| \rfloor + 1$, where $V(G)$ is the set of the vertices in G.

Proof: It is easy to see, that if we can guarantee this upper bound on a spanning tree of the graph, then the bound is valid for the original graph as well. Since the original graph has more (or at least equal number of) edges than the spanning tree, the nodes we selected based on the spanning tree still

cover all vertices in the original graph. In a tree the set we are looking for is identical to the well known minimum vertex cover from graph theory (since in this case to cover all vertices we have to cover all the edges as well). From the Gallai theorem we know that in a graph the sum of the minimum vertex cover and the maximum independent set is equal to the number of vertices. In a tree the number of nodes in a maximum independent set is equal to or greater than $\lceil |V(G)|/2 \rceil$, since trees are bipartite graphs, thus, can be divided into two distinct sets and the vertices of either set are independent from each other. Therefore the number of nodes in the minimum vertex cover is equal to or less then $\lfloor |V(G)/2| \rfloor$. The additional 1 vertex is added to the upper bound because by using a heuristic algorithm in trees it may happen that it selects the wrong one from the two distinct sets (this happens only when the difference between the two sets is only 1).

In this section we have proposed an algorithm able to select a subset of vertices from a connected G graph covering all other vertices in the graph through their edges. We have proved that the algorithm does the selection in polynomial time and we could define an upper bound for the number of selected vertices. Thus, with the help of this algorithm we are able to determine the agents in the system where the local auctions have to be performed in order to find the optimal task allocation.

C. The tree-based multicast algorithm

We have seen previously that the main problem was divided into 3 sub problems and in Section III-B we have provided a heuristic algorithm to approximate the solution for the first subproblem. In this subsection our goal is to identify the appropriate tree-based multicast algorithm to solve the remaining two problems based on the aforementioned selection process result. Thus, the role of the multicast algorithm is to deliver the tokens from the source node to the selected local auctioneer nodes relying on a multicast tree.

The problem of finding a minimum cost multicast tree is well-known as the minimum Steiner tree problem. According to [18] this problem is also NP-complete, even when every edge has the same cost, by reduction from the exact cover by 3-set. Although it is widely assumed that a Steiner tree is the minimal cost multicast tree, it is not generally true in multihop wireless networks [19]. The problem of minimizing the cost of a multicast tree in an ad hoc network needs to be re-formulated in terms of minimizing the number of the data transmissions. Since in a broadcast medium, the transmission of a data packet from a given node to any number of its neighbors can be done with a single transmission, thus, the minimum cost tree is the one which connects sources and receivers by issuing a minimum number of transmissions, rather than having a minimal edge cost. However, finding this optimal tree in a wireless network is also NP-hard [19], therefore, a heuristic algorithm will be used in this case as well.

There are many heuristic algorithms to compute minimal Steiner trees: for instance, the MST algorithm in [20] provides a 2-approximation, and Zelikovsky [21] proposed an algorithm which obtains a 11/6-approximation. However, these algorithms try to solve the minimum cost tree in general instead in a broadcast manner. Authors in [19] proposed two heuristics (a centralized and a distributed one), both resulting in a tree with lower or equal data-overhead then the MST Steiner tree. Since in our case each node knows the whole network (because of the locally maintained adjacency matrices), the centralized solution seems to be an appropriate choice.

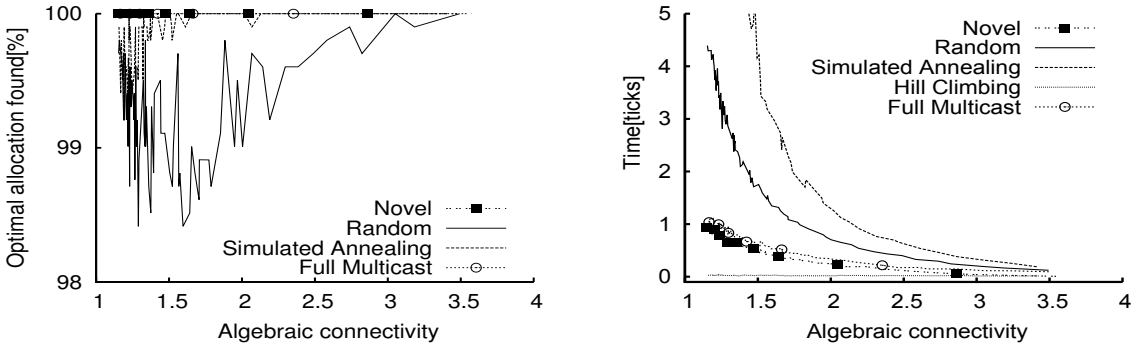
Thus, in the first step of the algorithm it selects the most appropriate auctioneer nodes, in the second step it creates a multicast tree where the multicast receivers are the nodes from the first step. This tree is not only usable for spreading the tokens from the source to the receivers (local auctioneers) but for the opposite direction as well. Therefore, if an auctioneer finishes the local auction, it transfers its token back on the multicast tree to the source. If a node on the path from auctioneer to source possesses two or more tokens at a time, it merges them in order to decrease the number of broadcast transmissions. Finally the first node (in worst case the source) which obtains all the tokens, immediately calculates the winner nodes and instructs them to carry out the tasks formulated in the token. Thus basically, only the auctioneer node selection step runs locally on a single node (where the task allocation event has been triggered) and then the algorithm uses distributed auction backed by multicast trees in order to increase the efficiency of the local auctions.

IV. Performance through Simulation

In the previous section we have proven that our algorithm can find the optimal allocation and does it in polynomial time, but our objective is to evaluate the effect of the proposed auction based task allocation algorithm in the context of a flock environment. Thus we have chosen to conduct a simulation study using the SimPy process-based discrete-event simulation framework [22] in a realistic scenario.

The simulations are based on a static context, thus the agents in our simulation do not move, since primarily our goal is to prove the algorithm can operate optimally in cases where the dynamics of the agents in the system are negligible compared to the communication speed. The nodes in the simulation communicate with each other via a radio interface, such as Wi-Fi, Bluetooth or anything else in a predefined communication range. It is important to note that in the simulation a perfect communication model is assumed as well, therefore, interference and packet loss are not considered.

To model the communication graph we used instances of the connected Watts and Strogatz graph [23], which is a random graph and can be parametrized to resemble the communication graph of flock. The communication network of a flocking system is a type of graph which is similar to a regular grid, therefore the valency of the nodes are nearly the same (since the agents in the flock are placed almost the same distance from each other). We generate a random connected Watts and Strogatz graph by setting the algebraic connectivity parameter as well. The algebraic connectivity of a graph is the second-smallest eigenvalue of the Laplacian matrix of the graph and



(a) For a clearer picture only the optimal or near to optimal solutions are depicted.

(b) Speed of the selected algorithms.

Fig. 1. Simulation results of the autonomous task allocation from all the aspects.

the magnitude of this value reflects how well connected the overall graph is. It depends on the number of vertices as well as the way in which they are connected. In random graphs, the algebraic connectivity decreases with the number of vertices, and increases with the average degree (and is greater than 0 if and only if the graph is connected). The simulations are performed with different algebraic connectivities in order to simulate different kind of flocks. Every graph with a unique algebraic connectivity value was tested 100 times and the results were aggregated in order to minimize the impact of outlier cases.

As it has been stated before, the requirements generated by an input event are translated into a token. However, our framework is capable of handling multiple types of requirements (and cost functions as well) since the algorithm does not restrict it. Thus it is really important to note that the proposed system does not depend on the type of the requirements or the cost function, instead these can be defined differently from case to case and the algorithms will adopt dynamically. In this paper we have considered a simple scenario where a token only contains a minimal energy level as requirement. An energy level is generated for each agent before the simulation, following a uniform distribution between 70% and 100%. In the generated event, the energy requirement is 65%, therefore every node can perform the task to ensure that a solution exists. In this configuration there are optimal and suboptimal allocations based on the generated energy level distribution.

Five different algorithms have been compared against each other in the simulator. The random walk algorithm, which transfers the token (the translated request) on randomly selected edges until some conditions (such as all nodes placed a bid or the token has been transferred at a given threshold) are fulfilled. The well known hill climbing algorithm, that transfers the token only to a node that has a higher bid than the current one, therefore, in our case it means that the tokens move towards nodes with higher energy. The simulated annealing algorithm: it tries to avoid suboptimal selection by allowing temporary worse states (so the selected node's bid is lower than the current highest). The algorithm uses acceptance probabilities of making the transition from a current node to a candidate one, which depends on the energy of the nodes and on a globally-varying time parameter, called the temperature.

A. Simulation Results

The results will be shown in two different aspects. The first is the main focus of the paper, that relates to the success of the task allocation. Measuring the success ratio of finding the optimal allocation among the nodes, 0% to 100% indicates the effectiveness of the used solution. And the second aspect focuses on the simulation time, which indicates the speed of the algorithm. Therefore, the goal is to minimize the time and select the most appropriate agents as quickly as possible.

Figure 1a presents the results from the first aspect. It has to be noticed that the hill climbing algorithm is not depicted since it performs really bad (around 40% at lower density and improves to around 80% in more connected networks) because its performance heavily depends on the location of the initial request event. And obviously the hill climbing algorithm produces the worst selections as it chooses nodes greedily, therefore it terminates at suboptimal nodes (local optimums). Therefore, for a clearer picture only solutions with close to the optimal performance are shown. When observing the percentage of the optimal coverage it could be seen that there are only two algorithms able to achieve the optimal selection regardless of the structure of the network graph: our proposed algorithm and the full multicast algorithm. The full multicast algorithm finds the optimal allocation in any case by definition, since it transfers the token to every node. The simulated annealing algorithm results in an almost optimal selection in all cases, but in really sparse environments (graphs with lower algebraic connectivity) it has some errors. The relatively good results of this algorithm is because it's parameters (such as the initial and the terminating temperature and the transfer rate) have been trained on many networks. The random solution (which passes the token to randomly selected nodes) also achieves a nearly optimal selection in all cases, since its input parameter (the number of token transfers) has also been trained on the same dataset. It is important to notice as well that the results

are improved when increasing the algebraic connectivity of the graphs, as higher algebraic connectivity values mean a denser and more connected network graph. Therefore local information becomes more global, in the extreme case of a fully connected graph each node has global knowledge about the whole graph.

Our final results are depicted on Figure 1b, and it shows the speed (the simulation time measured in ticks by the discrete simulator) of the different selection solutions. As it can be observed, all algorithms except the hill climbing are highly dependent on the algebraic connectivity. The relation between the algebraic connectivity and the speed is not linear but exponential, therefore, a less connected graph results in much slower selection. Hill climbing performs the best with respect to speed due to the greedy optimization: blindly passing tokens to agents with higher utility. On one hand this algorithm can terminate in local optimum instead of a global one (see Figure 1) but on the other hand it provides the fastest coverage. The other four examined solutions perform better if the underlying network graph is more connected, since in that case the network can be covered within less time. The second most efficient solution is our proposed algorithm, which can outperform the other three optimization techniques even by 30-40% in less connected environments. Based on the results from all the aspects, we can conclude that our solution finds the optimal selection in any network and does it in the fastest way among the algorithms which find the optimal or nearly optimal allocations. Only the hill climbing search technique does faster allocation but it can reach an optimal selection only in strongly connected flocking systems.

V. Conclusions

We have proposed and described a novel distributed auction based task allocation algorithm to enhance flocking systems. This task allocation algorithm exploits the network graph maintained locally by nodes in order to build a (sub) optimal multicast path from the source node to the chosen local auctioneers. Using this multicast tree our solution can find an optimal allocation based on the generated requirements.

Real-life flocking examples were used as a case study to evaluate how our novel algorithm can solve the distributed task allocation problem in flocking systems. We evaluated the performance based on multiple environments by varying the algebraic connectivity of the generated graphs. Our experimental results indicate that the proposed algorithm can find the optimal allocation regardless of the algebraic connectivity of the graph and does it relatively fast compared to the other examined solutions such as random walk, hill climbing, simulated annealing and the full multicast algorithm.

References

[1] H. M. La, W. Sheng, and J. Chen, "Cooperative and active sensing in mobile sensor networks for scalar field mapping," *Systems, Man, and Cybernetics: Systems, IEEE Transactions on*, vol. 45, no. 1, pp. 1–12, 2015. http://dx.doi.org/10.1109/tsmc.2014.2318282

[2] X. Wang, J. Qin, and C. Yu, "Iss method for coordination control of nonlinear dynamical agents under directed topology," *Cybernetics, IEEE Transactions on*, vol. 44, no. 10, pp. 1832–1845, 2014. http://dx.doi.org/10.1109/tcyb.2013.2296311

[3] X. Zhu, C. Wei, H. Duan, and Q. Li, "Some new results on bees-mechanism-based flock control with neighbors chosen by topological distance," in *Guidance, Navigation and Control Conference (CGNCC), 2014 IEEE Chinese*, pp. 2681–2686, IEEE, 2014. http://dx.doi.org/10.1109/cgncc.2014.7007591

[4] S. H. Semnani and O. A. Basir, "Semi-flocking algorithm for motion control of mobile sensors in large-scale surveillance systems," *Cybernetics, IEEE Transactions on*, vol. 45, no. 1, pp. 129–137, 2015. http://dx.doi.org/10.1109/tcyb.2014.2328659

[5] S. K. Lee, "Distributed space coverage for exploration, localization, and navigation in unknown environments," 2015.

[6] B. Wiandt, A. Kokuti, and V. Simon, "Application of collective movement in real life scenarios: Overview of current flocking solutions," *Scalable Computing: Practice and Experience*, vol. 16, no. 3, pp. 233–248, 2015. http://dx.doi.org/10.12694/scpe.v16i3.1099

[7] B. P. Gerkey and M. J. Matari, "Sold!: Auction methods for multirobot coordination," *Robotics and Automation, IEEE Transactions on*, vol. 18, no. 5, pp. 758–768, 2002. http://dx.doi.org/10.1109/tra.2002.803462

[8] M. Badreldin, A. Hussein, and A. Khamis, "A comparative study between optimization and market-based approaches to multi-robot task allocation," *Advances in Artificial Intelligence*, vol. 2013, p. 12, 2013. http://dx.doi.org/10.1155/2013/256524

[9] M. Koes, K. Sycara, and I. Nourbakhsh, "A constraint optimization framework for fractured robot teams," in *Proceedings of the fifth international joint conference on Autonomous agents and multiagent systems*, pp. 491–493, ACM, 2006. http://dx.doi.org/10.1145/1160633.1160724

[10] P. Cramton, Y. Shoham, and R. Steinberg, "Combinatorial auctions," 2006.

[11] M. Mito and S. Fujita, "On heuristics for solving winner determination problem in combinatorial auctions," *Journal of Heuristics*, vol. 10, no. 5, pp. 507–523, 2004. http://dx.doi.org/10.1023/b:heur.0000045322.51784.2a

[12] K. Zhang, E. G. Collins Jr, and D. Shi, "Centralized and distributed task allocation in multi-robot teams via a stochastic clustering auction," *ACM Transactions on Autonomous and Adaptive Systems (TAAS)*, vol. 7, no. 2, p. 21, 2012. http://dx.doi.org/10.1145/2240166.2240171

[13] T. Sandholm, "Algorithm for optimal winner determination in combinatorial auctions," *Artificial intelligence*, vol. 135, no. 1, pp. 1–54, 2002. http://dx.doi.org/10.1016/s0004-3702(01)00159-x

[14] C. M. Clark, R. Morton, and G. A. Bekey, "Altruistic relationships for optimizing task fulfillment in robot communities," in *Distributed Autonomous Robotic Systems 8*, pp. 261–270, Springer, 2009. http://dx.doi.org/10.1007/978-3-642-00644-9_23

[15] A. Wagner and R. Arkin, "Multi-robot communication-sensitive reconnaissance," in *Robotics and Automation, 2004. Proceedings. ICRA'04. 2004 IEEE International Conference on*, vol. 5, pp. 4674–4681, IEEE, 2004. http://dx.doi.org/10.1109/robot.2004.1302455

[16] R. Powers and Y. Shoham, "New criteria and a new algorithm for learning in multi-agent systems," in *Advances in neural information processing systems*, pp. 1089–1096, 2004.

[17] C. Perkins, E. Belding-Royer, and S. Das, "Ad hoc on-demand distance vector (aodv) routing," tech. rep., 2003. http://dx.doi.org/10.17487/rfc3561

[18] R. M. Karp, *Reducibility among combinatorial problems*. Springer, 1972. http://dx.doi.org/10.1007/978-1-4684-2001-2_9

[19] P. M. Ruiz and A. F. Gomez-Skarmeta, "Approximating optimal multicast trees in wireless multihop networks," in *Computers and Communications, 2005. ISCC 2005. Proceedings. 10th IEEE Symposium on*, pp. 686–691, IEEE, 2005. http://dx.doi.org/10.1109/iscc.2005.34

[20] L. Kou, G. Markowsky, and L. Berman, "A fast algorithm for steiner trees," *Acta informatica*, vol. 15, no. 2, pp. 141–145, 1981. http://dx.doi.org/10.1007/bf00288961

[21] A. Z. Zelikovsky, "An 11/6-approximation algorithm for the network steiner problem," *Algorithmica*, vol. 9, no. 5, pp. 463–470, 1993. http://dx.doi.org/10.1007/bf01187035

[22] N. Matloff, "Introduction to discrete-event simulation and the simpy language," *Davis, CA. Dept of Computer Science. University of California at Davis. Retrieved on August*, vol. 2, p. 2009, 2008.

[23] D. J. Watts and S. H. Strogatz, "Collective dynamics of small-world networks," *nature*, vol. 393, no. 6684, pp. 440–442, 1998.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1467–1470
DOI: 10.15439/2016F373
ACSIS, Vol. 8. ISSN 2300-5963

Simulating the Fractional Reserve Banking using Agent-based Modelling with NetLogo

Dagmar Monett
Computer Science Dept.
Faculty of Cooperative Studies
Berlin School of Economics and Law, Germany
Email: Dagmar.Monett-Diaz@hwr-berlin.de

Jesus Emeterio Navarro-Barrientos
Computer Science Dept.
Faculty of Cooperative Studies
Berlin School of Economics and Law, Germany
Email: e_navarrobarrientos@doz.hwr-berlin.de

***Abstract*—This work presents a multi-agent-based computational model of an artificial fractional reserve banking system. The model is implemented in NetLogo. The computational experiments and simulations we performed to analyse the proposed model show that different scenarios can lead to bank insolvency. We show that both the minimum reserve rate and the loss of confidence have large contributions to the insolvency of a bank, suggesting them as likely destabilizing economic forces driving the dynamics of the model.**

***Index Terms*—agent-based model, agent-based simulation, BDI agents, fractional reserve banking, NetLogo.**

I. Introduction

AGENT-BASED models (ABMs) are computational models consisting of a set of autonomous, self-driven agents that exhibit complex behaviours emerging from their interactions rather than from the complexity of the individual agents. ABMs are usually simulated in frameworks specially developed for these purposes [1]. They have already been applied to the study of emergent phenomena in a variety of domains that include social, political, and economic sciences.

ABMs have several interesting properties. Among them are the following: they are relatively easy to implement, are very practical for analysing the evolution of the simulations step by step, and can show emergent properties that could be difficult to predict. By stepping the simulations, it is possible to analyse the emergence of stylised facts and new equilibrium states, as well as the conditions under which they occur. For example, it is possible to analyse the emergence of pernicious domino effects, which may be achieved by increasing the degree of interdependence between the agents. The domino effects are of special interest to the analysis of financial fragility, in particular bankruptcies cascades, due to the intricate structures of liabilities among heterogeneous agents.

One of the goals of the ABM we present in this paper is to describe a methodological tool that can reproduce some of the stylised facts in fractional reserve banking (FRB) systems. *Fractional reserve banking*[1] is a banking system "in which banks hold only a fraction of their deposits in reserves, so that the reserve-deposit ratio is less than 1" [2]. In other words, some of the deposits are further used by the banks to be loaned out at interest-earning rates to other parties. Yet FRB has received a lot of criticism. For example, there are studies that show the viability of ending fractional reserve banking, as is the case of the FRB in Iceland [3].

Why then a computational agent-based model to simulate FRB? We believe that understanding FRB better could be one of the most important outcomes. Simulating artificial scenarios could help suggesting possible improvements or new policies. This is especially important for scenarios that could eventually be avoided if anticipated by a computational economic model for FRB. The FRB agent-based model presented in this paper can then be used to analyse possible scenarios that arise from evaluating different initial parameter settings of the model. The major purpose of the model is to provide artificial ways to represent and to simulate the impact of the fractional reserve banking system on a time period. It defines a very simple modern banking world that is, by no means, an example of real bank operations or of federal restrictions or monetary exchanges. Instead, it could serve as a basic playground setting, for example, to drive the policies and behaviours of banks before testing their validity in the real world. It could also be useful to find out the sufficient conditions for a banking system to become fragile and unstable.

II. Related Work

Traditional simulation approaches mainly use historical data [4] to analyse the interbank payment interactions. For example, Bedford, Millard, and Yang apply some stochasticity to test different bank behaviours under different hypotheses on the operational rules [5]. They propose a simulation-based framework to analyse large-value payment systems for a variety of worst-case scenarios. The framework shows many similarities to the stress-testing methods that are used to evaluate the robustness of banking systems to financial shocks.

Other researchers have used computer simulations to analyse interbank lending for scenarios with homogeneous and heterogeneous agents. Iori, Jafarey, and Padilla [6] show that, if the banks are homogeneous in size and risk exposure, then the interbank market has strong effects to avoid cascades and stabilise the system. However, if the agents are heterogeneous, then the system may present some cascade effects.

Modern simulation approaches like ABM have also been used to study economic and social systems, where the main

[1] Also known as *fractional deposit lending*.

idea is to describe the behaviour of the agents in the system and to reconstruct the aggregate behaviour by simulating their interactions. Also, some works allow behaviour adaptation based on changes in the different scenarios [7]. By this means, ABM is a methodology bringing together verbal descriptions of component systems and equation-based models [8]. In particular for our investigations, we are interested in ABM for analysing the credit, liquidity, and operational risks of settlement systems. In this type of system, banks are modelled as software agents that follow some behavioural rules and act independently, which leads to stylised facts that result from their interactions in the simulated world.

Simulation tools like StartLogo have also been used to simulate behavioural rules for banks in Real Time Gross Settlement systems. Arciero and co-authors present a model with a money market [9] which, after a critical event, either blocks or limits the activity of the bank. In their model, banks are perfectly informed on all payment requests. Thus, when delays in payments start accumulating, some banks start adjusting their expectations accordingly until the turbulence spills over in the market, needing the intervention of the central bank.

III. An Agent-based Computational Model for Fractional Reserve Banking

A computational model that describes an FRB system using ABM was introduced in a previous work [10]. The model basically consists of three main groups of artificial entities that are simulated by three types of agents, i.e., depositors or investors, debtors or borrowers, and banks, which interact through communication in a multi-agent system. When compared to the approach of Mallet of simply managing a list of accounts with deposits and loans [11], our model differs in that it simulates not only the bank behaviour, but also other parties and the interactions involved. In this paper, we focus on initial experiments with our model rather on extending it.

Each agent in our model pursues different interests. They are modelled in NetLogo,[2] a multi-agent programmable modelling environment [12], by following the BDI paradigm [13]. In other words, they are artificial agents with *beliefs* (B), *desires* (D), and *intentions* (I) that are defined using the NetLogo BDI add-on [14].

All agents follow a deliberation process that determines their subsequent actions and interactions with other agents. For example, depositors aim to create as much capital as possible without running the risk of losing their assets due to insolvency of the bank. They can retrieve the entire deposit or a specific, lower amount. They can also deposit money or do nothing. Figure 1 shows the deliberation process of a depositor agent. After updating her knowledge about the world and depending on both her preferences and trust in the bank, a depositor decides on whether to deposit money or to retrieve it, partially or totally.

[2]Jonathan Wiens implemented the first version of the NetLogo model. Eric Faustmann and Damian Rhein extended it.

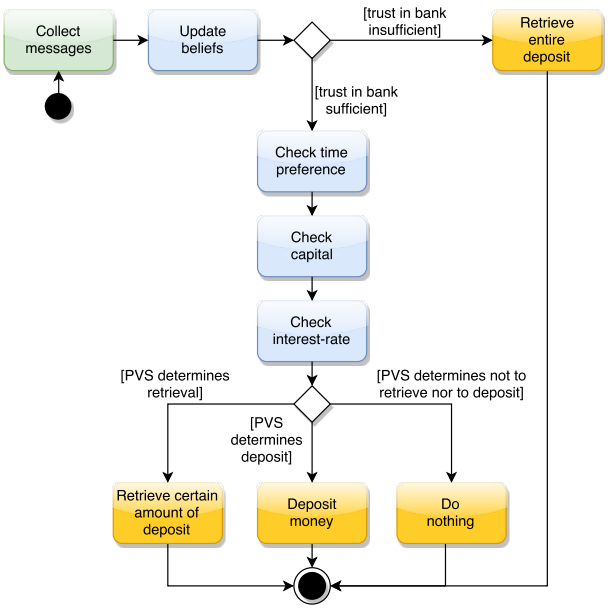


Fig. 1. Deliberation process of the depositor agent. PVS: personal value scale.

Depositors act depending on their own personal value scale. It would not be reasonable for a depositor to deposit 80% of her on-hand cash into a bank she has 10% of trust in, for instance. Other variables influence the decision process, too, like the time preference p, the current capital C, the deposit interest rate β, and the trust t in the bank. Algorithm 1 shows the pseudo-code that drives the depositors' actions, where D is the amount to deposit, W is the amount to withdraw, and S is the current deposits or savings in the bank. If confidence in the bank is lost, i.e., the trust in the bank is less than 30%, then a depositor might withdraw her entire deposits. She might deposit money, however, if the trust has a greater value and depending on both the time preference (a random parameter to simulate the possibly non-deterministic character of each agent's operations) and the interest rate.

Debtors or borrowers behave similarly, only that they have other local variables as well as actions in their repertoire, like borrowing a specific amount of money from the bank. The bank agent serves as a contact partner for depositors and debtors. It accepts or rejects requests from the agents depending on its own state. For example, a bank would award no credit to debtors if its reserves fall below the minimum permissible reserve amount because otherwise it could lead to insolvency. All agents update their information and knowledge about the world, i.e., their beliefs, iteratively, which determines the construction of new desires and intentions that are transformed in later actions.

IV. Experimental Settings

We are interested in finding which parameter values lead to either bank insolvency or to a stationary scenario with no insolvency of the bank, over iterations.

input : Incoming messages IM
output: Action a

begin
  process messages IM;
  update beliefs;
  update desires;
  if $0.5 \leq t \leq 1$ **then**
    $D = C \cdot (p + \beta)$;
  else if $0.3 \leq t < 0.5$ **then**
    $W = S \cdot (1 - p)$;
  else
    $W = S$;
  end
  update intentions;
  $a = selectBestOption()$;
  return a;
end

Algorithm 1: Depositor agent: Pseudo-code of the deliberation process at every iteration.

A bank becomes insolvent when its reserves are lower than the money one or more depositors want to withdraw back from their deposits. Such illiquid state scenarios are reached when both the trust of the depositors decreases, leading to withdrawals, and the reserves of the bank are lower than the withdrawals. Intuitively, the following three scenarios could lead to an illiquid state: (i) the bank does not invest at least some part of the deposits and converts them into profits from the loan interests to at least cover the deposit interests; (ii) the trust of the debtors decreases so that they do not want to get a loan. Thus, the profits of the bank would decrease and the bank will not be able to pay the deposit interests back; (iii) the bank uses a large part of the reserves but even a small withdrawal from a depositor can lead to an insolvency of the bank.

In order to analyse these scenarios, we first perform three computer experiments, i.e., *E1*, *E2*, and *E3*, where we change the values of both the minimum reserve rate and the average loss of confidence rate parameters, while leaving the rest of the parameter fixed, which are: number of depositors (5), number of debtors (5), average income (2000), start capital (5000), starting loan (credit) interest (10%), starting deposit interest (0.2%), and average win of confidence rate (30%). See Table I for those parameter values that differ among the experiments.

TABLE I
PARAMETER VALUES DIFFERING IN ALL EXPERIMENTS *E1*, *E2*, AND *E3*.

Experiment	minimum reserve rate	average loss of confidence rate
E1	1%	30%
E2	1%	50%
E3	5%	30%

V. RESULTS AND DISCUSSION

Different experimental results lead to the insolvency of the bank because of the scenarios mentioned in Section IV. We start by investigating the role of the average loss of confidence rate in the dynamics of the model. Figures 2 and 3 show an example of the evolution of the money warehouse receipts and the reserves over some iterations for an average loss of confidence rate of 30% and 50%, respectively. Note that in Figure 2 both the average loss of confidence rate and the average win of confidence rate are equal, whereas in Figure 3 the average loss of confidence rate is greater than the one depicted in Figure 2.

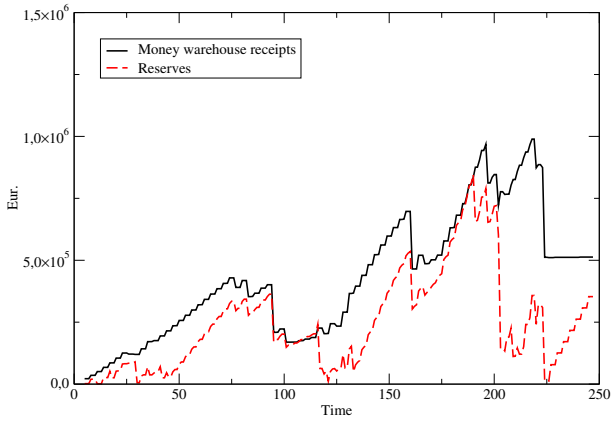


Fig. 2. Experiment *E1* with an average loss of confidence rate of 30% and a minimum reserve rate of 1%.

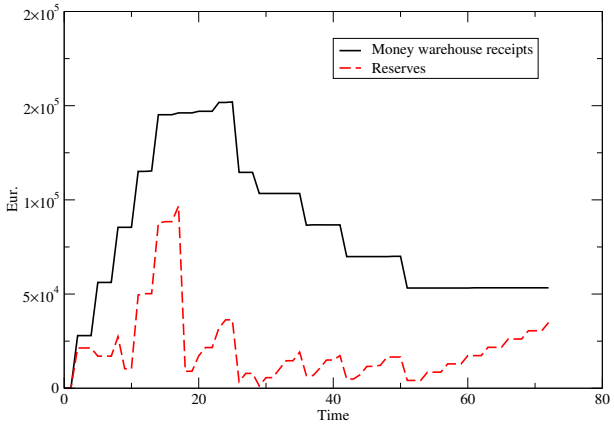


Fig. 3. Experiment *E2* with an average loss of confidence rate of 50% and a minimum reserve rate of 1%.

It can be seen that an insolvency is more probable when the loss of confidence increases, leading to a lesser number of iterations needed for a bank to become insolvent. Thus, when the average loss of confidence is greater than the average win of confidence, then both the depositors and the debtors lose their trust in the bank. Furthermore, if the reserves are low because of a large loan and the bank is not able to pay a deposit back, then all depositors start trying to withdraw their deposits, leading to an insolvency of the bank. This can be

seen in Figures 2 and 3 at the end of the simulations starting at iteration 220 and 50, respectively.

Furthermore, the larger the difference between the average loss and the win of confidence, the faster the depositors start to withdraw their deposits and the faster the debtors stop asking for loans from the bank. Note also that if the average loss of confidence is too great, then the difference between the deposits and the reserves can be so small that the bank may reach a state in which it does not become insolvent but neither does it have any depositors or debtors any more.

Figure 4 shows an example of the evolution of both the money warehouse receipts and the reserves for some iterations for the computer experiment *E3* with a minimum reserve rate of 5%. When comparing this result with the one from the computer experiment shown in Figure 2 where the minimum reserve rate is 1%, we can observe that the amount in Euros is greater in the computer experiment with higher minimum reserves (i.e., the one from Figure 4). It can be seen that the reason for the bank insolvency was the loss of confidence of a single depositor, which led to a cascade of confidence loss and withdrawals from all other depositors.

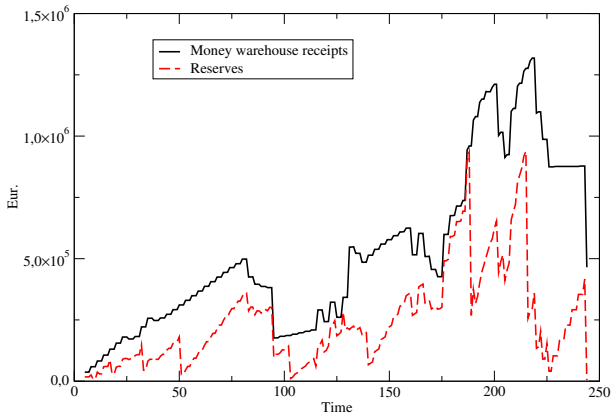


Fig. 4. Experiment *E3* with minimum reserve rate of 5% and an average loss of confidence rate of 30%.

We performed more computer experiments with larger minimum reserve values leading to fewer bank insolvency scenarios. The larger the minimum reserves are, the fewer the cases where the bank is not able to pay the withdrawals of the depositors. Therefore, it is not probable that a bank becomes insolvent unless a large number of depositors decide to withdraw their deposits at the same time.

These computer experiments also show that the minimum reserves do not have a great influence in the model. The reason for this is that there are not so many depositors. Thus, each agent has a large influence in the dynamics of the model. In this context, the loss of confidence of a single depositor can only lead to a withdrawal of 10 to 20% of the deposits in the bank. Furthermore, it was observed that the bank rarely becomes insolvent when the confidence loss has values between 30 and 60% and the re-utilised percentage of money is between 10 and 40%. It was also observed that, for a confidence loss between 30 and 40%, the bank becomes insolvent when a depositor withdraws her deposits just after a loan was granted.

VI. Conclusions

The results of the computer experiments show that the variable that has the greatest influence in the dynamics of the model is the average loss of confidence. This variable is determined randomly in the current version of the model. Further work is to consider different trust reputation mechanisms to make it adaptable, together with the average win of confidence, according to the number of deposits and loans over time.

Moreover, it would be of interest to further investigate the influence of the number of depositor and debtor agents in the model, i.e., whether the dynamics of the model scale in size or not. With respect to the bank agent, it would be interesting to analyse the impact of excluding external sources that would help a bank to pay to the depositors and keep it solvent.

References

[1] N. Gilbert and S. Bankes, "Platforms and methods for agent-based modeling," *Proceedings of the National Academy of Sciences*, vol. 99, no. 3, pp. 7197–7198, 2002.

[2] A. Abel and B. Bernanke, *Macroeconomics*, 5th ed. Pearson, 2005.

[3] F. Sigurjónsson, "Monetary Reform - A better monetary system for Iceland," Reykjavík, Iceland, Tech. Rep. 1.0, March 2015.

[4] K. Soramäki, M. L. Bech, J. Arnold, R. J. Glass, and W. E. Beyeler, "The topology of interbank payment flows," Federal Reserve Bank of New York, Tech. Rep. Staff Report no. 243, March 2006.

[5] P. Bedford, S. Millard, and J. Yang, "Analysing the impact of operational incidents in large-value payment systems: A simulation approach," in *Liquidity, risks and speed in payment and settlement systems âĂŞ a simulation approach*, H. Leinonen, Ed. Helsinki: Bank of Finland Studies, E:31, 2005, ch. 9, pp. 249–276.

[6] G. Iori, S. Jafarey, and F. Padilla, "Interbank Lending and Systemic Risk," *Journal of Economic Behavior and Organization*, vol. 61, pp. 525–542, 2006.

[7] M. Galbiati and K. Soramäki, "An agent-based model of payment systems," *Journal of Economic Dynamics and Control*, vol. 35, no. 6, pp. 859–875, 2011.

[8] N. Gilbert and P. Terna, "How to build an use agent-based models in social science," *Mind & Society*, vol. 1, pp. 57–72, 2000.

[9] L. Arciero, C. Biancotti, D. L., and C. Impenna, "Exploring agent-based methods for the analysis of payment systems: A crisis model for StarLogo TNG," *Journal of Artificial Societies and Social Simulation*, vol. 12, no. 1–2, 2009.

[10] J. Wiens and D. Monett, "Using BDI-extended NetLogo Agents in Undergraduate CS Research and Teaching," in *Proceedings of The* 9^{th} *International Conference on Frontiers in Education: Computer Science and Computer Engineering, FECS'2013*, H. Arabnia, A. Bahrami, V. Clincy, L. Deligiannidis, and G. Jandieri, Eds. CSREA Press U.S.A., July 2013, pp. 396–402.

[11] J. Mallett, "Analysing the behaviour of the textbook fractional reserve banking model as a complex dynamic system," in *Proceedings of the* 8^{th} *International Conference on Complex Systems, ICCS'2011*, H. Sayama, A. Minai, D. Braha, and Y. Bar-Yam, Eds., vol. 8. NECSI Knowledge Press, 2011, pp. 1141–1155.

[12] U. Wilensky, "NetLogo," Evanston, IL, U.S.A., 1999, available online at http://ccl.northwestern.edu/netlogo/, retrieved January 3, 2013.

[13] A. Rao and M. Georgeff, "BDI Agents: From Theory to Practice," in *Proceedings of the First International Conference on Multi-Agent Systems, ICMAS'95*. San Francisco, CA, U.S.A.: AAAI Press / The MIT Press, June 12–14 1995, pp. 312–319.

[14] I. Sakellariou, P. Kefalas, and I. Stamatopoulou, "Enhancing NetLogo to Simulate BDI Communicating Agents," in *Proceedings of the* 5^{th} *Hellenic Conference on Artificial Intelligence, SETN'08*, ser. Lecture Notes in Artificial Intelligence (LNAI), J. Darzentas, G. Vouros, S. Vosinakis, and A. Arnellos, Eds., vol. 5138. Syros, Greece: Springer Berlin Heidelberg, October 2008, pp. 263–275.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1471–1480
DOI: 10.15439/2016F126
ACSIS, Vol. 8. ISSN 2300-5963

Self-Organizing Redistribution of Bicycles in a Bike-Sharing System based on Decentralized Control

Thomas Preisler, Tim Dethlefs and Wolfgang Renz
Faculty of Engineering and Computer Science,
Hamburg University of Applied Sciences,
Berliner Tor 7, 20099 Hamburg, Germany
Email: {thomas.preisler,tim.dethlefs,wolfgang.renz}@haw-hamburg.de

***Abstract*—Currently, bike-sharing systems undergo a rapid expansion due to technical improvements in the operation combined with an increased environmental and health awareness of people. When it comes to the acceptance of such systems the reliability is of great importance. It depends heavily on the availability of bicycles at the stations. But, in spite of truck-based redistribution efforts by the operators, stations still tend to become full or empty, especially in rush-hour situations. This paper builds upon an incentive scheme that encourages users to approach nearby stations for renting and returning bikes, thereby redistributing them in a self-organized fashion. A cooperativeness parameter is determined by the fraction of users that respond to an incentive by choosing the proposed stations. It uses a decentralized control process to calculate alternative rent and return stations for each of the stations. These alternatives are then proposed to the users when they approach an empty or full station. The approach is based on a decentralized control framework that allows to equipping different distributed software systems with the control capabilities needed to realize the coordination efforts required to achieve the desired self-organizing properties.**

I. Introduction

RECENT challenges like climate changes, declining supplies of fossil fuels, noise emissions and congestion lead to discussions about individual means of transportation in urban areas. Especially bicycles (bikes in the following) have received an increased attention in city transportation, as they offer a healthy and environment-friendly way of transportation and allow to reach areas in cities that do not have direct access to public transportation. Combined with technical improvements of the underlying information systems, this results in a rapid extension of bike-sharing systems worldwide [1]. Obviously, bikes have drawbacks in comparison to other modes of transportation, as the usage of bikes strongly depends on weather conditions and the topography of the targeted area. This makes bikes more suitable for short trips [2]. As mentioned before, the increasing success of bike-sharing systems depends strongly on the introduction of information systems supporting the whole renting process (finding available bikes in the departure area as well as renting and returning them) [3]. Today, many cities aim at implementing bike-sharing systems in order to improve inner-city air quality and to reduce congestion [4].

The main challenge for the operation of modern bike-sharing systems in big cities is to ensure the *availability* of bikes at the stations. In rush-hour situations, stations may run out of bikes while others become full, thus reducing the overall *reliability* of the systems. Therefore, the planning and operation of redistribution attempts is essential to ensure reliability and user satisfaction. There are several attempts that have been tested to overcome these problems in scientific research as well as in practice [1] (cf. Section II).

This paper extends previous work [5], where an incentive scheme was investigated, that encourages users to approach nearby stations for renting and returning bikes, thus redistributing them in a self-organizing way. This work is extended by two aspects: First, a decentralized control framework is introduced that allows the declarative description of decentralized coordination processes to control the required coordination efforts among the participating entities in order to achieve the desired self-organizing behavior. Thereby, it shall support different types of heterogeneous applications and systems. The framework is used to replace the coordination processes that were tailored especially for the used RinSim simulator [6] presented in [5] by declarative, generic ones. Second, based on the redesigned control processes, the efficiency of the self-organizing redistribution approach depending on a circular communication range coordination parameter is examined. The communication range determines which other bike stations are within reach of a certain bike station and therefore, receive the status updates emitted from this station as part of the coordination process. There is a direct 1:1 mapping between the communication range and the maximum distance a user is detoured when an alternative rent or return station is proposed to him. Thus, minimizing the communication range is of concern when to ensure user cooperation and satisfaction.

Following the approach described in [5] a microscopic simulation of an idealized Monday based on data from Washington, D.C.'s bike-sharing system (2014) is realized. The simulation will be used to describe the application of the decentralized control framework and to measure the impact of the communication range as a coordination parameter. Washington, D.C.'s bike-sharing system was chosen as a base for the simulation

978-83-60810-90-3/$25.00

as all data concerning the system is freely accessible over the *Capital Bikeshare Dashboard* [7]. The *Capital Bikeshare* system has been started in September 2010 and until May 2013 it was the largest bike-sharing service offered throughout the United States [8]. In 2014, the system had 345 stations and about between 2400 and 2900 bikes were available for usage. Like many other bike-sharing system the pricing is based on the principle that the first 30 minutes of a rental are for free (except a fixed membership fee). Each additional 30 minutes require an extra fee. To ensure the reliability of the system and therefore, both the availability of bikes and free docks at the stations, Capital Bikeshare uses trucks to redistribute the bikes [9]. Fig. 1 shows the rebalancing efforts ventured by Capital Bikeshare in 2014. The figure shows that the operation of such a bike-sharing systems requires a significant amount of redistribution efforts.

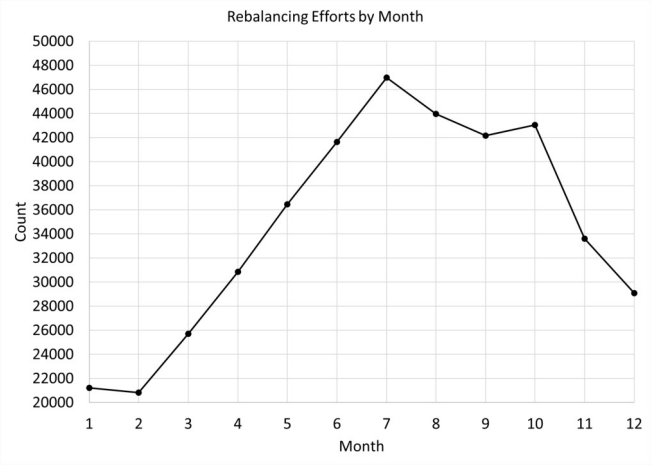


Fig. 1. Capital Bikeshare rebalancing efforts in 2014 (data taken from [7]).

However as shown in Fig. 2 stations still tend to become empty or full. Here the number of full and empty instances per month in 2014 are shown. From the data it becomes visibly, that the amount of empty stations is higher than the amount of full stations, indicating that the stations may be designed to have spare capacities to increase the chance that a bike can be returned at a station. In conclusion, the figure shows that, despite the redistribution efforts that are already carried out, there is potential for further improvement. This can either be an increasing number of truck-based redistributions or the introduction of new approaches, especially to overcome the problem of empty stations, e.g., by a self-organizing approach for the redistribution of bikes by the users.

The remainder of this paper is structured as follows: the next section will introduce related work in terms of the operation and planning of redistribution attempts as well as approaches dealing with the realization of self-organizing behavior based on decentralized control. Section III presents a decentralized control framework designed to allow the construction of decentralized coordination processes for different types of applications and systems. In Section IV, a Multi-Agent-based

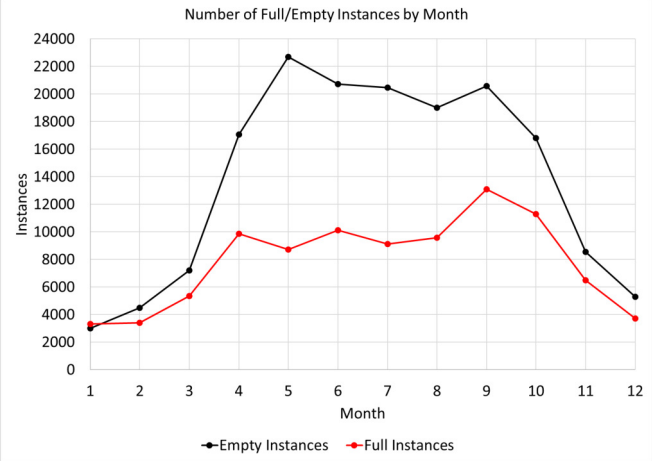


Fig. 2. Number of full/empty instances in 2014 (data taken from [7]).

simulation system of Washington D.C.'s bike-sharing system is introduced and it is described how the decentralized control framework is used to realize the self-organizing redistribution of bikes in the simulation system. Section V presents the results of the simulation and evaluates the impact of the communication range coordination parameter to the self-organizing strategy. Finally, Section VI concludes the paper and presents an outlook on future work.

II. Related Work

Several attempts have been ventured out to overcome the previous mentioned problem of maintaining reliability in bike-sharing systems. The authors of [10] used clustering techniques to identify shared behaviors across stations in order to predict short-term station usage for Barcelona's *Bicing* system. Thereby, they provide a spatiotemporal analysis of 13 weeks of bike station usage to sense and predict the rush on certain stations depending on their location and the time of day. Thus, supporting the operation of the system by providing planning support for the distribution of bikes. Similar work, also focusing on Barcelona's Bicing system was done in [11]. The authors analyzed human mobility data in an urban area using the amount of available bikes at the stations. Based on the data sampled by the system operator's website temporal and geographic mobility patterns within the city have been detected. These patterns were applied to predict the number of available bikes at the stations ahead. The timeliness of the whole bike-sharing topic is elaborated in [12] where the characteristics and commonalities between particular bike-sharing systems with a view to deriving influences on the sustainability of systems is explored. The empirical study analyzes bike-sharing systems in five Chinese cities. As China is suffering from severe negative consequence of high private vehicle usage in large and densely populated cities, it would greatly benefit from an environmentally-friendly way of transportation like bicycling. China has a long history of bike usage in the country and therefore, it provides a great potential for such a green form of travel to be part of public and

private transportation. Therefore, the authors of [12] analyze the effect of different bike-sharing systems in China and draw conclusions based on their success for the development of new systems.

In terms of building self-adaptive and self-organizing systems, there are several approaches which deal with the associated challenges. Research areas like Autonomic [13] or Organic Computing [14] provide approaches to solve these challenges. Both approaches rely on different types of feedback loops based on (usually) centralized control elements. According to [15] feedback loops are a key design element within a distributed system in order to exhibit adaptivity. Feedback loops normally consist of three main components: (1) Sensors are in charge of observing the behavior and the (current) status of the component, respectively the environment it is situated in. (2) Actuators can change the configuration of the system, which can lead to changes in the component's behavior. (3) A computing entity serves as a connector between the system input (sensor) and the output (actuator). It can be very different with regards to its internal architecture and abilities (cf. [16]). The importance of decentralized control to achieve requirements like resilience, robustness and scalability in large distributed systems has been identified in [17]. The work presented there distinguishes decentralized self-adaptive solutions from their centralized counterparts and also proofs some of the key research challenges for the realization of decentralized self-adaptation.

Related work in terms of decentralized coordination is among others presented in [18]. There a framework for the decentralized coordination of ubiquitous web services is proposed. It is based on an Event-Condition-Action (ECA) approach and relies on an XML-based language for describing ECA rules that are embedded in web service-enabled devices. Another example for a self-organizing infrastructure that offers coordination capabilities, inspired by chemical reactions is the *TuCSoN* coordination space concept [19]. It relies on a multiplicity of independent communication abstractions, called tuple centers. These can be spread over Internet nodes and are used by agents to interact with each other. TuCSoN exploits tuple centers as its coordination media, where a tuple center enhances a tuple space with a behavior specification. Therefore, the tuple centers are a communication abstraction whose behavior can be defined to embed an overall law of coordination. This is similar to the approach presented in this paper which utilizes coordination media as communication abstractions. Also similar is the propagation of a clean separation of concerns between application and coordination logic as introduced by [20]. The authors of [20] propagate a loose coupling between the core functionality of an application (computation) and the coordination. Thereby coordination is an orthogonal aspect w.r.t. to the computation when it comes to the realization of distributed systems. According to [20] this increases the generality when the coordination is swapped in a separate model.

However, there is a lack of approaches that support decentralized coordination in general regardless of the used technology and design patterns. The authors of [21] for example present a decentralized framework for the dynamic composition and coordination of autonomic agent applications. As a first step toward such a general decentralized control framework, that will be proposed in the next section, previous work dealt with a middleware supporting the construction of decentralized control in self-organizing system based on the concept of Active Components [22]. Active Components combine the autonomous behavior known from software agents with the service provider paradigm from the Service Component Architecture (SCA). A more detailed description about the concept of Active Components is given in [23].

III. DeCoF: A decentralized Coordination Framework

Today's distributed systems are characterized by an increasing size and complexity, which requires novel engineering approaches. The utilization of self-organizing processes has been proposed to enable adaptiveness of inherently decentralized systems [24]. Self-organization refers to physical, biological and social phenomena, where global structures arise from local interactions of autonomous individuals [25]. It has turned out to be a promising paradigm for the development of advanced distributed applications and systems with strongly decentralized control and high demands for self-adaptive behavior.

The designed decentralized coordination framework is based on the concept that the self-organizing dynamic that causes a system to adapt to external and internal influences is controlled by decentralized coordination processes. The processes describe the self-organizing behavior that continuously structures, adapts and regulates aspects of the application. They instruct a set of decentralized Coordination Media and Coordination Endpoints. Coordination Media deal with the interactions between the components (information propagation), while the Coordination Endpoints handle the adaptation of the components (local entity adaptation). Together they control the microscopic activities of the components, which on a macroscopic level lead to the manifestation of the intended self-organizing dynamic. The integration of the Coordination Endpoints and Media is prescribed by declarative defined coordination processes which structure and instruct their operations (cf. Section IV-A for an example of such a declarative coordination process description).

The Decentralized Coordination Framework (DeCoF) emerges from a tailored programming model for the software-technical utilization of coordination processes as reusable design elements in Multi-Agent Systems (MAS). The DeCoMAS (Decentralized Coordination for Multi-Agent Systems) [26] architecture introduces concepts like Coordination Media for the propagation of Coordination Information and Coordination Endpoints for the observation and adaptation of the local entities. But while the DeCoMAS architecture is especially

designed to equip BDI[1]-agent system with coordination processes and therefore, is limited to such systems, DeCoF aims at supporting distributed systems in general. Thereby, different and also heterogeneous software components in general are supported, allowing to equip not only MAS but component based systems in general with decentralized coordination processes to extend them with self-organizing capabilities. The current reference implementation supports software components written in *Java*. But, it is also applicable to other programming languages in general.

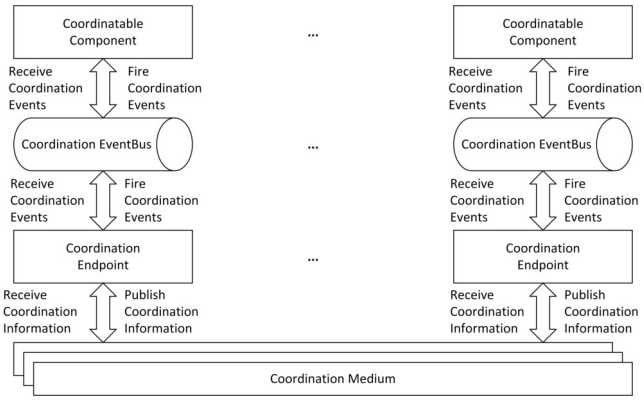


Fig. 3. Architecture of the Decentralized Coordination Framework

Fig. 3 shows the conceptual architecture of the proposed framework. Components respectively agents that should be equipped with coordination capabilities to realize self-organizing behavior based on the aforementioned concepts are labeled as *Coordinatable Components*. As the framework aims at supporting various types of MAS resp. distributed systems in general, there are no inherent characteristics that could be used to monitor or control the behavior of the agents respectively components, e.g. different types of MAS use different scheduling and life-cycle mechanisms, while component-based systems might lack them at all. Therefore, the concept of *Coordination Events* is introduced. These are events that are fired by a coordinatable component, whenever something relevant for the coordination happened inside the component. A coordination event (ce) is a tuple with the length 2 (double or 2-tuple) containing contextual data about the specific coordination event (cd) as well as a representation of the event's originator (eo). A coordination event is thus defined as: $ce = (cd, eo)$.

Following a separation of concerns between the application and the coordination logic as propagated by [20], the actual processing of the coordination events is handled by a related *Coordination Endpoint*. The coordination endpoints are loosely coupled to the coordinatable component via a so called *Coordination Event Bus*. An event bus[2] allows publish-subscribe-style communication between components without requiring the components to explicitly register with one another (and thus be aware of each others). The separation of concerns requirement is fulfilled by the loosely coupling between the coordinatable component and the coordination endpoint realized by the coordination event bus. Thus, the component is only responsible for realizing the application logic and do not need to have knowledge about (the present of) the coordination endpoint.

When a coordinatable component fires a coordination event, it is received by the related coordination endpoint via the coordination event bus. The endpoint then processes the event according to a prescribed coordination process definition. The process definitions defines how different coordination events have to handled by the endpoints. These descriptions contain instructions on *how* to distribute the coordination event to *which* other coordinatable components. The *how* is described by indicating what kind of *coordination medium* should be used for the information dissemination. As described before, coordination media deal with the information propagation among the components. The *which* is realized with a role-concept. A coordination process definition specifies various roles that components might adopt. Thereby, a component can have multiple roles and a role may be carried out by various component types. So to process a coordination event the endpoint encapsulates it and enriches it with additional information about the originating coordination endpoint. The resulting *Coordination Information* (ci) is a 2-tuple containing the coordination event (ce) and information about the originating endpoint (oe), thus is defined as: $ci = (ce, oe)$. Besides prescribing which coordination event, originating from which coordinatable components should be published to which other components, a coordination process definition also prescribes which type of coordination event should be triggered in the receiving components. How the coordination information are actually propagated is part of the implementation of the actual coordination medium. This regards the technical realization of how the information should be distributed, as well as how the subset of receivers is selected. Therefore, simple coordination medium relying on a network-topology for the information dissemination as well as complex ones, where the dissemination of the information relies on, e.g. diffusion processes in an (virtual) environment are possible.

Fig. 4 shows an UML class diagram of the relevant classes and interfaces of the framework. A component that should be equipped with coordination capabilities has to implement the `ICoordinatable` interface. It requires the component to implement two functions. The `getId` function returns an unique string that identifies the component. The `handleCoordinationEvent` function is called whenever a coordination event relevant for the component has been received by its coordination endpoint. Here the

[1]The Belief, Desire, Intention software model is developed for programming intelligent agents. It is characterized by the implementation of an agent's beliefs, desires and intentions and uses these concepts to solve a particular problem in agent programming [27].

[2]See: https://www.github.com/google/guava/wiki/EventBusExplained (accessed April 20, 2016)

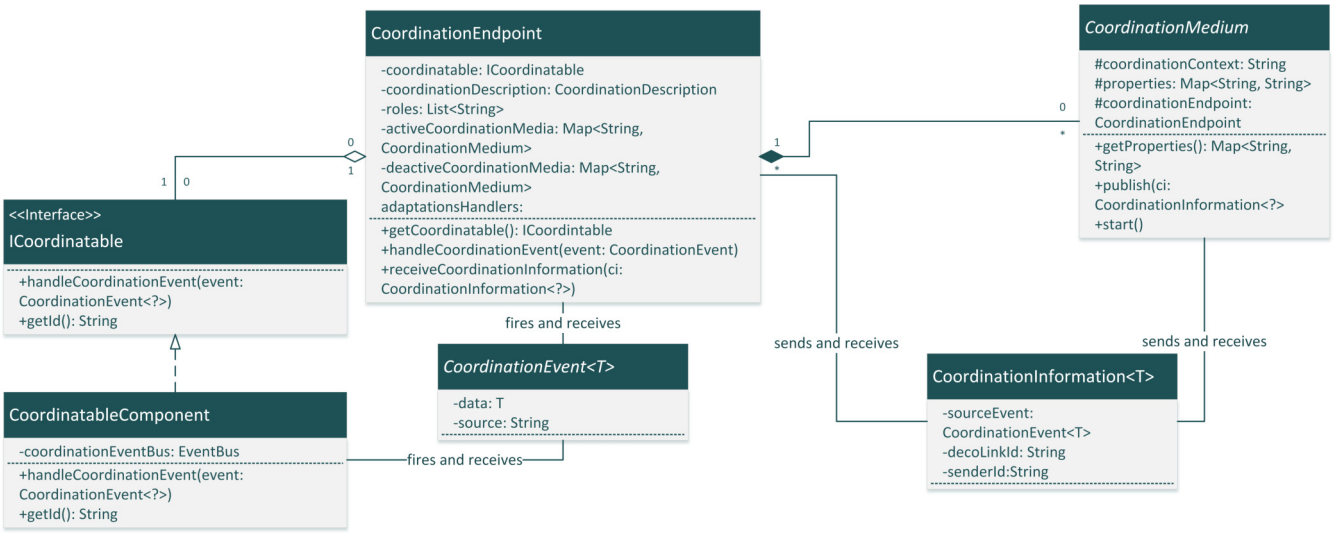


Fig. 4. UML class diagram of the Decentralized Coordination Framework.

component-specific coordination event handling should be implemented. Also, an according coordination endpoint has to created for the component. The framework provides a helper function that creates such a coordination endpoint and connects it to the component with a coordination event bus. The framework provides a ready-to-use implementation of the `CoordinationEndpoint` class as well as a generic implementation of the `CoordinationInformation` class. Abstract super classes exist to implement specific `CoordinationMedium` and `CoordinationEvent` classes. Therefore, if an application should be equipped with coordination capabilities the following steps have to be performed:

1) Writing the XML-based declarative coordination process description that instructs the coordination endpoints.
2) Implementing the `ICoordinatable` interface for the components that should be coordinated.
3) Identifying the relevant coordination events and implementing them using the abstract `CoordinationEvent` super class.
4) Implementing the coordination logic for the information propagation by extending the abstract `CoordinationMedium` super class.

IV. Simulation: Self-Organizing Redistribution of Bikes

The decentralized coordination framework described in the previous section has been used to implement a coordination process in order to control the self-organizing redistribution of bikes in a simulated bike-sharing system. This section will give an overview about the implementation details of the simulation system as well as the usage of the coordination framework. Also, the simulation scenario will be described.

A. Implementation Details

The simulation system is realized using RinSim [6], a logistics simulator written in Java. It supports (de)centralized algorithms for dynamic pickup-and-delivery problems (PDP). The cyclists renting bikes at stations were realized as agents. They are created at a specific station, where they try to rent a bike and if a bike is available at the station, they drive it to their designated destination stations, where they return it. In order to map the road model of Washington, D.C., the corresponding area was extracted from Open-StreepMap (OSM) [28] and transformed into the graph-based road model supported by RinSim. Thus, the movement of the cyclists along the roads can be simulated by moving them on the edges of the resulting graph. Following the PDP-modeling approach, the bikes were modeled as parcels and the bike stations as depots. Time in the simulation system is simulated in a discrete manner, divided into ticks of 60 seconds length. Therefore, the simulation of a whole day consists out of 1440 simulation ticks.

In order to rebalance the availability of bikes at the stations, as a possible addition to the truck-based redistribution efforts ventured by Capital Bikeshare, an incentive scheme for the users to stimulate them to re-distribute bikes in a self-organizing fashion is proposed. The approach is based on the concept that whenever a user tries to rent a bike at an empty or nearly empty station, an alternative rent station with a sufficient amount of bikes is suggested to the user. Equivalent, whenever a user tries to return a bike at a full or nearly full station, an alternative return station with a sufficient amount of free docks is suggested to the user. Thus, the distribution of bikes among the stations will be balanced in a self-organizing way, as users renting a bike are detoured from empty stations to, preferably full or nearly full ones or at least non-empty ones. The same goes for the returning of bikes, where users are detoured from full or nearly full stations to preferably empty or at least non-full ones.

For the simulation the alternative stations are proposed to the user agents whenever they approach a station, in a real world scenario a mobile phone application is imaginable that informs users about alternative rent or return stations before they actually approach them in case they state their intent to approach a station a priori. Possible incentives for a detour are, e.g., free minutes that are added to the users next trips or virtual bonus points that they can exchange for goodies in web shop, similar to the bonus programs of many retail stores. Fig. 5 depicts the whole process from a user's point of view.

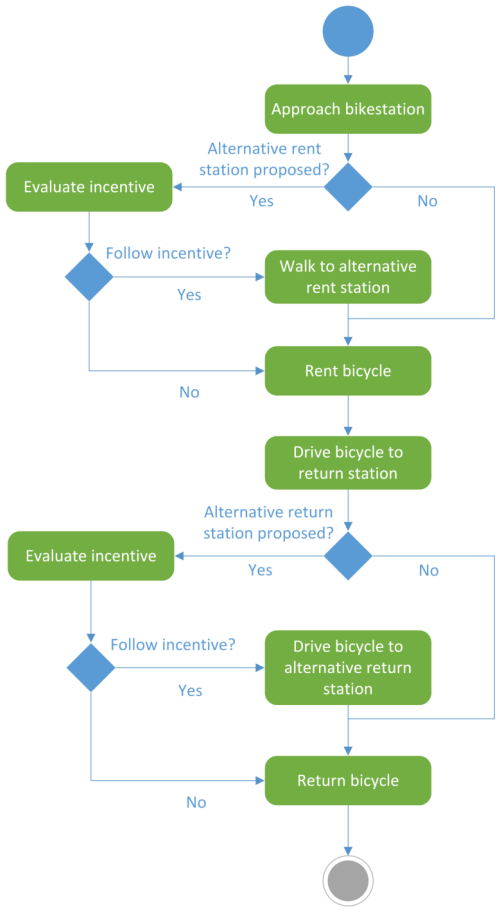


Fig. 5. UML activity diagram describing the rent/return process from a user's perspective.

A decentralized coordination approach is used to calculate the alternative rent and return stations that are suggested to the users. It is realized using DeCoF. Bike stations periodically send their current occupancy rate to all other bike stations within a certain circular communication range. Stations receiving such status updates from other stations collect them and use them to calculate alternative rent and return stations in a decentralized way. Whenever such status updates are received, the receiving bike station determines the station with the lowest and the highest occupancy rate from the list of stations. The station with the lowest occupancy rate is selected as the alternative return station and the station with the highest occupancy rate is selected as the alternative rent station. These alternative stations are suggested to a user whenever it tries to rent a bike at the station when the station is currently empty, respectively when the user tries to return a bike at the station when it is currently full or critical occupied. So, the maximum detour distance equals the communication range of the stations, as only stations that are within a station's communication range are considered. For the simulated scenarios, a critical occupancy rate of 75% is used.

Following the usage instructions of the DeCoF the bike stations that are implemented as RinSim specific depot agents (`Bikestation`), implement the `ICoordinatable` interface and are equipped with a coordination endpoint. Every minute of the simulated time they fire an `BikeStationStatusUpdateEvent` as a coordination event which contains their current occupancy rate. The according coordination endpoint publishes this as part of a coordination information over a `RoadBasedCoordinationMedium`. This medium extends the abstract `CoordinationMedium` super class with RinSim specific coordination logic. Therefore, it has a reference to the simulator's road model, so it has knowledge about the simulation environment and the position of all the bike stations. The circular communication range is a configurable coordination parameter of the medium. Based on the road model the medium selects all bike stations within the communication range and publishes the coordination information to their according coordination endpoints. When receiving such coordination information the endpoints trigger a `BikeStationStatusUpdateEvent` in the coordinatable bike stations and thus, initialize the calculation of the alternative rent and return stations. Listing 1 shows the declarative coordination process description for this application. It shows how the `Bikestation` agent is mapped to the `bikestation` role. This role is used to instruct a decentralized coordination link (`deco-link`). The link definition contains information about the affected roles, the `bikestation` role in this case, and how the coordination events should be mapped to each other as well as which coordination medium should be used for the information propagation. In this case the `RoadBasedCoordinationMedium` is used and it is configured with a `maxCommRange` coordination parameter limiting the information dissemination. In case of the listing a communication range of 1 km is used.

```
1  <coordination-description context="BikeSharing">
2   <role-definitions>
3    <role name="bikestation">
4     <components>
5      <component class="de.haw.c4das.bikesharing.
           simulation.Bikestation" />
6     </components>
7    </role>
8   </role-definitions>
9   <deco-link-definitions>
10   <!-- Bikestation status update link -->
11   <deco-link id="bs-update">
12    <from role="bikestation" event="de.haw.c4das.
          bikesharing.simulation.coordination.
```

```
        BikeStationStatusUpdateEvent" />
13    <medium class="de.haw.c4das.bikesharing.
        simulation.coordination.
        RoadBasedCoordinationMedium">
14     <properties>
15      <property key="maxCommRange" value="1" />
16     </properties>
17    </medium>
18    <to role="bikestation" event="de.haw.c4das.
        bikesharing.simulation.coordination.
        BikeStationStatusUpdateEvent" />
19   </deco-link>
20  </deco-link-definitions>
21 </coordination-description>
```

Listing 1. Bikesharing coordination process description (XML).

Fig. 6 visualizes the system's self-organizing dynamic that results from the previous described concepts. The dynamic is composed of two parts, the *Bikestation Coordination* as well as the *User Cooperativeness*, and it results from the interactions between them. The coordination among the bikestations generates the alternative rent or return stations that are proposed to the users. Then the self-organizing dynamic of the systems depends on the users' willingness to follow such a proposal.

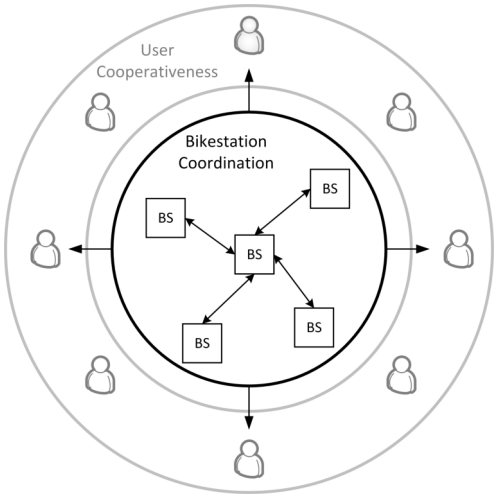


Fig. 6. Self-organizing dynamic of the bikesharing-system.

B. Simulation Scenario

A typical Monday was examined in order to simulate Washington D.C.'s bike-sharing system and to evaluate the impact of the proposed coordination strategy. Therefore, the trip history data of all Mondays (except holidays) from 2014 provided by Capital Bikeshare (the system's operator) was analyzed. To do so, the day was divided into 24 time slices. For each of these 24 time slices the departure probabilities q and the dependent destination probabilities q for the bike-stations were calculated based on the ventured trips: The departure probability p_A for a station A is denoted by

$$p_A = \frac{n_A}{N},$$

with n_A as the number of all trips started at station A while N is the total number of trips. The dependent destination probability q_B is denoted by

$$q_B = \frac{d_B}{D_A}.$$

It is characterized by the fraction of numbers of departures to station B from station A denoted as d_B and the total number of departures from station A denoted as D_A. The simulated scenario starts at 12 a.m. In order to simulate the different rush at different times of the day, the mean total number of trips for each of the 24 hours of the day was determined based on the trip history data. During the execution the simulator generates the number of cyclist agents specified by the rush equally distributed for the currently simulated time slice. As a simplification, all cyclists move with a constant speed along the graph-based road model. In order to find a route from the departure to the destination stations, they use a shortest path approach and traverse the edges of the graph road model, considering the edge weight as the distance to the next node. The simulation was configured to allow an overcrowding of bike stations, when no free docks are available. If a cyclist agent tries to rent a bike at and empty station, this incident is reported and the total number of rides that did not take place is returned as part of the simulation results for evaluation purposes. Fig. 7 shows an extract of the simulated map with the road model and some of the cyclists and bike stations as well as the chosen communication range.

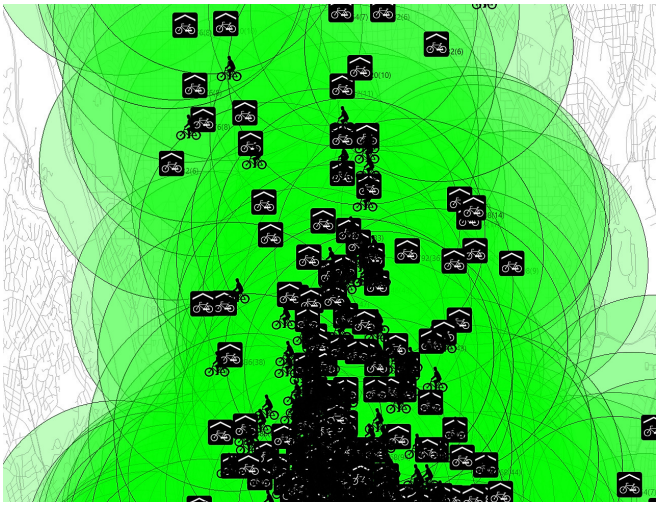


Fig. 7. Extract of the simulated map showing the road model, some of the cyclists and bike stations as well as the chosen communication range.

V. Simulation: Self-Organizing Redistribution of Bikes

In order to evaluate the impact of the self-organizing redistribution strategy depending on the communication range as the relevant coordination parameter, different scenarios with fluctuating communication ranges as well as a reference scenario without any self-organizing behavior were simulated. In all of this scenarios the cyclists moved with a fixed speed of 18km/h. Initially, half of the docks of the bike stations were

occupied. The bike stations' number of docks was extracted from the data provided by operator via the *Capital Bikeshare Dashboard* [7]. For all the scenarios simulating the self-organized redistribution of bikes, a user cooperativeness rate of a 100% was assumed. This means that the users always follow a proposed detour to an alternative rent respectively return station. This assumption was made for a more precise evaluation of the impact of the communication range. A detailed analysis of the impact of the users' cooperativeness rate on the proposed self-organizing redistribution strategy is presented in [5].

Fig. 8 shows and compares the results of a simulation scenario with no self-organizing redistribution of bikes and one with a communication range limited to 1,5 km. The parameters of the simulated scenario are summarized in Table I. The figure depicts the number of stations that are in a normal state (neither full nor empty) for both scenarios. It is observable for both cases, how the number of normal stations declines with the morning rush-hour beginning at around 7 a.m. (minute 420). Over the day, these numbers fluctuate only a little. In the late afternoon (around minute 1000) the number of normal stations recovers a bit. This behavior can be explained by the rush-hour movements of commuters. In the morning they drive from the suburbs to the city center and return in the afternoon. Stations in the suburbs tend to become empty during the morning rush-hour, while stations in the city center tend to become full or over-crowded. The returning commuters in the afternoon take bikes from the overcrowded stations in the city center and refill the empty ones in the suburbs when they return, thus increasing the number of normal stations. The figure gives a first impression about how the self-organizing redistribution of bikes improves the number of normal stations over the whole day.

TABLE I
PARAMETERS OF THE CONDUCTED SIMULATION.

Simulation Parameter	Value
No. of cyclists	7433
No. of stations	345
Cyclist speed	18 km/h
Initial bikestation utilization rate	50%
User cooperativeness rate	100%
Bikestation communication range	1.5 km

In order to further measure the impact of the self-organizing redistribution with regards to the maximum communication range, 6 simulations with different communications ranges (from 0,5 km to 3 km) were performed and compared to the reference scenario. Apart from the fluctuating communication range the same simulation parameters as displayed in Table I were used. The according results are shown in Fig. 9. It contains the mean deviation of the number of normal stations in comparison to the previous described reference scenario with no self-organizing behavior. This value states how many more stations in average over the simulated day are in the normal state in comparison to the reference scenario. The figure shows that with a communication range of 1,5 km a

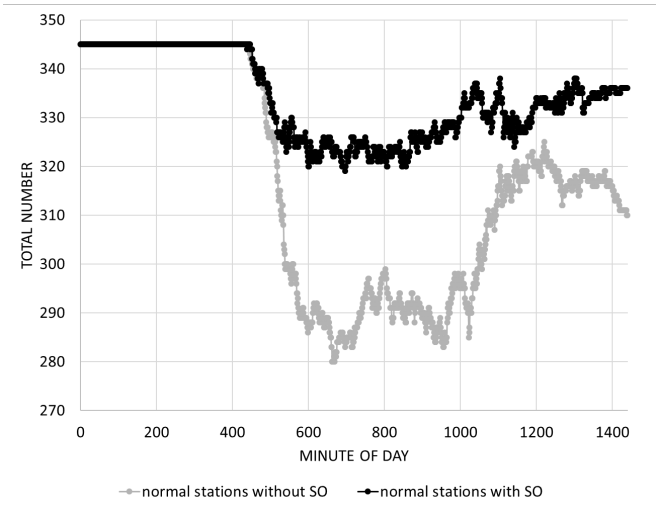


Fig. 8. Results of a simulation scenario with enabled and disabled self-organization.

maximum deviation of 10,7% can be achieved. This means that about 10% more of the total amount of stations are in the normal state in comparison to the scenario with no self-organizing behavior. Considering 345 stations in total, an improvement of 10% of the stations means that about 35 more stations are in the normal state. Fig. 8 shows that without any self-organizing redistribution efforts at least about 280 stations are in the normal state. This results in 65 stations being empty or full/overflown. By increasing the number of normal stations by 10% of all stations to about 315 stations, the number of stations that are empty or full/overflown is reduced by about 55% to 30 stations.

As the communication range is directly related to the distance a user is detoured when he/she wants to rent or return a bike, it is obvious that this value should be as low as possible to increase user acceptance. Also the figure shows that a higher communication range actually may have a negative impact on the results. One potential reason is that the trip time is lengthened and therefore, also the time the bikes are in usage and thus, not available at the stations. In addition, a higher communication ranges may decrease the density of bikes in the area where they are actually needed. For example, they might be detoured from the city center back to the suburbs if the communication range is too high and therefore, no longer available to meet the higher demand in the city center.

VI. CONCLUSION AND FUTURE WORK

The acceptance of bike-sharing systems depends heavily on the availability of bikes at the stations. In spite of truck-based redistribution efforts by the operators, stations still tend to become empty or full, especially in rush-hour situations. In this paper, we explored an approach for the self-organizing redistribution of bike by the users and presented a decentralized coordination framework for the realization of self-organizing systems. It is based on the concept that the self-organizing dynamic that causes a system to adapt to external

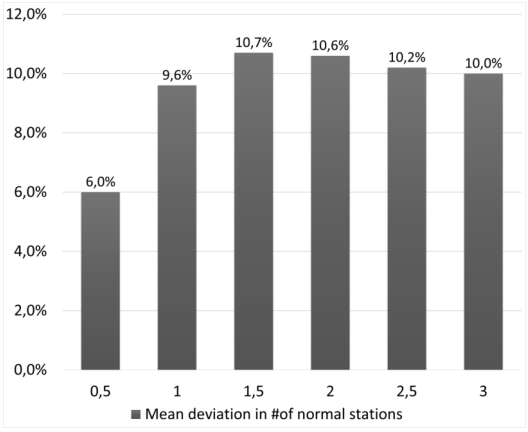


Fig. 9. Mean deviation of the number of normal stations in comparison to a reference scenario with no self-organizing behavior. X-axis denotes communication ranges in km.

and internal influences is controlled by decentralized coordination processes. The processes describe the self-organizing behavior that continuously structures, adapts and regulates aspects of the application. They instruct a set of decentralized coordination media and coordination endpoints. Coordination media deal with the interactions between the components (information propagation), while the coordination endpoints handle the adaptation of the components (local entity adaptation). Together they control the microscopic activities of the components, which on a macroscopic level lead to the manifestation of the intended self-organizing dynamic. The integration of the coordination endpoints and media is prescribed by declarative defined coordination processes which structure and instruct their operations. A microscopic simulation of a bike-sharing system based on data taken from Washington, D.C. (2014) was realized to show how the presented framework can be used to build self-organizing applications. In case of the presented bike-sharing system, a decentralized coordination process was introduced that allowed the bike stations to communicate their current occupancy rate to other stations within a certain circular communication range to calculate alternative rent and return stations for the users. Thereby, we measured and evaluated the impact of the communication range as the relevant coordination parameter for the performance of the self-organizing behavior.

Future work will deal with the structural adaptation of coordination processes. Under certain conditions self-adaptive systems may exhibit a behavior where performance decreases and/or starvation may occur. Structural adaptations are a promising concept under such conditions, that can be used to realize the dynamic exchange or reconfiguration of self-organizing coordination processes. By facilitating the concept of structural adaptations presented in [29] for the proposed decentralized coordination framework it is envisioned to realize the dynamic exchange of the described coordination process. Thereby, another coordination process based on the formation of clusters among the stations, where a random station will take on the role of a super station for this cluster and receive status updates from all stations within the cluster, so it can calculate the optimal alternative rent and return stations for all stations within the cluster, will dynamically replace the existing coordination process at runtime.

REFERENCES

[1] P. Vogel and D. Mattfeld, "Modeling of repositioning activities in bike-sharing systems," in *Transport Research (WCTR), 2010 World Conference on*, 2010.

[2] P. DeMaio, "Bike-sharing: History, impacts, models of provision, and fu-ture," *Journal of Public Transportation*, vol. 12, no. 4, pp. 41–56, 2009.

[3] S. Bührmann, "Bicycles as public-individual transport - european developments." Rupprecht Consult Forschung und Beratung GmbH, Cologne, Germany, Tech. Rep., 2008.

[4] P. Midgley, "The role of smart bike-sharing systems," *Urban Mobility Journeys*, vol. 2, pp. 23–31, 2009.

[5] T. Preisler, T. Dethlefs, and W. Renz, "Data-adaptive simulation: Cooperativeness of users in bike-sharing systems," in *Logistics (HICL), 2015 Hamburg International Conference of*, W. Kersten, T. Blecker, and C. M. Ringle, Eds., vol. 20. epubli GmbH, 2015, pp. 201–228.

[6] R. van Lon and T. Holvoet, "Rinsim: A simulator for collective adaptive systems in transportation and logistics," in *Self-Adaptive and Self-Organizing Systems (SASO), 2012 IEEE Sixth International Conference on*, Sept 2012. doi: 10.1109/SASO.2012.41. ISSN 1949-3673 pp. 231–232.

[7] "Capital bikeshare dashboard," http://cabidashboard.ddot.dc.gov, accessed: August 4, 2016.

[8] M. Martinez, "Washington, d.c. launches the nation's largest bike sharing program," *Grist Magazin*, 2010.

[9] J. Maus. (2013) Behind the scenes of capital bikeshare. Available at: http://bikeportland.org/2013/03/10/behind-the-scenes-of-capital-bikeshare-84006 (retrieved August 4, 2016).

[10] J. Froehlich, J. Neumann, and N. Oliver, "Sensing and predicting the pulse of the city through shared bicycling," in *Artifical Intelligence (IJCAI), 2009 International Joint Conference on*, ser. IJCAI'09. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2009, pp. 1420–1426. [Online]. Available: http://dl.acm.org/citation.cfm?id=1661445.1661673

[11] A. Kaltenbrunner, R. Meza, J. Grivolla, J. Codina, and R. Banchs, "Urban cycles and mobility patterns: Exploring and predicting trends in a bicycle-based public transport system," *Pervasive and Mobile Computing*, vol. 6, no. 4, pp. 455–466, Aug. 2010. doi: 10.1016/j.pmcj.2010.07.002. [Online]. Available: http://dx.doi.org/10.1016/j.pmcj.2010.07.002

[12] L. Zhang, J. Zhang, Z. yu Duan, and D. Bryde, "Sustainable bike-sharing systems: Characteristics and commonalities across cases in urban china," *Journal of Cleaner Production*, vol. 97, pp. 124–133, June 2015.

[13] J. O. Kephart and D. M. Chess, "The vision of autonomic computing," *Computer*, vol. 36, no. 1, pp. 41–50, 2003.

[14] J. Branke, M. Mnif, C. Müller-Schloer, H. Prothmann, U. Richter, F. Rochner, and H. Schmeck, "Organic computing - addressing complexity by controlled self-organization," in *Leveraging Applications of Formal Methods, Verification and Validation (ISoLA), 2006 International Symposium on*, ser. ISOLA '06. IEEE Comp. Soc., 2006. ISBN 978-0-7695-3071-0 pp. 185–191.

[15] Y. Brun, G. Marzo Serugendo, C. Gacek, H. Giese, H. Kienle, M. Litoiu, H. Müller, M. Pezzè, and M. Shaw, "Software engineering for self-adaptive systems through feedback loops," B. H. Cheng, R. Lemos, H. Giese, P. Inverardi, and J. Magee, Eds. Berlin, Heidelberg: Springer-Verlag, 2009, ch. Engineering Self-Adaptive Systems through Feedback Loops, pp. 48–70. ISBN 978-3-642-02160-2

[16] J.-P. Mano, C. Bourjot, G. A. Lopardo, and P. Glize, "Bio-inspired mechanisms for artificial self-organised systems," *Informatica (Slovenia)*, vol. 30, no. 1, pp. 55–62, 2006.

[17] D. Weyns, S. Malek, and J. Andersson, "On decentralized self-adaptation: Lessons from the trenches and challenges for the future," in *Software Engineering for Adaptive and Self-Managing Systems (SEAMS), 2010 ICSE Workshop of*, ser. SEAMS '10. New York, NY, USA: ACM, 2010. doi: 10.1145/1808984.1808994.

ISBN 978-1-60558-971-8 pp. 84–93. [Online]. Available: http://doi.acm.org/10.1145/1808984.1808994

[18] J.-Y. Jung, J. Park, S.-K. Han, and K. Lee, "An eca-based framework for decentralized coordination of ubiquitous web services," *Information and Software Technology*, vol. 49, no. 11-12, pp. 1141 – 1161, 2007. doi: http://dx.doi.org/10.1016/j.infsof.2006.11.008. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0950584906001959

[19] E. Nardini, M. Viroli, M. Casadei, and A. Omicini, "A self-organising infrastructure for chemical-semantic coordination: Experiments in tucson," in *Proceedings of the 11th WOA 2010 Workshop, Dagli Oggetti Agli Agenti, Rimini, Italy, September 5-7, 2010.*, 2010. [Online]. Available: http://ceur-ws.org/Vol-621/paper17.pdf

[20] D. Gelernter and N. Carriero, "Coordination languages and their significance," *Commun. ACM*, vol. 35, no. 2, pp. 97–107, 1992. doi: 10.1145/129630.129635. [Online]. Available: http://doi.acm.org/10.1145/129630.129635

[21] Z. Li and M. Parashar, "A decentralized agent framework for dynamic composition and coordination for autonomic applications," in *Database and Expert Systems Applications, 2005 Sixteenth International Workshop on*, Aug 2005. doi: 10.1109/DEXA.2005.10. ISSN 1529-4188 pp. 165–169.

[22] T. Preisler, T. Dethlefs, and W. Renz, "Middleware for constructing decentralized control in self-organizing systems," in *Autonomic Computing (ICAC), 2015 IEEE International Conference on*, July 2015. doi: 10.1109/ICAC.2015.56 pp. 325–330.

[23] A. Pokahr and L. Braubach, "The active components approach for distributed systems development," *International Journal of Parallel, Emergent and Distributed Systems*, vol. 28, no. 4, pp. 321–369, 2013. doi: 10.1080/17445760.2013.785546. [Online]. Available: http://dx.doi.org/10.1080/17445760.2013.785546

[24] G. Di Marzo Serugendo, M.-P. Gleizes, and A. Karageorgos, "Self-organization in multi-agent systems," *The Knowledge Engineering Review*, vol. 20, no. 2, pp. 165–189, Jun. 2005. doi: 10.1017/S0269888905000494. [Online]. Available: http://dx.doi.org/10.1017/S0269888905000494

[25] M. Prokopenko, "Design vs. self-organization," in *Advances in Applied Self-organizing Systems*, ser. Advanced Information and Knowledge Processing, M. Prokopenko, Ed. Springer London, 2008, pp. 3–17. ISBN 978-1-84628-981-1. [Online]. Available: http://dx.doi.org/10.1007/978-1-84628-982-8_1

[26] J. Sudeikat and W. Renz, "Decomas: An architecture for supplementing mas with systemic models of decentralized agent coordination," in *Web Intelligence and Intelligent Agent Technologies (WI-IAT), 2009 IEEE/WIC/ACM International Joint Conferences on*, vol. 2, Sept 2009. doi: 10.1109/WI-IAT.2009.137 pp. 104–107.

[27] A. S. Rao, M. P. Georgeff *et al.*, "Bdi agents: From theory to practice." in *Proceedings of the First International Conference on Multiagent Systems (ICMAS'95)*, 1995, pp. 312–319.

[28] M. Haklay and P. Weber, "Openstreetmap: User-generated street maps," *Pervasive Computing, IEEE*, vol. 7, no. 4, pp. 12–18, Oct 2008. doi: 10.1109/MPRV.2008.80

[29] T. Preisler, W. Renz, and T. Dethlefs, "Structural adaptation for self-organizing multi-agent systems: Engineering and evaluation," *International Journal on Advances in Intelligent Systems*, vol. 8, no. 3&4, pp. 413–425, 2015.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1481–1490
DOI: 10.15439/2016F549
ACSIS, Vol. 8. ISSN 2300-5963

Run-time Injection of Norms in Simulated Smart Environments

Patrizia Ribino, Carmelo Lodato, Antonella Cavaleri, Massimo Cossentino
Istituto di Reti e Calcolo ad Alte Prestazioni
Consiglio Nazionale delle Ricerche
Palermo, Italy
Email: {ribino, c.lodato, a.cavaleri, cossentino}@pa.icar.cnr.it

***Abstract*—Smart systems have to deal with environmental changes and react for adapting their behavior to changes in the operating conditions, so to always meet users' expectations. This is fundamental for those systems operating in open environments that may change frequently. Smart environments are complex systems that more than others are affected by these issues. In this paper, we propose a normative framework for regulating at run-time system behavior when some situations occur, thus providing system flexibility. The proposed approach includes also mechanisms to identify anomalous situations that can occur in the system due to the run-time injection of new norms.**

I. Introduction

A GREAT challenge in complex systems is to adequately deal with the unpredictability and the dynamics changing of the application context the systems are plugged in. Smart environments are complex systems that more than others are affected by these issues. A way to provide system flexibility is given by implementing statically normative frameworks in which norms for regulating the behavior of the system are specified at design time. This kind of solution is not effective when we face with unpredictable and unexpected situations that have not been considered at design time. Hence, solutions for modifying at run-time norms or injecting new ones into the systems are mandatory.

It is widely accepted that multi-agent systems provide relevant features for implementing smart environments [1][2][3][4]. In such field, norms have been widely employed for regulating the ideal behavior of the system. Norms are considered as a mechanism for controlling multi-agent societies and for ensuring order and predictability [5]. Such norms define standards of behaviors that have to be adopted in the society as well as undesired behaviors that have to be avoided. Hence, normative frameworks rely on a representation of norms by means of permission, obligations and prohibitions that ensure the agents behave within predefined boundaries. In particular, several works have been conducted for addressing theoretical and practical aspects about norm change [6][7][8] providing agents with mechanisms for enacting behavior modification. Typically, new plans/actions have been created for agents to be complied with new norms.

In this paper, we propose a normative framework to be coupled with goal oriented systems in order to introduce more flexibility in smart environment by means of run-time injection of norms regulating system goal fulfillment. In our approach, norms and goals allow to cope with two main aspects of a smart environment. Goals express what is the desired state of the world the system has to result in. Norms regulate the desired state of the world according to the normative context in which the system works.

In order to modify system behaviors when new norms are injected into the system, we implemented some algorithms in agents life cycle for reasoning about new goals to be pursued according to new norms. In so doing, we also manage some anomalous situations that can occur in the system due to the run-time injection of new norms. In particular, we identified three kinds of anomalies:

- the injection of an inconsistent norm, namely containing logical contradictions;
- the injection of a norm that is structurally incompatible with a goal;
- finally, the injection of a norm that creates an antinomy.

Thus, the main contributions of this work are:

- an algorithm for modifying agent goal fulfillment according to norm changing;
- an algorithm for checking the consistency of norms with the system.

In order to show in semi-natural languages some practical examples, we use GoalSpec [9] and SBVR [10] for modeling goals and norms.

Moreover, the normative framework with algorithms was implemented in MUSA [11]. It is a Middleware for User-driven Service Adaptation that provides means for supporting run-time adaptation of a process together with a multi-agent system for executing the activities of the process.

In order to validate our approach, we simulated a smart environment for improving services of a health care home. The simulation has been developed by integrating MUSA with ICasa simulator [12].

The rest of the paper is organized as follows. Section II introduces the theoretical background of the paper. In Sections III and IV, related works and motivation are respectively presented. Section V and Section VI presents the key concepts and the algorithms the proposed normative framework is based on. Section VII illustrates an application scenario of the proposed approach in a simulated smart environment. Finally, in Section VIII conclusions are drawn.

978-83-60810-90-3/$25.00

II. BACKGROUND

The main purpose of this work is to introduce more flexibility in smart environments by means of run-time injection of norms for adapting the behavior of the system to new situations. This section introduces the theoretical background of the paper. In particular, it is organized in four parts: the first one presents the features making multi-agent systems more suitable for implementing smart environments; the second one introduces two modeling languages (GoalSpec [9] and SBVR [10]) we use for modeling goals and norms for specifying functionalities and setting regulations the smart environment has to own. The third one provides an overview of MUSA [11], a multi-agent system for the dynamic composition and the orchestration of services in a distributed and open environment we adopt for implementing the functionality of the smart environment. The last one introduces ICasa [12], a dynamic pervasive environment simulator, we use for simulating the injection of norms in a smart environment.

A. Multi-Agent System features for Smart Environment

A Multi-Agent System (MAS) is a distributed system composed of autonomous entities, called agents. The decentralized and loosely coupled nature of such kind of systems makes it possible to design applications that are highly flexible, scalable and adaptive. The multi-agent paradigm provides several features that make them more suitable in order to fit smart environment requirements. Among them:

- *Decentralization* - MASs provide decentralized control based on distributed autonomous entities.
- *Interactions* - MASs support complex interactions between entities, using high level semantic languages. It is essential for smart environments that commonly deal with various, heterogeneous information from physical sensors, services or users preferences.
- *Coordination* - In a MAS, individual entities with limited capabilities are able to coordinate in order to achieve complex tasks. Flexible organization patterns enable groups of agents to create and dynamically reconfigure applications depending on current conditions. In an open, and dynamic smart environment, this feature is highly suitable for dynamic composition of elementary functionality in order to accomplish more sophisticated processes.
- *Heterogeneity* - A MAS is a society of agents with different capacities and roles.

B. GoalSpec & SBVR

GoalSPEC [9] is a language designed for specifying user-goals and enabling at the same time goal injection and software agent reasoning. The concept of goal is central in GoalSpec. Goals are described as states of the world that the user desires to achieve [13]. In GoalSpec, a goal is composed of a *Trigger Condition* and a *Final State*. The trigger condition is an event that must occur in order to start acting for addressing the goal. The final state is the desired state of the world that must be addressed. Trigger conditions and final states must be expressed by using domain ontology predicates. In GoalSpec, uppercase words represent the keywords of the language, and lowercase words represent the predicates constrained by the domain ontology.

Let us suppose we want define in our smart environment system the following common user goal:

If at 07:00 the living-room temperature is below 20 C, the shutter should be closed and the air conditioning turned on at level 6.

In GoalSpec, it could be written as follow:

WHEN on(07:00 pm) AND temperature(T) AND T<20 THE system SHALL ADDRESS closed(shutter) AND turned_on(air_conditioning, level(6)).

In order to define norms in natural language, we adopted the SBVR [10] standard as a modeling language for our system.

The Semantics of Business Vocabulary and Business Rules (SBVR) is an adopted standard of the Object Management Group (OMG). It is designed for business domains for formalizing complex business rules. In business domains, a business rule describes the conditions of a business process execution. A business rule may define the semantics of business concepts, reactions to business events, constraints and preconditions on tasks and activities, as well as the prohibitions, permissions and obligations of business actors and activities. In other words, business rules guide and constrain various aspects of business, including the sequence and timing of activities [14]. SBVR uses 'semantic formulation', which is a way of describing the semantic structure of statements and definitions.

In our approach, we use GoalSPEC for defining expected results of the system in terms of functionality and we adopt SBVR for specifying the normative context the system works in. Hence, by using the same kind of abstraction of SBVR, we can see a smart environment as an organization of entities that are involved in activities for reaching goals. The entities of the smart environment, sharing the same domain ontology (the SBVR vocabulary), constitute a community's body of shared meanings that can understand SBVR rules.
An example of SBVR rules for the previously defined goal could be: *It is prohibited that the system closes shutters if John is at home*.

C. MUSA- A Middleware for User-driven Service Adaptation

The Middleware for User-driven Service Adaptation (MUSA) proposed in [11] is intended to provide a means for supporting run-time adaptation of a process based on a multi agent system for executing the activities of the process. The core element of the approach is the use of Goals for explicitly representing user-preferences into the system (what to address). The injection of goals triggers the re-organization of the agents in hierarchical groups. These self-adaptive structures allow for dealing with dynamic composition and orchestration of services. MUSA provides a platform in which *(i)* it is possible to deploy some capabilities that wrap real services,

completing them with a semantic layer for their smart use; *(ii)* users can inject their goals for satisfying their specific needs. Under the hypothesis that both goals and capabilities refer to the same semantic layer (described as an ontology), then the agents of the system are able to conduct a proactive means-end reasoning for composing available capabilities into tasks for addressing the user request.

D. iCASA - a dynamic pervasive environment simulator

iCasa [12] is a set of integrated tools for the development and administration of pervasive applications. It includes a smart home simulator that allows the creation and removal of a wide range of devices that can be used by the applications. Specifically, iCasa Simulator provides:

- A graphical user interface that displays a map of the house and the localization of the different devices. It allows developers to create and configure devices, create and move physical users, and watch their actual configurations;
- Scripting facilities for controlling the environment and to test the applications under reproducible conditions.
- Notification facilities for notifying users of any modifications in the environment.

III. Related Works

Norms like obligations, permissions and prohibitions have been implemented in multi-agent systems in order to specify (un)desired behavior of agents so that the goals of the system can be reached. They also provide means for coordinating agent activities in order to reach the overall objective of the system they are part of [15]. Norm-governed systems are also known as Normative systems.

Normative systems are commonly defined as systems that specify every possible system transition, whether or not that transition is considered to be legal or not. In other words, Normative Systems specify which actions or which states should be achieved or avoided [16][17][18].

A lot of work has been done about normative frameworks in the field of Electronic Institutions or Virtual Organizations where norms have found a natural implementation. For instance, Alechina *et.al* [19] present a programming framework for developing normative organizations. Such framework is based on N-2APL, a BDI-based agent programming language for implementing norm-aware agents. N-2APL supports normative concepts such as obligations, prohibitions and sanctions. In such a work, the normative system is conceived in such away that the interaction between agents and the environment is regulated by a "normative exogenous organizations", which is defined by means of a set of conditional norms (i.e: conditional obligation and conditional prohibition). Such norms have the form *obligation(l, o, d, s)* that means "*agent l is obliged to establish an environment state satisfying o before deadline d, otherwise it will be sanctioned by updating the environment with s*"[19]. A norm-aware deliberation approach is also proposed. It allows agents to determine the set of plans (to be adopted in order to satisfy a goal) of highest priority which do not violate higher priority prohibitions.

In [20] Kollingbaum and Norman proposed the NoA Normative Agent Architecture. It supports the implementation of norm-governed practical reasoning agents. NoA agents are motivated by norms to act. In the NoA language, all the effects of a plan are declared in a plan specification. These effects are considered by agents for reasoning about plan selection and execution. Moreover, the norms governing the behavior of a NoA agent refer to either actions that are obligatory, permitted, forbidden, or states of affairs that are obligatory, permitted or forbidden. The NoA language enables an agent to be programmed in terms of plans and norms. Normative statements formulated in the NoA language express obligations, permissions and prohibitions of an agent: *Obligations* motivate the agent to achieve either a state of affairs or to perform a specific action. Prohibitions require the agent to not achieve a state of affairs or to not perform an action. The agent is forbidden to pursue a specific activity. Prohibitions represent restrictions on what capabilities the agent is allowed to reach a certain goal. Finally, Permissions allow the achievement of a state of affairs or the performance of an action.

Some works have been conducted for addressing norm change and norm consistency providing agents with mechanisms for enacting behavior modification. Typically, new plans/actions have been created to comply with new norms [8][21][22]. In [23], Jiang *et.al* propose a normative structure, named *Norm Nets* (NNs) for modeling sets of interrelated regulations. NNs aims at verifying whether executions of business process are compliance with process regulations. Authors define a norm as a tuple of elements that specify the type of deontic operator, the pair role-action (the target) to which the deontic modality is assigned, a deadline of norm validity and a precondition that determines when the target is initiated. A formal method for checking norm compliance by using Colored Petri Nets is proposed. In [24] [25], authors propose a means for automatically detecting and solving conflict and inconsistency in norm-regulated Virtual Organization and Electronic Institution.

In the next section, we provide the reasons that motivate our work with respect to the related works.

IV. Motivation

Although a considerable literature exists about norms, it is mainly directed to explore the role of norms inside Virtual Organization and Electronic Institution that are tightly coupled with agents. We take inspiration from those works and we are trying to introduce norms in smart systems as mechanisms that regulate the system at a higher level of abstraction than that where agents work.

The normative framework we defined adopts a norm specification similar to the previous ones. But in our framework, we employ norms at a higher level of abstraction by moving them from activity's regulations to goal's regulations. This choice is motivated by the context of self-adaptive and self-organized systems (SASO) we are working on. Commonly this

kind of systems are able to effectively adapt their behavior to environment changes and self-organize their internal structure for finding composed solutions in order to achieve collaborative goals. In particular, the kind of SASO systems we are considering owns the following features [11]:

- *Openness* - They are open systems that evolve at run-time because: *(i)* new services could be made available for satisfying user requirements; *(ii)* the satisfaction of new user requirements may be demanded to the system;
- *Goal-directed* - They are goal-directed systems. Goals are motivators for these systems providing them the reason for doing something. Goals express user requirements to be satisfied.

The aim of this work is to increase the openness of the system by allowing the run-time introduction of new regulations thus improving the flexibility of such systems. In so doing, we look at norms from a different perspective with respect to the classical one. Starting from the consideration that goals are key elements for the systems we focus on and they are motivators of their behavior, we look at goals as commitments the system engages with users. In other word, goals can be seen as a particular kind of obligation that has to be satisfied when some conditions occur.

Hence, in the normative framework we developed norms that are directly linked to goals where a permission norm relaxes the conditions under which the goal has to be satisfied and a prohibition norm nullifies the commitment under the circumstances expressed by the prohibition. The effect is that norms may act for increasing the opportunity for the system to pursue the goal it is committed to (Permission) or, on the contrary, norms may inhibit system intentions to pursue the goal it is committed to (Prohibition).

By adopting this perspective, the norms operate for regulating *what* the system has to satisfy and not *how* it does that.

In order to provide an example, let us to suppose there is a system that is committed for satisfying the user goal *have lunch* and at the moment of the commitment, the system owns two means to satisfy the *have lunch* goal. Fig. 1 shows a goal diagram where the goal "*have lunch*" can be satisfied by performing the tasks "*book a restaurant*" or "*take a pizza*". Let us suppose that such smart system has to provide its assistance to a diabetic patient. A norm in the system states that "*It is prohibited that a diabetic patient has lunch before taking insulin*". The norm (at the goal level) constraints the requirement that the system can provide by inhibiting its intentions rather than disabling all the possible ways the system can follow in order to satisfy that requirement.

Let us suppose now that a new task "*cook with microwave oven*" is introduced at run time for satisfying the *"have lunch"* goal (see Fig.2). Considering that the norm at the goal level spreads to the task level, we do not need to add/change any system regulations to adapt the behavior of the system to manage the change of its operative context.

Moreover let us suppose that a new norm for the goal "*have lunch*" is injected at run-time in the system. The simultaneous presence of interrelated norms may cause some system conflict or inconsistency. Indeed, when two norms are interrelated, it may happen that being compliant with a norm may cause to be uncompliant with the second one, thus generating conflicts or inconsistencies. In classical approaches, conflicts are generated when an agent wants to perform an action that is simultaneously allowed and forbidden. Inconsistencies, instead, occur when an agent may be forbidden to perform an action that may be essential for fulfilling one of its obligations [26].

In our approach, we characterize the definition of conflict and inconsistency to deal with dynamically changing environments, where the conflicting state of two norms may change according to the particular execution context. In order to address this concern, we introduce some new definitions about conflicts and inconsistencies that are based on a representation of the execution context.

Resuming the previous example, let us suppose that a new norm regulating the "*have lunch*" goal is introduced into the system, that is "*It is permitted that guests have lunch if they have performed sports activities*". This latter norm is interrelated to the previous one because they both refer to the same goal "*have lunch*". The injection of the second norm could cause a system deadlock because if the conditions of both norms are simultaneously valid an antinomy is generated and the system does not know how to behave.

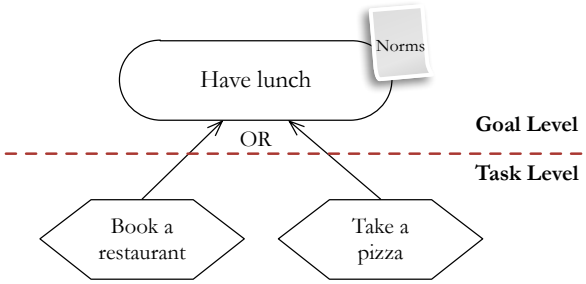


Fig. 1: An example of Means-End Analysis [27]. It introduces two tasks "*Book a restaurant*" and "*Take a pizza*" to indicate two particular ways to fulfill the goal "*Have lunch*".

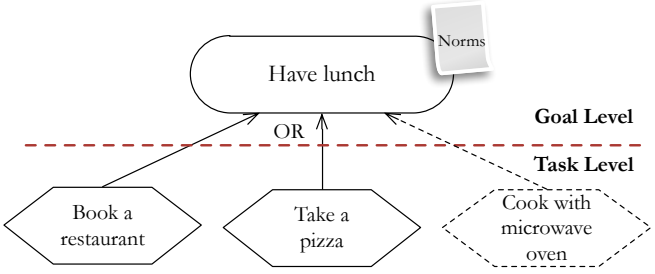


Fig. 2: The availability of the new task "*cook with microwave oven*" gives the system a new mean for satisfying the goal "*Have lunch*".

In the following section, we introduce the key concepts our approach is based on.

V. Theoretical Foundations

The normative framework for introducing more flexibility in smart environments, we propose in this work is based on three key concepts: state of the world, goal and norm. In the following, we formally introduce some definitions.

▼ Definition 1 — *State of the world*

Let $\mathcal{D}$ be the set of concepts defining a business domain. Let $\mathcal{L}$ be a first-order logic defined on $\mathcal{D}$ with $\top$ a tautology and $\bot$ a logical contradiction, where an atomic formula $p(t_1, t_2..., t_n) \in \mathcal{L}$ is represented by a predicate applied to a tuple of terms $(t_1, t_2..., t_n) \in \mathcal{D}$ and the predicate is a property of or relation between such terms that can be true or false.

A *state of the world* in a given time t ($\mathcal{W}^t$) is a subset of atomic formulae whose values are true at the time t:

$$\mathcal{W}^t = [p_1(t_1, t_2, ..., t_h), ..., p_n(t_1, t_2, ..., t_m)]$$

The *state of the world* represents a set of declarative information concerning events occurred within the environment and relations among events at a specific time. An event can be defined as the occurrence of some fact that can be perceived by or be communicated to the smart system. Events can be used to represent any information that can characterize the situation of an interacting user as well as a set of circumstances in which the smart system operates at a specific time. Definition 1 is based on the close world hypothesis[28] that assumes all facts that are not in the state of the world are considered false.

▼ Definition 2 — *Goal*

Let $\mathcal{D}$, $\mathcal{L}$ and $p(t_1, t_2..., t_n) \in \mathcal{L}$ be as previously introduced in the definition 1. Let $t_c \in \mathcal{L}$ and $f_s \in \mathcal{L}$ be formulae that may be composed of atomic formulae by means of logic connectives AND($\wedge$), OR ($\vee$) and NOT ($\neg$).

A *Goal* is a pair $\langle t_c, f_s \rangle$ where t_c (*trigger condition*) is a condition to evaluate over a state of the world $\mathcal{W}^t$ when the goal may be actively pursued and f_s (*final state*) is a condition to evaluate over a state of the world $W^{t+\Delta t}$ when it is eventually addressed:

- *a goal is active iff* $t_c(\mathcal{W}^t) \wedge \neg f_s(\mathcal{W}^t) = true$
- *a goal is addressed iff* $f_s(\mathcal{W}^{t+\Delta t}) = true$

Goals express what is the desired state of the world the system has to result in. Conversely, *Norms* regulate the desired state of the world according to the normative context in which the system works.

▼ Definition 3 — *Norm*

Let $\mathcal{D}$, $\mathcal{L}$, $p(t_1, t_2..., t_n) \in \mathcal{L}$ and $\mathcal{W}^t$ be as previously introduced in the definition 1. Let $\phi \in \mathcal{L}$ and $\rho \in \mathcal{L}$ be formulae composed of atomic formula by means of logic connectives AND($\wedge$), OR ($\vee$) and NOT ($\neg$). Moreover, let $D_{op} = \{permission, obligation, prohibition\}$ be the set of deontic operators. A *Norm* is defined by the elements of the following tuple:

$$n = \langle r, g, \rho, \phi, d \rangle$$

where

- $r \in \mathcal{R}$ is the *Role* the norm refers to. The special character "_" indicates that the norm refers any role.
- $g \in \mathcal{G}$ is the *Goal* the norm refers to. The special character "_" indicates that the norm refers to any goal.
- $\rho \in \mathcal{L}$ is a formula expressing the set of actions and state of affairs that the norm disciplines.
- $\phi \in \mathcal{L}$ is a logic condition (to evaluate over a state of the world $\mathcal{W}^t$) under which the norm is applicable;
- $d \in D_{op}$ is the deontic operator applied to ρ that the norm prescribes to the couple $(r, g) \in \mathcal{R} \times \mathcal{G}$.

$$\text{In particular } d(\rho) = \begin{cases} \rho & \text{iff } d = \textit{obligation} \\ \neg\rho, & \text{iff } d = \textit{prohibition} \\ \rho \vee \neg\rho & \text{iff } d = \textit{permission} \end{cases}$$

In other words, let $\mathcal{W}^t$ be a state of the world, a norm prescribes to a couple (r, g) the deontic operator d applied to ρ if ϕ is true in $\mathcal{W}^t$ [1].

▼ Definition 4 — *State of Norm*

Let a norm $n = \langle r, g, \rho, \phi, d \rangle$ where $g = \langle t_c, f_s \rangle$ and let a state of the world in a given time t ($\mathcal{W}^t$)

A norm can assume the following states:

- n is *applicable at time t* if $\phi(\mathcal{W}^t) = true \vee \phi = \top$
- n is *active at time t* if n is applicable and $t_c(\mathcal{W}^t) = true$
- n is *logically contradictory* if ϕ *is* $\bot$
- n is *in opposition to goal* if $f_s \wedge d(\rho)$ *is* $\bot$

Moreover, let a state of the world (W^t) and let two norms $n_1 = \langle r_1, g_1, \rho_1, \phi_1, d_1 \rangle$ and $n_2 = \langle r_2, g_2, \rho_2, \phi_2, d_2 \rangle$ where $r_1 = r_2$, $g_1 = g_2$, $\rho_1 = \rho_2$

- n_1 and n_2 are *deontically contradictory* iff

$$\begin{cases} \phi_1(\mathcal{W}^t) \wedge \phi_2(\mathcal{W}^t) = true \\ d_1 \neq d_2 \end{cases}$$

It is worth noting that we talk about *logically contradictory* when the contradiction concerns the logical conditions ($\phi \in \mathcal{L}$) under which the norms are applicable. On the contrary, we talk about *deontically contradictory* when the contradiction concerns the semantic meaning of the deontic operator ($d \in D_{op}$) the norms apply to.

Moreover, a norm is in opposition to a goal when pursuing that goal violates always the prescribed norm.

In the next section, we present the algorithms for the injection of norms in a running smart environment.

VI. Algorithms for run-time injections

The aim of the normative framework is to provide some mechanisms that allow to modify the behavior of the smart environment in order to adapt it to unexpected situations that can occur by introducing new norms. The approach we propose is illustrated by the Algorithm 1.

The triple of elements the Algorithm 1 works on is composed of: a *state of the world* $\mathcal{W}^t$ that characterizes the system in a given time, a *set of goal* $\mathcal{G}$ representing the requirements the system is able to satisfy and finally a *set*

[1] It is worth noting that in order to be compliant in $\mathcal{W}^t$ with 1) an obligation ρ must be true, 2) a prohibition $\neg\rho$ must be true 3) a permission ρ or $\neg\rho$ may be true. In the context of this paper, we assume that the system does not violate norms.

Algorithm 1: RunTime injection

Data: $\mathcal{W}^t$, $\mathcal{G}$, $\mathcal{N}$
while *system is running* **do**
  $\mathcal{N}^t_{injected} \leftarrow inject\_new\_norms$;
  ①**for** $j \leftarrow 1$ **to** $length(\mathcal{N}^t_{injected})$ **do**
    $\langle r, g, \rho, \phi, d\rangle \leftarrow n_j$;
    $\langle t_c, f_s\rangle \leftarrow g$;
    if $(\phi \neq \bot) \wedge (f_s \wedge d(\rho) \neq \bot)$ **then**
      add $\langle r, g, \rho, \phi, d\rangle$ to $\mathcal{N}$;
    else
      $revise(n_j)$
  ②**foreach** $g_i \in \mathcal{G}$ **do**
    $\langle t_{ci}, f_{si}\rangle \leftarrow g_i$;
    if $\neg f_{si}(\mathcal{W}^t)$ **then**
      $\mathcal{N}_i \leftarrow \{n \in \mathcal{N} : n = \langle r, g_i, \rho, \phi, d\rangle \wedge \phi(\mathcal{W}^t) = true\}$;
      Ⓐ **if** $card\{\mathcal{N}_i\} = 0 \wedge t_{ci}(\mathcal{W}^t) = true$; **then**
        $pursue(g_i)$;
      Ⓑ **if** $card\{\mathcal{N}_i\} = 1$; **then**
        $\langle r, g_i, \rho, \phi, d\rangle \leftarrow n$;
        if $(d = Permission)$ **then**
          $pursue(g_i)$;
      Ⓒ **if** $card\{\mathcal{N}_i\} > 1$; **then**
        $\mathcal{N}_i \leftarrow Check\_Norms(\mathcal{N}_i, \mathcal{W}^t)$; (see Alg.3)
        $(\phi_{OR}, \phi_{AND}) \leftarrow Compose\_Norm\_Condition(\mathcal{N}_i)$; (see Alg.2)
        $t^{'}_{ci} \leftarrow OR\_composition(t_{ci}, \phi_{OR})$;
        $t^{'}_{ci} \leftarrow AND\_composition(t^{'}_{ci}, \phi_{AND})$;
        $g_i^{'} \leftarrow \langle t^{'}_{ci}, f_{si}\rangle$;
        if $t^{'}_{ci}(\mathcal{W}^t) = true$ **then**
          $pursue(g_i)$;

of norms $\mathcal{N}$ the system has to obey in order to deal with some specific situations. Both $\mathcal{N}$ and $\mathcal{W}^t$ may change during system execution. In particular, the state of the world may change due to some events that can occur or some actions that can be performed in the environment. The set of norms may change due to norm injection.

While the system is running, new norms can be injected. Step ① makes a preliminary check on new injected norms. Such step ensures that among injecting norms neither norms are in opposition to the goal they refer nor they are logically contradictory.

Step ② is the core of the algorithm. The system is in the state of the world ($\mathcal{W}^t$), the norms have effects on the system goals only if they are not addressed yet. Thus, for each goal the system has to satisfy, the set of applicable norms ($\phi(\mathcal{W}^t) = true$) is processed.

Hence, three situations can occur. The most simple one (Ⓐ) is that there are no applicable norms for an active goal (see *Definition 2*). In such a case the system can pursue the goal without restrictions. The second situation (Ⓑ) is a basic case in which the set of applicable norms for a single goal is composed of only one norm. In such case, *(i)* if the norm is a permission it actives the related goal and the system can fulfill that goal despite $t_c(\mathcal{W}^t) = false$. This is because the permissions relax system constraints, giving alternatives; *(ii)* if the goal is regulated by a prohibition, it further constraints the goal activation. The system cannot pursue that goal until $\phi(\mathcal{W}^t) = true$. It is worth noting that generally speaking, an applicable norm influences a goal when it is active. However, a permission norm can influence a goal also when it is inactive.

The last situation (Ⓒ) is a general case in which norms are more than one and they can have different deontic operators. In this case Algorithm 1 allows to modify goals, making them norm compliant. By encapsulating the condition expressed by the norms inside the goal they refer to, it is possible to modify the activation of that goal thus making it compliant with the norms. Such composition (see Algorithm 2) takes into consideration different types of norms and it accordingly modifies the activation of a goal.

Algorithm 2: Compose_Norm_Condition

Data: a list of norms $NormList$
Result: a couple $(\phi_{mergedOR}, \phi_{mergedAND})$
$List\phi\_OR \leftarrow \varnothing$;
$List\phi\_AND \leftarrow \varnothing$;
// Identification of norm types
for $j \leftarrow 1$ **to** $size(NormList)$ **do**
  $\langle r, g, \rho, \phi, d\rangle \leftarrow NormList[j]$;
  switch d **do**
    case *Obligation* **do**
      break;
    case *Prohibition* **do**
      add $\neg\phi$ to $List\phi\_AND$;
    case *Permission* **do**
      add ϕ to $List\phi\_OR$;
// Permissions give alternatives (OR)
if $Size(List\phi\_OR) \neq 0$ **then**
  $\phi_{mergedOR} \leftarrow List\phi\_OR[1]$;
  for $h \leftarrow 2$ **to** $Size(List\phi\_OR)$ **do**
    $\phi_{mergedOR} \leftarrow OR\_composition(\phi_{mergedOR}, List\phi\_OR[h])$;
// Prohibition are mandatory (AND)
if $Size(List\phi\_AND) \neq 0$ **then**
  $\phi_{mergedAND} \leftarrow List\phi\_AND[1]$;
  for $h \leftarrow 2$ **to** $Size(List\phi\_AND)$ **do**
    $\phi_{mergedAND} \leftarrow AND\_composition(\phi_{mergedAND}, List\phi\_AND[h])$;

It is worth noting that, when there are more than one applicable norm in the system (Ⓒ), it is necessary to check for deontological contradictions among norms (see Definition 4) and to remove them. This is performed by Algorithm 3. Deontological contradictions are known in legislative environments as *antinomy*. For instance, if there is an applicable norm *n1* that prohibits to pursue a goal *g1* and another applicable

norm *n2* that obliges to pursue the same goal *g1*, then *n1* and *n2* generate an antinomy.

In legal theory, several criteria exist for solving such antinomy [29]: *legis posterior*, the most recent norms takes precedence; *legis superior* the norm imposed by the strongest institutional power takes precedence; and *legis specialis* the most specific norm takes precedence. In this paper, we assume to work with hierarchically equal norms (i.e: norms with the same authority) thus we adopt the *legis posterior* criterion in order to choice among conflicting norms.

In the following, we show our approach in a simulated smart environment.

Algorithm 3: Check_Norms

Data: a list of applicable norms $\mathcal{N}$ related to a single goal, a state of the world $\mathcal{W}^t$
Result: a list $\mathcal{N}_{out}$ of consistent norms
$\mathcal{N}_{out} \leftarrow chronological\_order(\mathcal{N})$;
$\mathcal{M}_{conflicts} \leftarrow \varnothing$;
for $i \leftarrow 1$ **to** $length(\mathcal{N}_{out})$ **do**
  $\langle r, g, \rho, \phi_i, d_i \rangle \leftarrow n_i$;
  for $j \leftarrow i+1$ **to** $length(\mathcal{N}_{out})$ **do**
    $\mathcal{M}_{conflicts}[i][i] \leftarrow 1$;
    if $r_i = r_j \wedge \rho_i = \rho_j \wedge d_i \neq d_j$ **then**
      $\mathcal{M}_{conflicts}[i][j] \leftarrow 1$;
    else
      $\mathcal{M}_{conflicts}[i][j] \leftarrow 0$;
for $i \leftarrow length(\mathcal{N}_{out})$ **to** 1 **do**
  $n_{current} \leftarrow \mathcal{N}_{out}[i]$;
  for $j \leftarrow i-1$ **to** 1 **do**
    if $\mathcal{M}_{conflicts}[i][j] = 1$ **then**
      $delete(\mathcal{N}_{out}, oldest(\mathcal{N}_{out}[j], n_{current}))$;
      $n_{current} \leftarrow \mathcal{N}_{out}[j]$;

VII. A Simulated Health-care Home

In order to show our approach, we simulated an health-care home provided with a smart system for supporting guest activities.

The simulation framework is responsible for time advancing. In particular, we adopt a discrete time-stepped simulation. The virtual time advances with a fixed interval (i.e.: each time step is 30 minutes).

In the simulated health-care home, each guest is provided with some devices for individual recognition and with wearable sensors for monitoring physiological parameters (i.e: body temperature, heart rate, blood pressure etc...). The smart environment owns a knowledge base containing information about health-care home guests (i.e: diseases, pharmacological therapy, clinical investigations etc...). The health-care home is endowed with a plurality of sensors and devices. The smart system provides each guest with a personal virtual tutor that supports daily activities such as taking medicines, doing medical examinations and so on.

Fig. 3 shows a screen-shot of health-care house simulated with ICasa[12]. It is composed of several bedrooms for guests,

Fig. 3: The simulated health-care house

a medical room and a restaurant. Each room is endowed with a presence sensor that detects the occupancy of a space by people. Some monitors are located in common areas in order to show personalized advertisements to guests.

An excerpt of user goals the system can satisfy is described in the following by adopting the GoalSpec specification.

goal_1: *WHEN is_Time_to_WakeUp(guest) THE system SHALL ADDRESS light(guest_room, on) AND alarm(guest_room, on) AND guest(awake).*

goal_2: *WHEN is_Time_to_Have_Lunch(guest) THE system SHALL ADDRESS restaurant_service(guest, available).*

goal 3: *WHEN is_Time_to_Take_Medicine(guest) THE system SHALL ADDRESS at(guest, medical_room) AND done(took_medicine).*

...

goal n: *WHEN is_Time_to_Meet_Doctor THE system SHALL ADDRESS done(checked, guest)*

The first goal means that the Smart Environment has to reach the desired state of the world in which the light and the alarm of guest's room are turned on and guest is awake. The second one makes the restaurant available to guests. The third one indicated that the smart environment has to fulfill the state of the world in which the guest is at medical room and he takes his medicine. Finally, the system allows to periodically make medical checks for monitoring the wellness of guests.

For the sake of clarity, in the following scenarios we exemplify norms in SBVR language. It is worth noting that our application translates SBVR rule according to Definition 3. In the following we provide an example:

SBVR norm: *It is permitted that guests have lunch if they have performed sports activities*

System norm: *norm(type(permission), role(guest), goal (have_launch), condition(is(guest,diabetic))).*

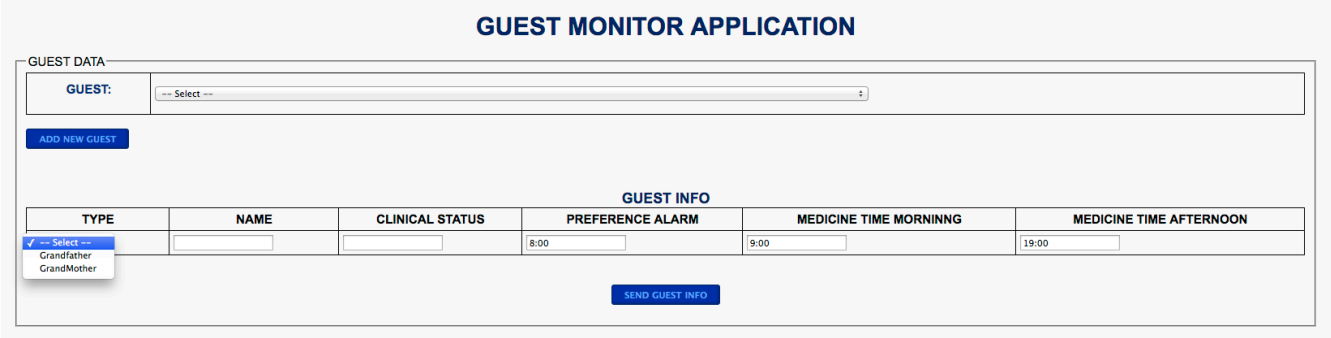


Fig. 4: Guest Monitor Application

We initially assume that the guests of the health-care home are individuals without particular diseases. The system behaves in the same way for each guest thus satisfying the previous goals. Over time, other people with particular illnesses are received and their preferences and clinical status are registered in the system. The manager of the health-care home inserts new norms in the system in order to change system behavior accordingly to the new situations. In particular, we simulated the following scenarios.

a) Permission Norm Injection: During the normal execution of the system, the manager of the health-care home gives the permission to guests that get already had some sports activities to have lunch before the established time. Thus, he introduces into the system the following norm:

norm_1: *It is permitted that guests have lunch if they have performed sports activities.*

In such case, the system differently behaves according to the particular guest. It allows only sportive guests to go to the restaurant before the regular time for lunch (i.e.: 13 o'clock).

In Fig.5 two guests (Paul and Maria) want to go to the restaurant at midday. In such scenario, Paul went to jogging in the morning. Thus, the system permits Paul to go to the restaurant before the regular time for lunch. Conversely, it is prohibited for Maria that has not performed any sport activities.

Fig. 5: The system behaves differently for different users.

b) Diabetic Guest: A new guest (Mark) is received in the health-care home. It is the first diabetic patient of the house. Fig.4 shows a screen-shot of the simulation framework that allows to introduce new guest into the simulated environment along with some guest information and preferences. The following norm is injected at run-time into the system:

norm_2: *It is prohibited that a guest has lunch before taking insulin if the guest is diabetic.*

In such scenario, the diabetic guest goes to the restaurant at 13 o'clock but before taking insulin. In such case, the system does not allow Mark to eat at the restaurant and its virtual tutor advises him that has to take the medicine before lunch (see Fig.6 (a)). Then, Mark goes to the medical area and he takes insulin (see Fig.6 (b)). Thus, the system updating the state of the world will permit Mark to have lunch.

c) Heart-Rate Control: The manager of the health-care home wants to provide better services to his guests. Thus, he introduces a new norm for allowing the system to suggest a medical check when some physiological parameters are out of their normal range.

norm_3: *It is permitted to do a medical check if a guest has an irregular heart rate.*

The introduction of the previous norm modifies the normal system behavior, thus allowing not only periodic medical check but also appropriate ones. In such case, the system that is able to monitor the health of guests, when it perceives a heart rate out of normal range (see Fig.7 (a)), it suggests the guest to go to see a doctor and it alerts the nurses (Fig.7 (b)).

d) Conflict Scenario: In such scenario, the diabetic guest goes to the restaurant after having had some gym and before taking insulin. This particular state of the world cause a conflicting situation among two norms. In such case, the system notifies the manager that there is a conflicting situation due to the simultaneous application of the norm_1 and norm_2. If the manager does not modify anything, the system applies the *legis posterior* criterion thus prohibiting the user to go to the restaurant before taking insulin (see Fig.8).

In the following section, we draw some conclusions.

(a) The guest does not use restaurant services. The virtual tutor advises the guest to take insulin.

(b) The guest goes to medical area and he takes insulin. The system updates the new state of the world.

Fig. 6: Diabetic Guest scenario

(a) The system perceives an irregular heart rate and it alerts the nurses. The virtual tutor advises the guest to go to make a check.

(b) The guest goes to medical area.

Fig. 7: Heart-Rate Control scenario

VIII. Conclusions

Norms are well know means for regulating the behavior of multi-agent systems, thus ensuring the fulfillment of the overall objective of the society the agents live in. This work takes place in the context of self-adaptive and self-organized systems that consider goals as key elements. They are systems conceived for satisfying user requirements that can be also established at run-time. These systems may evolve over time by increasing the objectives they are able to satisfy. In such context, we defined a normative framework that owns a direct link with the goals the system is able to pursue thus hiding the agent level. In our approach we consider the goals as a particular kind of obligation that have to be satisfied under certain conditions. Besides, we see permission and prohibition norms as promoters or inhibitors of the system in pursuing its goals. By introducing norms at run-time we also make the system more flexible to environment changes and able to self-adapt to new normative contexts in which it could be employed.

Moreover, the proposed algorithm takes into consideration the simultaneous presence of multiple norms related to the same goal, thus determining their joint effect. The approach for run-time injection also provides a means for checking norm conflicts and inconsistencies. Such an approach also implements a recovery mechanism based on the *legis posterior* criterion for solving inconsistencies among norms.

Finally, the simulated environment provides us a means for testing new hypotheses about a real system. In particular, in our future works, it will used for studying the consequences of norms injection in order to discover undesired behaviors of the system or new kinds of conflicting situations.

References

[1] Diane J Cook. Multi-agent smart environments. *Journal of Ambient Intelligence and Smart Environments*, 1(1):51–55, 2009.

[2] Diane J Cook, Michael Youngblood, and Sajal K Das. A multi-agent approach to controlling a smart environment. *Designing smart homes*, 4008:165–182, 2006.

[3] Mathieu Vallée, Fano Ramparany, and Laurent Vercouter. *A multi-agent system for dynamic service composition in ambient intelligence environments*. Citeseer, 2005.

[4] Laura Klein, Jun-young Kwak, Geoffrey Kavulya, Farrokh Jazizadeh, Burcin Becerik-Gerber, Pradeep Varakantham, and Milind Tambe. Coordinating occupant behavior for building energy and comfort management using multi-agent systems. *Automation in Construction*, 22:525–536, 2012.

[5] Huib Aldewereld, Frank Dignum, Andrés García-Camino, Pablo Noriega, Juan Antonio Rodríguez-Aguilar, and Carles Sierra. Operationalisation of norms for electronic institutions. In *Coordination, Organizations, Institutions, and Norms in Agent Systems II*, pages 163–176. Springer, 2007.

[6] Guido Boella and Leendert WN van der Torre. Regulative and constitutive norms in normative multiagent systems. *KR*, 4:255–265, 2004.

[7] Mehdi Dastani, John-Jules Meyer, and Nick Tinnemeier. Programming norm change. *Journal of Applied Non-Classical Logics*, 22(1-2):151–180, 2012.

[8] Felipe Meneguzzi and Michael Luck. Norm-based behaviour modification in bdi agents. In *Proceedings of The 8th International Conference*

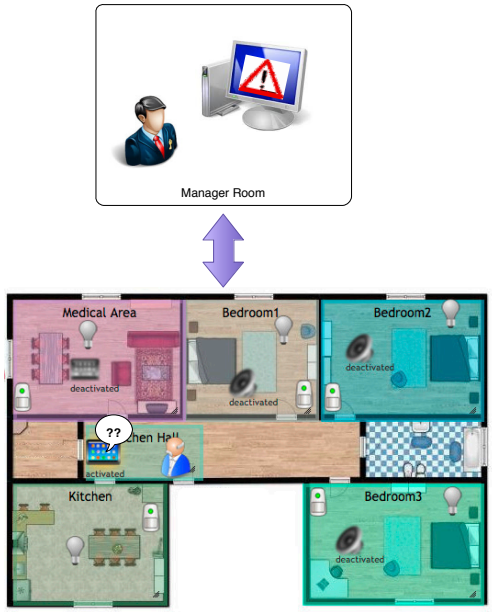


Fig. 8: A conflicting situation is detected by the system. It is notified to the manager. The *legis posterior* criterion is applied.

on Autonomous Agents and Multiagent Systems-Volume 1, pages 177–184. International Foundation for Autonomous Agents and Multiagent Systems, 2009.

[9] Luca Sabatucci, Patrizia Ribino, Carmelo Lodato, Salvatore Lopes, and Massimo Cossentino. Goalspec: A goal specification language supporting adaptivity and evolution. In *Engineering Multi-Agent Systems*, pages 235–254. Springer, 2013.

[10] Object Management Group. Semantics of business vocabulary and business rules (sbvr). version 1.3. may 2015.

[11] Massimo Cossentino, Carmelo Lodato, Salvatore Lopes, and Luca Sabatucci. Musa: a middleware for user-driven service adaptation. *in proc. of XVI Workshop "Dagli Ogetti agli Agenti", Napoli, June, 17-19, 2015*, 1382, 2015.

[12] Icasa : a dynamic pervasive environment simulator http://adele.imag.fr/icasa-a-dynamic-pervasive-environment-simulator/.

[13] Paolo Bresciani, Anna Perini, Paolo Giorgini, Fausto Giunchiglia, and John Mylopoulos. Tropos: An agent-oriented software development methodology. *Autonomous Agents and Multi-Agent Systems*, 8(3):203–236, 2004.

[14] Graham Witt. *Writing Effective Business Rules: A Practical Method*. Elsevier, 2012.

[15] Frank Dignum. Autonomous agents with norms. *Artificial Intelligence and Law*, 7(1):69–79, 1999.

[16] Thomas Agotnes, Wiebe Van Der Hoek, JA Rodriguez-Aguilar, Carles Sierra, and Michael Wooldridge. On the logic of normative systems. In *Proceedings of the Twentieth International Joint Conference on Artificial Intelligence (IJCAI'07)*, pages 1181–1186, 2007.

[17] Marek Sergot. Action and agency in norm-governed multi-agent systems. In *Engineering Societies in the Agents World VIII*, pages 1–54. Springer, 2007.

[18] Mehdi Dastani, Nick AM Tinnemeier, and John-Jules Ch Meyer. A programming language for normative multi-agent systems. *Multi-Agent Systems: Semantics and Dynamics of Organizational Models*, pages 397–417, 2009.

[19] Natasha Alechina, Mehdi Dastani, and Brian Logan. Programming norm-aware agents. In *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems-Volume 2*, pages 1057–1064. International Foundation for Autonomous Agents and Multiagent Systems, 2012.

[20] Martin J Kollingbaum and Timothy J Norman. A contract management framework for supervised interaction. In *Working Notes of the 5th UK Workshop on Multi-Agent Systems UKMAS 2002*, 2002.

[21] Nick Tinnemeier, Mehdi Dastani, and John-Jules Meyer. Programming norm change. In *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems: volume 1-Volume 1*, pages 957–964. International Foundation for Autonomous Agents and Multiagent Systems, 2010.

[22] Max Knobbout, Mehdi Dastani, and John-Jules Ch Meyer. Reasoning about dynamic normative systems. In *Logics in Artificial Intelligence*, pages 628–636. Springer, 2014.

[23] Jie Jiang, Huib Aldewereld, Virginia Dignum, and Yao-Hua Tan. Compliance checking of organizational interactions. *ACM Transactions on Management Information Systems (TMIS)*, 5(4):23, 2015.

[24] Wamberto Vasconcelos, Martin J Kollingbaum, and Timothy J Norman. Resolving conflict and inconsistency in norm-regulated virtual organizations. In *Proceedings of the 6th international joint conference on Autonomous agents and multiagent systems*, page 91. ACM, 2007.

[25] Marc Esteva, Wamberto Vasconcelos, Carles Sierra, and Juan A Rodriguez-Aguilar. Norm consistency in electronic institutions. In *Advances in Artificial Intelligence–SBIA 2004*, pages 494–505. Springer, 2004.

[26] Martin J Kollingbaum, Timothy J Norman, Alun Preece, and Derek Sleeman. Norm conflicts and inconsistencies in virtual organisations. In *Coordination, organizations, institutions, and norms in agent systems II*, pages 245–258. Springer, 2007.

[27] Eric Yu. Modelling strategic relationships for process reengineering. *Social Modeling for Requirements Engineering*, 11:2011, 2011.

[28] Raymond Reiter. *On closed world data bases*. Springer, 1978.

[29] Andrés García-Camino, Pablo Noriega, and Juan-Antonio Rodríguez-Aguilar. An algorithm for conflict resolution in regulated compound activities. In *Engineering Societies in the Agents World VII*, pages 193–208. Springer, 2006.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1491–1494
DOI: 10.15439/2016F506
ACSIS, Vol. 8. ISSN 2300-5963

Simulation Goals and Metrics Identification

Valeria Seidita[§‡]
[§]Dip. di Ing. Chimica,Gestionale,Informatica,Meccanica
University of Palermo
Palermo, Italy
Email: {valeria.seidita}@unipa.it

Patrizia Ribino[‡], Massimo Cossentino[‡], Carmelo Lodato[‡]
[‡] Istituto di Reti e Calcolo ad Alte Prestazioni
Consiglio Nazionale delle Ricerche
Palermo, Italy
Email: {ribino,cossentino,c.lodato}
@pa.icar.cnr.it

***Abstract*—Agent-Based Modeling and Simulation (ABMS) is a very useful means for producing high quality models during simulation studies. When ABMS is part of a methodological approach it becomes important to have a method for identifying the objectives of the simulation study in a disciplined fashion. In this work we propose a set of guidelines for properly capturing and representing the goals of the simulations and the metrics, allowing and evaluating the achievement of a simulation objective. We take inspiration from the goal-question-metric approach and with the aid of a specific problem formalization we are able to derive the right questions for relating simulation goals and metrics.**

I. Introduction

AGENT-BASED simulation studies are effective tools for addressing problems that in real life require several resources. Despite their power, the key of success in a simulation study is to follow a comprehensive process for building simulated systems that produce acceptable and credible results.

In particular, it is widely known that a very important activity of a simulation study is to clearly establish the simulation objectives. Such objectives, or simulation goals, are the guiding element of a simulation study, thus they have to be defined at the beginning of a simulation process [1][2][3][4][5].

In classical simulation studies, the simulation team should develop a list of specific questions that the simulation model should address and develop a list of performance metrics for evaluating goal satisfaction [1][6]. How to perform this activity is not clear enough. It is often committed to the experience of the simulation team.

We claim that problem formalization assumes a very important role in the identification of simulation goals and in this paper, we propose to identify simulation goals, starting from an ontological representation of the problem domain, and then to identify the performance metrics for evaluating goals. Our approach takes inspiration from the *Goal Question Metric* (GQM) approach proposed by Basili et al. [7].

GQM is an approach for expressing the goal of a software engineering study. It provides a method for defining goals, refining them into questions and finally developing a set of metrics needed to answer such questions.

For a couple of years we have been working on the development of a methodological approach covering the entire life cycle of a multi-agent simulation study taking into account each facet of the agent based modeling paradigm. Such a methodology is founded on the premise that a simulation study is conducted for testing new hypothesis on a model of a real system (i.e.: what happens if). In so doing, we are building our approach according to the phases of classical simulation studies [1][2] but providing specialized guidelines for using agent based simulation techniques. So far, we developed the initial activities of Simulation Problem Analysis phase, namely Problem Formalization [8]. Problem formalization aims at formalizing simulation problems by adopting an ontological representation of the simulation problem domain.

In this paper, we employ such a representation (namely Problem Ontology) as a guide for generating questions and specifying the metrics (and also parameters) for answering those questions thus providing a structured way in order to perform a GQM-like approach in the field of simulation studies.

The Problem Ontology has to be intended as a representation of the domain on which the simulation model is being constructed. Such a domain contains, the objects involved in the problem, their behaviors, rules and organizational aspects; all these elements find a perfect mapping with the elements Basili et al. define for applying the GQM approach.

For the scope of this paper, we want to focus on the elements that can be defined *Problem Ontology*:

- *Active Entity*: an active entity allows to represent concepts that are specified in the problem statement that perform dynamic actions.
- *Action*: an action in the problem ontology represents concepts in the problem statement describing the fact or process of doing something, typically to achieve an aim.
- *Object*: it represents something to which a specified action is directed.
- *Predicate*: It allows to express a property, a state or more generally it allows to specify a concept.

The main contribution of this work is a set of guidelines for applying the GQM approach for simulation studies that use the *Problem Ontology* as an input. Such guidelines result in two workproducts, the Simulation Goal document and the Simulation Metrics document.

The rest of the paper is organized as follows. Section II presents an overview of the GQM approach and points out the main elements and the rationale we used for applying the GQM to our case. Section III introduces the proposed approach for simulation goal and metrics identification by describing the steps to perform for obtaining metrics and

978-83-60810-90-3/$25.00

parameters. Finally, in section IV some discussions and conclusions are drawn.

II. Goal Question Metric Approach

The *Goal Question Metric* (GQM) approach, proposed by Basili et al. [7][9][10][11], is a paradigm that links metrics with the general goal of a process or project. The aim is to generate questions, whose answers are known, thus establishing if all the goals have been reached. Metrics are the means for answering the questions.

GQM approach relates to software metrics, indeed it is based on three main principles, classifying the *entities* to be examined, determining relevant measurement goals and finally determining the level of maturity of the organization. This latter principle is obviously strictly related to software production and has not been considered in this paper. The initial step of a software measurement is to identify the *objects* (i.e: processes, products and resources) and the *attributes* (i.e.: internal or external) to measure. A process is an activity related to software, products are artefacts or deliverable resulting from an activity and a resource is an entity required by a process. Attributes characterize an entity. An internal attribute refers to the entity itself and may be measured by investigating only the entity without referring to its behavior; an external attribute refers to how an entity relates to its environment.

The GQM model is based on three levels: Conceptual, Operational and Quantitative. The *Conceptual level* concerns the definition of *Goals*. In GQM, a goal is defined for an object with respect to various models of quality, from several perspectives and according to a particular environment. The *Operational level* concerns the identification of *Questions*. The questions are articulated for defining models of the object under study and for characterizing the object with respect to a selected quality issue. The *Quantitative level* is related to the definition of *Metrics*. A metric is associated with every question in order to answer it in a measurable way.

GQM is typically described as a six-step process. The first three steps concern how to use business goals for driving the identification of the right metrics. The last three steps are about gathering the measurement data and making effective use of the measurement results to drive decision making and improvements.

Hence, in order to characterize a goal in a quantifiable way, it is necessary to relate the *goal* to an *object*, or an *entity*, that possesses some kind of *attributes* related to itself or to the environment. A goal or a question may be ascribed to, at least, a couple entity - attribute. The answer to a question means that we give a measurement to an attribute. In particular, the GQM approach uses three elements for specifying the goal: purpose, issue and object. We base our work on this assumption and extend the first three steps of GQM. In particular, we provide guidelines for identifying elements the simulation goals are related to by adopting an ontological formalization of the problem domain.

III. Simulation Goal and Metrics Identification

All simulation studies, in general, and agent simulation study, in particular, greatly depend on how the simulation problem analysis is conducted. Simulation problem analysis aims at identifying the goals of the simulation study with respect to the domain under study.

During simulation studies some hypothesis are investigated through manipulating some independent variables that affect dependent variables and consequently the validation of the initial hypothesis. In this process, the identification of the simulation goals during the early analysis phase provides a prerequisite for identifying metrics, or parameters, for the evaluation of the simulation hypothesis.

Sentences like "*Improve the timeliness of change request processing during the maintenance phase of the life cycle of a system*" that can be found in the description of the simulation study, hence the in some kind of *problem statement*, give a hint on what the simulation study is intended for, the objective that has to be reached. Sometimes, it is not obvious or immediate to identify the independent or observable variable related to the simulation goal.

In this section we show the rationale we propose for conducting two activities: the *Simulation Goal Identification* and then the *Entity Parameter Identification*. These activities aim at identifying and modeling the goal(s) of the simulation study and the related parameters to be considered for the agent-based software development. These two activities are strictly tied and are part of a complete methodological approach for supporting agent simulation studies. We claim that the identification and the representation of simulation goals greatly descend from the problem formalization.

Problem formalization is the key element in all engineering activities for producing software products of high quality. Complete and well defined simulation models require a careful analysis of the problem domain and a detailed problem formalization phase [1][12].

We propose to join the problem formalization to the *Goal Question Metric* approach. We already discussed how to perform problem formalization and the proposed results are shown in [8][13]. The result of the *Problem Formalization* activity is a work product, the Problem Ontology diagram, containing all the elements representing the domain under study (see section I). In what we propose, the Problem Ontology may be used as a guide for generating questions and then specifying the measures for answering questions in a GQM like approach.

Using GQM, a goal is expressed through the triplet <*purpose, issue, object*>; it defines an object with respect to a purpose and some quality issues. This kind of triplet, in the GQM, is the starting point for extracting all the useful parameters for a specific situation.

In a simulation problem statement the goal is generally expressed through a sentence and, from the analysis of this sentence, it is possible to identify the previous triplet for applying GQM. Suppose the Problem formalization activity

resulted in a Problem Ontology diagram, where all the active entities, the objects, the actions and the predicates, representing the simulation study, have been represented [1]. The steps we propose are:

- Identify goals - the inputs of this step is the Problem Statement. An analysis of sentences in terms of nouns and verbs is required
  The work to be done starts with the identification of the goal, or the goals, of the simulation study; this activity implies a careful analysis of the Problem Statement in order to identify sentences from which goals may be extracted. Since, as Basili et al. [7] highlight, *"typical goals are expressed in terms of productivity, quality, risk, satisfaction"*, useful sentences are those containing verbs such as assess, improve, evaluate and so on.
  Sometimes, the sentence may be in the affirmative or in the interrogative form and some others it may not be simple to identify the sentence related to the simulation goal. In this latter case a refinement of the problem statement should be required.
- Identify <*purpose, issue, object*> - This step implies to analyze the syntax of the sentence containing the simulation goal. The verb is generally followed by a direct object and then by a genitive case (if not the sentence may be reformulated).
  For instance, the sentence *"Improve the timeliness of change request processing during the maintenance phase of the life cycle of a system"* leads to the triplet <improve, timeliness, change request processing>.
- Prepare questions - The main input of this step is the Problem Ontology diagram (POD). In order to obtain all the possible questions for one goal, it is required to explore all the relationships between the object in the triplet and all the other elements in the Problem Ontology diagram. If there exist a relation between an element of the POD and the object of the goal then there may be some parameters, attributes or metrics that once modified may affect the object of the goal and, at the same time, the overall simulation thus giving means for assessing or evaluating the initial hypothesis the simulation study is created for.
  For instance, suppose that the *change request processing* is related to an object like *time*, then it should be possible that the amount of time is a value of a parameter that may affect the improvement of the timeliness.

It is worth to note that two situations may happen. The POD may present the object of the simulation model or not. The first situation happens when the object is a product or a resource and it has been identified during the previous activity as an «object» or an «active entity». The second situation, instead, happens when the goal's object is a process or simply it has not been identified during *Problem Formalization*, in this latter case a refinement activity of the POD is necessary and should be performed.

[1]The reader may have a look at [8] for a detailed example.

The Problem Ontology diagram is very important in all the previous steps. The way we employ POD is the very contribution of this paper in fact it provides an important means for supporting the analysis and identification of goals and parameters. The efficiency increasing in the identification of the right elements to consider is also obtained.

The work products resulting from the *Simulation Goal Identification* and the *Entity Parameter Identification* are Simulation Goal (SG) document that is composed of two different diagrams and the Simulation Metric (SM) document. The first contains a diagram representing the view on the simulation goal(s), its relationships with the elements of the POD and some possible new goals coming from reasoning about the relationships among the main simulation goals and POD objects. In the second the simulation goal(s) is illustrated with all the new goals identified and the kind of contribution they give to the main goal(s).

A brief example of the resulting work products is given in Fig. 1.a) and Fig. 1.b) as regard the SG and Fig. 1.c) as regard SM. These documents are related to a specific case in which, from the Problem Ontology, the analyst has identified the sentences "How may we improve the value of the throughput of the warehouse?" and "Another aim of this study is to reduce the costs of warehouse managing". The related triplets are: <improve, value, throughput> and <reduce, cost, management>. By Applying the proposed approach and by iterating them on the elements of the POD a portion of the obtained results is shown in the figures. The second work product shows questions and parameters. Obviously, for space reasons, this example cannot be complete and exhaustive but is intended to give a view on the work products.

IV. Discussions and Conclusions

Determining the goals of a simulation study is a very important task because from these we may obtain the set of metrics and test parameters that are typical of scientific experiment, such as simulation ones.

Simulation goals identification is generally a mental and speculative activity tied to the skills and knowledge of the analysts and very few methodological recommendations exist in literature. Since we need to extract some metrics and parameters, that are formal elements, we want to apply a disciplined way in order to avoid the risk of forgetting or omitting some important elements of the problem domain.

In this paper, we have proposed a GQM like approach for determining such a parameters starting from the simulation goal identification. This work is a further step of a more complete one: a complete design methodology for developing agent based simulation studies, this methodology also includes the agent-based system development.

One of the most important parts, already developed, is the problem domain formalization that becomes the most important and necessary input to the simulation goals identification. In some previous works we described how to represent the problem domain as an ontological (POD) representation

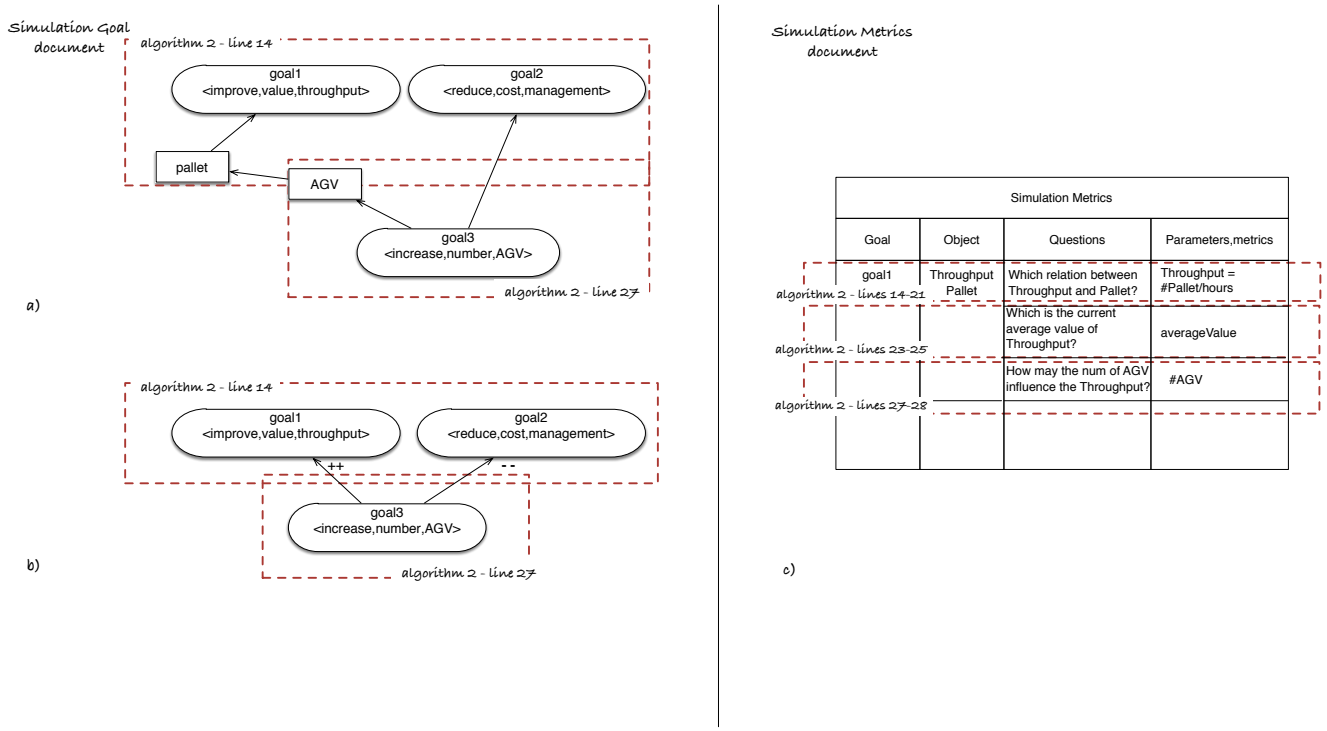


Fig. 1. SG and SM documents - an example

containing objects and entities from which it is possible to identify behaviors and interactions with the environment.

In the present work, we combine that with the well known goal-question-metric approach (GQM) in order to create a disciplined and structured method for the identification of the goals and the related metrics and parameters useful for understanding if the goals have been satisfied for the simulation system.

The use of a problem formalization and of the POD ensures a final result coherent with the domain under study and complete since the POD, while applying GQM, is explored in its entirety. For sure this approach depends on the presence of a good and well done Problem Ontology Diagram but, the GQM approach is a valid support that, once applied, also allows to reason on the problem domain and on the POD and potentially to refine it. The same we make with the Problem Formalization [8], both are iterative and incremental approaches that lead to the goal simulation model and at the same time allow to refine the POD towards a status where, reasonably, nothing has been omitted.

In the future, we plan to complete the methodological approach and to validate the whole approach using the true experimental set up coming from the simulation model.

REFERENCES

[1] I. Carson and S. John, "Introduction to modeling and simulation," in *Proceedings of the 36th conference on Winter simulation*. Winter Simulation Conference, 2004, pp. 9–16.

[2] O. Balci, "Guidelines for successful simluation studies (tutorial session)," in *Proceedings of the 22nd conference on Winter simulation*. IEEE Press, 1990, pp. 25–32.

[3] ——, "Validation, verification, and testing techniques throughout the life cycle of a simulation study," *Annals of operations research*, vol. 53, no. 1, pp. 121–173, 1994.

[4] F. Klügl, "Multiagent simulation model design strategies." in *MALLOW*, 2009.

[5] A. Garro and W. Russo, "easyabms: A domain-expert oriented methodology for agent-based modeling and simulation," *Simulation Modelling Practice and Theory*, vol. 18, no. 10, pp. 1453–1467, 2010.

[6] F. Klügl, "Engineering agent-based simulation models?" in *Agent-Oriented Software Engineering XIII*. Springer, 2012, pp. 179–196.

[7] R. Van Solingen, V. Basili, G. Caldiera, and H. D. Rombach, "Goal question metric (gqm) approach," *Encyclopedia of software engineering*, 2002.

[8] M. Cossentino, C. Lodato, P. Ribino, and V. Seidita, "A heuristic for problem formalization in agent based simulation studies," in *Computer Science and Information Systems (FedCSIS), 2015 Federated Conference on*. IEEE, 2015, pp. 1733–1743.

[9] V. R. Basili and D. M. Weiss, "A methodology for collecting valid software engineering data," *IEEE Transactions on Software Engineering*, vol. 6, no. SE-10, pp. 728–738, 1984.

[10] V. R. Basili, S. Green, O. Laitenberger, F. Lanubile, F. Shull, S. Sørumgård, and M. V. Zelkowitz, "The empirical investigation of perspective-based reading," *Empirical Software Engineering*, vol. 1, no. 2, pp. 133–164, 1996.

[11] N. Fenton and J. Bieman, *Software metrics: a rigorous and practical approach*. CRC Press, 2014.

[12] A. M. Law, "How to build valid and credible simulation models," in *Proceedings of the 40th Conference on Winter Simulation*. Winter Simulation Conference, 2008, pp. 39–47.

[13] P. Ribino, V. Seidita, C. Lodato, S. Lopes, and M. Cossentino, "Common and domain-specific metamodel elements for problem description in simulation problems," in *Computer Science and Information Systems (FedCSIS), 2014 Federated Conference on*. IEEE, 2014, pp. 1467–1476.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1495–1504
DOI: 10.15439/2016F407
ACSIS, Vol. 8. ISSN 2300-5963

Simulating Large-scale Aggregate MASs with Alchemist and Scala

Roberto Casadei
Università di Bologna, Italy
roberto.casadei12@studio.unibo.it

Danilo Pianini
Università di Bologna, Italy
danilo.pianini@unibo.it

Mirko Viroli
Università di Bologna, Italy
mirko.viroli@unibo.it

***Abstract*—Recent works in the context of large-scale adaptive systems, such as those based on opportunistic IoT-based applications, promote aggregate programming, a development approach for distributed systems in which the collectivity of devices is directly targeted, instead of individual ones. This makes the resulting behaviour highly insensitive to network size, density, and topology, and as such, intrinsically robust to failures and changes to working conditions (e.g., location of computational load, communication technology, and computational infrastructure). Most specifically, we argue that aggregate programming is particularly suitable for building models and simulations of complex large-scale reactive MASs. Accordingly, in this paper we describe SCAFI (Scala Fields), a Scala-based API and DSL for aggregate programming, and its integration with the ALCHEMIST simulator, and usage scenarios in the context of smart mobility.**
Keywords — aggregate programming, Scala, DSL, simulation.

I. INTRODUCTION

APPLYING multiagent systems (MASs) in the context of large-scale distributed systems is known to be hard in general, due to the ineluctable need to take into account issues such as communication, robustness, consistency and performance. The situation is then becoming harder and harder especially in recently emerging distributed computing scenarios, such as pervasive computing or IoT, due to the number of computational entities, the complexity of interactions, the presence of natural limitations related to energy, communication and processing, and the tight connection with the physical world and human users—quintessential source of unpredictability. Achieving a sound development of MAS applications in this context, so as to ensure desired properties of robustness and scalability, calls not just for better algorithms and computing frameworks, but possibly for whole new paradigms.

Recent research in collective adaptive software systems proposed *aggregate computing* [1] as a promising approach generalising over several prior models and languages addressing computations over collections of spatially-situated systems [2]. Essentially, aggregate computing allows one to express complex system-wide, global-level computations involving large sets of devices in a fully declarative way, promoting decomposition and resiliency. Aggregate computing can be formally grounded in the field calculus [3], a core language able to express complex patterns of information diffusion and aggregation. Resiliency is guaranteed by self-organisation: aggregate programs can be compiled into repetitive local tasks to be executed by the single agent, promoting identification of robust building blocks of aggregate behaviour [4].

Though naturally applicable to swarm-like reactive MASs, aggregate computing is also of interest when stronger notions of agency enter the picture: as discussed in [5], aggregate programs can be seen as "aggregate plans," namely, operational instructions for deliberative agents that guide the cooperative behaviour of a team.

In order to more deeply investigate the impact of this new paradigm to the mainstream development of large-scale MASs, in this paper we explore and propose an extension of the Alchemist simulator [6] with `scafi`[1] [7], a Scala framework that provides an internal domain-specific language (DSL) for specifying aggregate computations via a simple API that well integrates with the advanced typing features of Scala (inference, genericity, implicits) and its library. This allows to smoothly simulate complex collective adaptive behaviours on top of Scala mainstream programming.

The remainder of this paper is organised as follows: Section 2 introduces aggregate programming, Section 3 presents the `scafi` framework and example specifications, Section 4 depicts Alchemist and the integration of `scafi`, Section 5 discusses case studies in the context of smart mobility, and Section 6 presents related works before finally drawing conclusions.

II. AGGREGATE PROGRAMMING

Aggregate programming [1] is a novel approach to (large-scale) distributed systems engineering that supports the specification of collective behaviours in a simple, high-level, and composable way. The key idea is to shift programming from the traditional single-device viewpoint to a global viewpoint where the programmable entity is the *aggregate* body of computational elements constituting a system. This way, programmers are no longer required to solve the intricate local-to-global problem, i.e., building the desired emergent phenomenon by specifying how each component behaves and interacts with others in a fully bottom-up fashion; instead, it is possible to focus on *what* the system should exhibit, and let the computational platform define – under-the-hood – *how* the computation is carried on by the interaction of individual entities. It essentially solves the inverse, global-to-local mapping problem.

[1] http://scafi.apice.unibo.it

An immediate consequence is the independence of aggregate computations from the physical implementation details of systems, which is realised by suitably abstracting spatial distribution, topology and interaction. More specifically, as realised by space-time programming approaches [2], logical or physical neighbouring of nodes can be exploited to make interaction implicit. In addition, this notion is instrumental for the conceptual connection with systems where locality might play a major role in communication.

The main programming abstraction in aggregate programming is the *computational field* [3] (or field for short), a notion already used in the MAS community [8], [9]. Generalising the notion of (gravitational, electromagnetic) field in physics, a computational field is a function that, at a given moment in time, maps each point in space to a computational object; when considering the space discretised by a networked set of situated agents, each sample represents the outcome of computation for that agent. The key insight of the approach consists in the ability to specify collective behaviours (i.e., aggregate computations) by algorithms expressed as a functional composition of fields: from input fields representing information sensed from the environment, up to output fields representing some form of actuation. In other words, aggregate computations are represented by a declarative specification of functional operations involving collective data structures, though, under-the-hood, they are turned into repetitive, gossip-like interactions between individual agents—namely, relying on known low-level self-organisation patterns [10].

This approach is shown to support a solid engineering methodology, in which composable and reusable high-level library components of aggregate behaviour can be defined that are provably resilient [4].

III. Aggregate Programming in Scala

The aggregate computing idea can be naturally supported by a programming language, used to specify aggregate behaviours. Among different choices one can make to frame one such language (as a DSL, as an API, and so on), in this paper we explore the idea of using the Scala programming language [11] as the host language for building an aggregate programming platform. This is motivated by both technical and practical reasons.

Scala is a modern language for the JVM which integrates the object-oriented and functional paradigms in a seamless way, hence we can combine the typically rich and expressive OO libraries and data structures, with functional programming as promoted by aggregate computing. Scala has a powerful and expressive type system, combining the advantages of static type checking with a concise syntax for productivity—thanks to type inference, the implicits system, generic and functional programming features, and ad-hoc syntactic sugar. This allows library designers to create Scala APIs which actually have a DSL-like flavour, which are thus perceived by users as "embedded languages." Moreover, Scala is now becoming a standard de facto for the construction of platforms for distributed processing (frameworks like Akka actors, and Apache Kafka, Storm and Spark are essentially Scala-based): this makes it the ideal language to target a platform for aggregate computing—though this issue is not discussed here further.

Accordingly, we propose `scafi` (Scala fields) [7], a framework consisting of two main parts:

1) *aggregate programming support*, by a Scala-internal DSL that provides a syntax and the corresponding semantics for the constructs of the computational field calculus [3], by which aggregate computations are naturally expressed and seamlessly combined in code; and
2) *aggregate platform support*, allowing configuration and execution of such code in actual distributed setups.

`scafi` explicitly addresses the construction of concrete aggregate computing applications: however, it can also play the role of a language to validate complex MAS algorithms, as meta-model for simulations—in the next section, in fact, we shall describe its integration with a full featured simulator, Alchemist.

A. Computational field calculus in Scala

An aggregate system consists of a (possibly mega-scale) number of computational devices or agents, all executing the same aggregate program at asynchronous rounds of computation. According to contextual information (e.g, sensor values), different such computational devices may take different branches of computation, i.e., computing sub-fields in different domains of execution. Interaction depends on a notion of locality, i.e., a device can communicate to all its neighbours, as defined by an application-specific proximity relation. The communication is carried out by repeatedly broadcasting the latest computed state to the neighbourhood: the shape of this state, and how it affects and gets affected by computations, is precisely defined by our language semantics [3]. The basic primitives of the field calculus (described below, in turn) are declared in the `Constructs` trait, implemented by the framework and mixed-in in any library of user-defined aggregate functions:

```
trait Constructs {
  def rep[A](init: A)(fun: (A) => A): A
  def nbr[A](expr: => A): A
  def foldhood[A](init: => A)(acc: (A,A)=>A)(expr: => A): A
  def branch[A](cond: => Boolean)(th: => A)(el: => A): A
  def aggregate[A](f: => A): A
  def sense[A](name: LSNS): A
  def nbrvar[A](name: NSNS): A
}
```

The field calculus constructs can be understood and described according to two complementary viewpoints:

1) *Local viewpoint* – refers to the traditional device-centric interpretation where an aggregate computation is considered in the context of a single device. The operational semantics of the field calculus is implemented according to this stance.
2) *Global viewpoint* – it corresponds to the natural semantics and refers to the aggregate-level interpretation of programs as computations running on whole fields

(i.e., spatial data structures mapping each device to some computational object).

B. Working with basic constructs

The most trivial program is one that simply evaluates to a constant value, such as a boolean, a number or a string. For instance, value

```
"Hello, World"
```

should be interpreted as a constant field evaluating that value everywhere; concretely, it results in the string `"Hello, World"` being the return value of the local computation of every device in the system.

Construct `sense` provides the means for reading a value from a local sensor, which enables context-sensitive behaviours. Expression

```
sense[Double]("temperature")
```

gets in any device a double value from the temperature sensor, creating the field of temperatures.

Change over time of a field can be realised via the `rep` construct, which produces a dynamically evolving field by repeatedly applying a state-transformation function. For example, it could be used for counting how many rounds a device performed since the beginning of computation:

```
// Initially 0; state is incremented at each round
rep(0){ _+1 }  // or equivalently: rep(0){ x => x+1 }
```

Note that the frequency at which devices compute rounds and hence send messages to neighbours can vary over time and from agent to agent: the aggregate computing model generally assumes partial synchronicity [12], though in most cases even full asynchrony of rounds can be assumed.

Communication with the neighbourhood is achieved via `nbr`, which gives a map from neighbouring devices to their value of the argument—essentially an observation primitive. Construct `nbr` has to be nested inside a `foldhood` operation, which reduces one such map back to a single value via a monoidal reduction (on top of it, derived `minHood`, `sumHood` and others are defined and will be used in next sections). Example applications of `foldHood` are as follows:

```
// Counting number of neighbours at each device
foldhood(0)(_+_){ nbr{1} } // sum 1 across neighbours

// Is sensor "sns" active in every neighbour?
foldhood(true)(_&&_){ nbr{ sense[Boolean]("sns") } }
```

In addition to local sensors, there is a notion of "environmental" sensor. `nbrvar` allows to extract values from a neighbouring sensor, which gives a sample for each neighbour. Thus, similarly to `nbr`, `nbrvar` has to be used within a `foldhood` operation.

```
def nbrRange(): Double = nbrvar[Double](NBR_RANGE_NAME)

// Compute the maximum distance of a neighbour
foldhood(Double.MinValue)(max(_,_)){ nbrRange() }
// equivalently: maxHood{ nbrRange() }
```

Operation `branch` splits the spatial domain of devices into two parts, or rather two sub-teams, according to a boolean field expressing some condition. Each of the two parts, compute a different sub-field in complete isolation. For example, if we would like to execute some aggregate computation only on a subset of the devices of the network (the complementary subset must not participate), we need a partition:

```
branch(sense[Boolean]("flag")){
  Double.MaxValue // not computing
}{
  compute(...)    // sub-computation
}
```

Construct `aggregate` is used to define the body of a new function that should work on whole fields. The devices running a given aggregate function constitute a partition, i.e., they are able to interact with each other via `nbr`. As an example, `branch` could be entirely rewritten using `aggregate`:

```
def branch[A](cond: => Boolean)(th: => A)(el: => A): A =
  mux(cond)(() => aggregate{ th })(() => aggregate{ el })()
```

The symbol `mux` used in this example is a built-in operator that provides a purely functional multiplexer.

C. Working with combinations

More elaborate aggregate behaviours can be defined by compositions of the basic constructs. In addition, specific parts of the program logic can be encapsulated into Scala functions. Simple examples involving these features include counting neighbours except the device itself:

```
// mid is a special sensor yielding the device unique id
def isMe = nbr{ mid() } == mid()

sumHood{ mux(isMe){0}{1} }
```

The paradigmatic example of computational field is known as gradient [10], computing in each node the distance (hop-by-hop, or estimated) from the nearest node where a `source` field holds a true value:

```
def nbrDist = nbrvar[Double](NBR_RANGE_NAME)

def gradient(source: Boolean): Double =
  rep(Double.MaxValue) { dist =>
    mux(source){ 0.0 } { minHood{ nbr{dist} + nbrDist } }
  }
```

D. Scaling with complexity

Even though the basic constructs of the field calculus are somewhat low-level, they can be further composed so as to define reusable library components that provide higher-level behaviours—in fact, the functional character of the approach promotes systematic factorisation of behaviour into reusable layers of increasing abstraction.

An initial set of general coordination operators has been identified in [1], [4]. These operators capture common patterns of distributed computation and also enjoy the *self-stabilisation* property, which ensures that a system, independently of the current state, will eventually reach a stable state in finite time

that is not affected by transitory events. Moreover, as the self-stabilisation property is preserved by composition [13], it is formally guaranteed that also composite structures and algorithms self-stabilise.

On top of such resilient building blocks, a sound development API can be defined, which in turn can be used to implement application-specific aggregate behaviours.

1) Gradient-cast: `G` simultaneously performs two tasks: i) builds a distance-gradient from the source (`src`) according to `metric`, and ii) builds accumulated values via `acc` along the gradient starting from `init` at the `src`. In `scafi`, it can be encoded as follows:

```
def G[V](src: Boolean, field: V, acc: V=>V, metric: =>Double)
        (implicit ev: OrderingFoldable[V]): V =
  rep( (Double.MaxValue, field) ){  // (distance,value)
    dv => mux(src) {
      (0.0, field)       // ..on sources
    } {
      minHoodPlus { // minHood except myself
        val (d, v) = nbr { dv }
        (d + metric, acc(v))
      }
    }
  }._2 // yielding the resulting field of values
```

The generic type `V` (and, in this case, also `Tuple2[+A,+B]`) must have an (implicit or explicit) instance of the `OrderingFoldable` type-class available so that `minHoodPlus` can work out the minimum value on the neighbourhood (by convention, `*hoodPlus` operators exclude the device itself from the neighbours set).

A `broadcast` operation can easily get built on top of `G`:

```
def broadcast[V](source: Boolean, field: V)
                (implicit ev: OrderingFoldable[V]): V =
  G[V](source, field, x=>x, nbrRange())
```

In the following example, we leverage `broadcast`, to diffuse across the whole network the distance between two devices:

```
def distanceTo(source: Boolean): Double =
  G[Double](source, 0, _ + nbrRange(), nbrRange())

def distBetween(source: Boolean, target: Boolean): Double =
  broadcast(source, distanceTo(target))

def isSource = sense[Boolean]("source")
def isObstacle = sense[Boolean]("obstacle")

distBetween(isSource, isObstacle)
```

and, most notably, we could realise an algorithm that builds a `width`-wide channel connecting a source and a destination:

```
def channel(src: Boolean, dest: Boolean, width: Double) =
  distanceTo(src) + distanceTo(dest) <=
    distBetween(src, dest) + width
```

2) Converge-cast: `C` is the dual of `G`: it collects information distributed across space by `accumulating` values down a `potential` field, starting with `local` at the sources (i.e., devices with no parent, located at the edge of the `potential` field):

```
def C[V](potential: V, acc: (V,V)=>V, local: V, Null: V)
        (implicit ev: OrderingFoldable[V]): V = {
  rep(local){ v =>
    acc(local, foldhood(Null)(acc){
      mux(nbr(findParent(potential)) == mid()){
        nbr(v)
      } {
        nbr(Null)
      }
    })
  }
}

def findParent[V](potential: V)
                 (implicit ev: OrderingFoldable[V]): ID = {
  mux(ev.compare(minHood{ nbr(potential) }, potential)<0 ){
    minHood{ nbr{ Tuple2[V,ID](potential, mid()) } }._2
  }{
    Int.MaxValue
  }
}
```

`C` and `G` can be combined to originate a self-stabilising `summarise` operator that first collects information across the space and then propagates back the computed summary:

```
def summarize(sink: Boolean,
              acc: (Double,Double)=>Double,
              local: Double,
              Null: Double): Double =
  broadcast(sink, C(distanceTo(sink), acc, local, Null))

def average(sink: Boolean, value: Double): Double =
  summarize(sink, (a,b)=>{a+b}, value, 0.0) /
    summarize(sink, (a,b)=>a+b, 1, 0.0)
```

3) Time-decay: The `T` operator can be used to condense information across time by decreasing the `initial` field according to a `decay` function:

```
def T[V](initial: V, floor: V, decay: V=>V)
        (implicit ev: Numeric[V]): V = {
  rep(initial){ v =>
    ev.min(initial, ev.max(floor, decay(v)))
  }
}

def T[V](initial: V)
        (implicit ev: Numeric[V]): V = {
  T(initial, ev.zero, (t:V)=>ev.minus(t, ev.one))
}
```

Given such a function, the implementation of a `timer` is direct. With `timer`, a `limitedMemory` function can be defined, computing `value` until `timeout` has expired and `expValue` thereafter, effectively realising a memory limited in time.

```
def timer[V](length: V)
            (implicit ev: Numeric[V]) = T[V](length)

def limitedMemory[V,T](value: V, expValue: V, timeout: T)
                      (implicit ev: Numeric[T]) = {
  val t = timer[T](timeout)
  (mux(ev.gt(t, ev.zero)){value}{expValue}, t)
}
```

4) Sparse-choice: `S` can be used to create partitions and for selecting sparse subsets of devices in space. Essentially, it realises a local leader election, where `grain` is the mean distance between two leaders and `metric` represents the notion of distance.

```
def S(grain: Double,
      metric: Double): Boolean =
  breakUsingUids(randomUid, grain, metric)
```

The implementation uses `randomUid` to generate a field of unique identifiers:

```
def randomUid: (Double,ID) = rep((Math.random()), mid()) {
  v => (v._1, mid())
}
```

which is in turn exploited to break the network symmetry:

```
def breakUsingUids(uid: (Double,ID),
                   grain: Double,
                   metric: => Double): Boolean =
  uid == rep(uid) { lead:(Double,ID) =>
    val acc = (_:Double)+metric
    distanceCompetition(G[Double](uid==lead, 0, acc, metric),
                        lead, uid, grain, metric)
  }
```

by means of a competition between devices for leadership:

```
def distanceCompetition(d: Double,
                        lead: (Double,ID),
                        uid: (Double,ID),
                        grain: Double,
                        metric: => Double) = {
  val inf:(Double,ID) = (Double.PositiveInfinity, uid._2)
  mux(d > grain){ uid }{
    mux(d >= (0.5*grain)){ inf }{
      minHood {
        mux(nbr{d}+metric >= 0.5*grain){nbr{inf}}{nbr{lead}}
      }
    }
  }
}
```

5) Restriction in space: We have already encountered the operator for doing domain restriction, which is `branch`. With it, a number of interesting computations can be achieved:

```
// Compute distance from 'src', avoiding obstacles
def distAvoidObstacles(src: Boolean, obs: Boolean): Double =
  branch(obs){ Double.PositiveInfinity }{ distanceTo(src) }

// Perform a broadcast within a particular 'region'
def bcastRegion[V](region: Boolean, src: Boolean, v: V)
                  (implicit ev: OrderingFoldable[V]): Option[V] =
  branch[Option[V]](region){
    Some[V](broadcast(src, v))
  }{ None }

// Measure the size of connected components of a region
def groupSize(region: Boolean): Double =
  branch(region){ summarize(S(1,0),_+_,1,0) }{ Double.NaN }

// Remember whether an event has recently occurred
def recentEvent(event: Boolean, timeout: Int): Boolean =
  branch(event){ true } { timer(timeout)>0 }
```

IV. Alchemist as simulation platform

Testing, debugging, and performance assessment prior to actual deployment are key components of a good software engineering process, and aggregate programming makes no exception. However, since the primary target for this emerging paradigm are distributed and situated systems, conventional testing and debugging tools are hardly enough, in particular falling short at capturing the interaction among devices and between them and the underlying environment. In this situation, simulation emerges as a valuable tool for all the phases of software development: early testing and debugging, integration testing, and performance assessment. Of course, simulation cannot entirely capture the complexity of the real system (much like the classic unit testing cannot test every possible situation in classic application development), nevertheless the desiderata is to be as close as possible to a real situation. There are two relevant dimensions in this regard: first, the simulated environment must capture the most relevant aspect of the distributed system under design; second, the code that the simulator executes must resemble as closely as possible the production code. Considering both dimensions, we picked Alchemist [6].

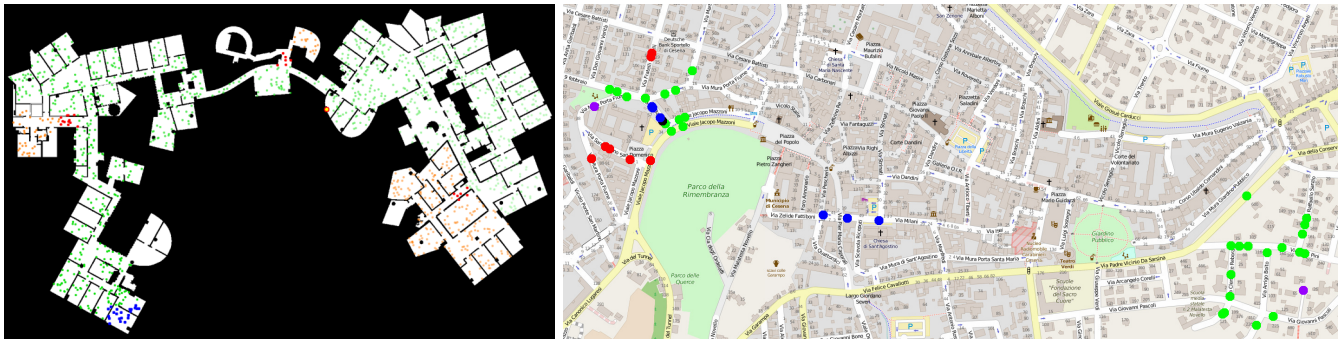
Alchemist is an event-driven simulator, mostly written in Java, tailored to the simulation of pervasive systems with a focus on performance. The model, albeit originally inspired by chemistry, supports complex environments, several flavors of node mobility, and advanced network models. Particularly interesting for real world applications is the possibility of exploiting map data from OpenStreetMap, navigating nodes along existing GPS traces as well as along roads (distinguishing among a handful of vehicles types). Also useful is the support for converting indoor images to Alchemist environments with physical obstacles. Figure 1 shows typical instances of environments frequently simulated with Alchemist.

The Alchemist computational model is generic (it is rather a meta-model), and requires a so called "incarnation" to be developed in order to actually execute simulations. An incarnation is a mapping between the Alchemist meta-model concepts and the concrete entities that the user is interested in simulating. Alchemist, at the time of writing, ships two incarnations, one of them tailored to aggregate programming supporting the simulation of (possibly mobile) Protelis [14] programmed devices.

A. Interfacing Alchemist and Scala

Our goal in interfacing `scafi` and Alchemist was to be able to feed it with the production Scala code, injecting it directly into the simulated environment. A few factors made such integration quite straightforward:

1) Scala and Java are both hosted in the JVM and feature full, bidirectional interoperability;
2) the `scafi` architecture neatly separates its core from the actor-based network backend, this design was key in our pursue to sharing interpreter and Scala code between the actor platform and the simulator;



(a) (b)

Fig. 1: Typical multi-agent scenarios supported by Alchemist: a very dense and intricate indoor environment (Figure 1a), and an urban environment (Figure 1b).

Alchemist meta-model	**Protelis incarnation**	**`scafi` incarnation**
Environment: container of nodes and network model	-	-
Network model	-	-
Node: container of reactions and molecules	Device: container of events and environment variables	Device: container of events and environment variables
Reaction: set of conditions that, if matched, triggers a set of actions with some time distribution	Event	Event
Condition	-	-
Action: any change of the environment	Any Alchemist action, or the execution of a computation round of a Protelis virtual machine, or the dispatch of results to neighbours	Any Alchemist action, or the execution of a `scafi` computation round, included message dispatching
Time distribution	-	-
Molecule, associated with a concentration	Environment variable, associated with a value	Environment variable, associated with a value
Concentration	Value: any Java object	Value: any Scala object

Fig. 2: Mapping between the Alchemist meta-model entities and the Protelis and `scafi` incarnation concepts. Omitted cells indicate that the Alchemist concept is inherited as-is, with no change. Similarities between the incarnations hosting the two aggregate languages are immediately clear.

3) the Protelis incarnation in Alchemist provided an architectural template, that led to a quick identification of the proper conceptual mapping to operate when building the incarnation.

Figure 2 summarizes the mapping effort between Alchemist entities and `scafi` ones, also comparing with the existing Protelis incarnation. As the reader may expect, several entities share the same meaning between the two aggregate programming languages, and as such we believe that an incarnation skeleton for any aggregate programming language could be crafted in Alchemist, further easing the integration of any JVM-hosted aggregate programming language interpreter. In particular, note that sensing of information is supported by means of environment variables in Alchemist, whose evolution over time is controlled when configuring an Alchemist simulation, and which binds to the behaviour of construct `sense` in `scafi`.

V. Case Study

Our goal in this work is to demonstrate the feasibility of using `scafi` in conjunction with Alchemist to realise complex simulations. As such, we do not aim at presenting novel scenarios, but rather we do explain how to recreate some of the results available in literature, validating the proposed framework and showing how convenient is to express complex collective adaptive systems with aggregate computing, `scafi`, and the available APIs. In particular, we focus on two examples that have already been implemented in Protelis and simulated in Alchemist:

1) an example application of smart vehicle counting, which showcases the usage of higher order functions and the possibility of tuning the computational round execution

frequency with Alchemist. The proposed code is derived from [15];
2) an urban crowd tracking scenario featuring the combined usage of all fundamental self-stabilising building blocks exposed in Section III-D [1].

The goal of these simulations, performed prior to actual deployment, would be to evaluate whether the propose collective adaptive system would provide an effective and efficient service for the application at hand.

A. Vehicle counting

Consider a highway, where a sensor is deployed to monitor the number of vehicles passing nearby. We suppose that the vehicles are equipped with electronic components that make them connected with all the other compatible devices within a certain range. We do not consider connectivity issues (which are beyond the scope of this work), and make instead the assumption that every device (vehicle or sensor) is able to communicate with every neighbour within a certain distance. We suppose that all devices are running the following `scafi` "virtual machine", which is able to import any injected procedure from the `snsInjectedFun` sensor:

```
// Dynamically import any computation
def snsInjectedFun: ()=>Double = sense("injectedFun")
// True where the data should get aggregated
def snsInjectionPoint: Boolean = sense("injectionPoint")
// 1 for patrons, 0 otherwise
def snsShouldCount: Double = mux(sense("patron")){1}{0}
// Detection range in meters
def snsRange: Double = 30

def virtualMachine(): Double = {
  deploy(snsRange,snsInjectionPoint,snsInjectedFun,()=>0.0)
}

def deploy[T](range:Double, source:Boolean,
              g: ()=>T, noOp: ()=>T)
             (implicit ev: OrderingFoldable[T]): T = {
  val f: ()=>T = branch(distanceTo(source) < range) {
    G(source, g, identity[()=>T], nbrRange())
  }{ noOp }
  f()
}
```

To monitor the number of nearby vehicles, the sensor device injects the following function, which relies on the `C` building block to realise a convergent sum:

```
def countPatrons() = {
  C[Double](potential = distanceTo(snsInjectionPoint),
      acc = _+_, local = snsShouldCount, Null = 0.0)
}
```

In this scenario, the sensor injects such function after two seconds of simulations, due to a congested block of traffic coming, and turns it off after 10 seconds (by injecting a function returning 0). We executed the same experiment multiple times, varying the round frequency. Figure 3 summarises the results. As expected, values get much closer to reality when the sampling frequency is very high. Moreover, the algorithm exploited to implement `C` is based on a spanning tree, that by its nature is sensible to changes in the network topology: these

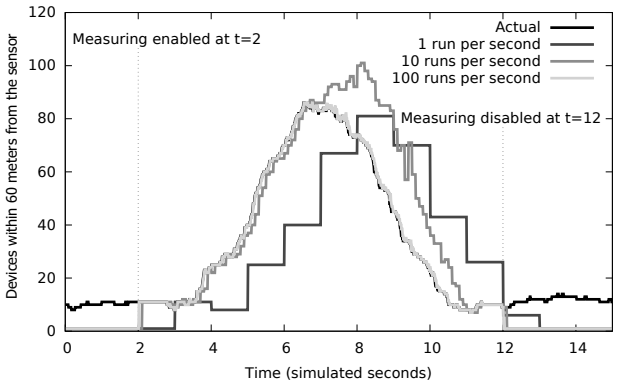


Fig. 3: Number of vehicles counted using the convergent, distributed sum based on `C`. As expected, higher frequencies lead to more precise measurements.

are responsible of the peak in the device count observed with devices running at 1Hz.

B. Crowd tracking in an urban scenario

As second example, consider an IoT environment that provides services for crowd safety at a mass public event, such as a marathon. Such events pose challenging safety issues, because the movement of people in crowded and constrained environments often creates emergent zones of dangerous overcrowding where any small incident can create a panic or stampede that injures or kills people. We simulate here a possible crowd safety service running in an IoT urban environment.

We define three possible crowding levels;

```
val (high,low,none) = (2,1,0) // crowd level
```

a function locally estimating crowd density;

```
def unionHoodPlus[A](expr: => A): List[A] =
  foldhoodPlus(List[A]())(_++_){ List[A](expr) }

def densityEst(p: Double, range: Double): Double = {
  val nearby = unionHoodPlus(
    mux (nbrRange < range) { nbr(List(mid())) } { List() }
  )
  nearby.size / p / (Math.PI * Math.pow(range,2))
}
```

a function mapping each local area to a danger level, depending on the average density sensed locally;

```
def managementRegions(grain: Double,
     metric: => Double): Boolean = S(gran,metric)

def dangerousDensity(p: Double, r: Double) = {
  val mr = managementRegions(r*2, () => { nbrRange })
  val danger = average(mr, densityEst(p, r)) > 2.17 &&
    summarize(mr, (_:Double)+(_:Double), 1 / p, 0) > 300
  mux(danger){ high }{ low }
}
```

and a function yielding true if a situation of danger has remained active for enough time.

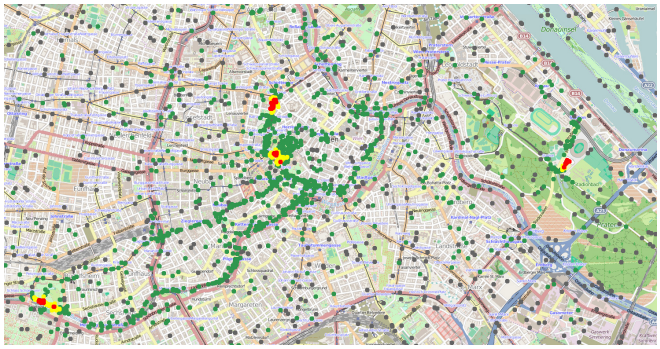


Fig. 4: A snapshot, taken from [1], of Alchemist executing the crowd warning application. Grey dots are stationary devices, not participating the system. Green dots are users in safe areas, red dots users in dangerous areas, and yellow dots are users who are being warned.

```
def recentTrue(state: Boolean, memTime: Double): Boolean = {
  branch(state) {
    true
  }{
    limitedMemory[Boolean,Double](started, false, memTime)._1
  }
}

def crowdTracking(p: Double, r: Double, t: Double) = {
  val crowdRgn = recentTrue(densityEst(p, r)>1.08, t)
  branch(crowdRgn){ dangerousDensity(p, r) }{ none }
}
```

With all those ingredients, we warn the users of those devices located near areas which have remained crowded for a long time.

```
def crowdWarning(p: Double, r: Double,
                 warn: Double, t: Double): Boolean = {
  distanceTo(crowdTracking(p,r,t) == high) < warn
}
```

For the sake of simplicity, the numbers we used for the estimates had been directly written in code. They could get extracted and substituted by parameters, the values proposed are derived from literature [16]. The actual simulation is composed of 1000 stationary devices embedded into the environment plus 1479 mobile personal devices, each following a smartphone position trace collected at the 2013 Vienna marathon [17], [18]. Figure 4 shows a sample screenshot of the system deployed.

VI. Related Work

A. Computational fields and aggregate programming

A wide range of existing approaches to aggregate programming have been proposed, including such diverse approaches as abstract graph processing (e.g., [19]), declarative logic (e.g., [20]), map-reduce (e.g., [21]), streaming databases (e.g., [22]), and knowledge-based ensembles (e.g., [23])—for a detailed review, see [24]. Most of them, however, have been too specialized for particular assumptions or applications to be able to address the aggregate programming challenge at its full complexity in a wide range of different environments.

A unifying model based on *computational fields* has been identified as a generalization of a wide range of existing approaches, surveyed in [24]. Formalized as the computational field calculus [3], this universal language provides a theoretical foundation on which real aggregate programming platforms can be built: both Protelis [14] and `scafi` are practical instances of such calculus.

B. Aggregate programming and multi-agent systems

Aggregate programming targets collective behaviour of systems, which in the MAS literature have typically been addressed in various ways. On the one hand, we have approaches facing the design of relatively small systems of deliberative/cognitive agents: coordination mechanisms and tools (e.g. via artifacts [25] or protocols [26]), social/organisational norms [27], commitments [28], and so on. They provide declarative constraints to agent interaction (abstracting away from the step-by-step operational behaviour of the aggregate), and strongly rely or deal with autonomy of agents, assuming agents have inner mechanisms to dynamically adapt their behaviour to the specific contingency, up to the point of deviating from a previously agreed cooperative behaviour.

On the other hand, collective behaviour of large-scale MASs is mostly studied in terms of swarm-intelligence techniques, assuming that agents are reactive and perform repetitive tasks, with the problem of designing local agent behaviours that can end up in the desired global tasks [29].

Aggregate programming somewhat sits in between these two apparently unreconcilable views of a self-adaptive MASs: it aims at devising a methodology by which the collective behaviour of large-scale ensembles of autonomous, deliberative agents, can be designed so as to manifest inherent properties of self-adaptation, resiliency, and openness.

C. Simulation of aggregates

We claimed in Section IV that simulation is a key step in the development of aggregate programming applications. There are many kinds of simulation tools available: they either provide programming/specification languages devoted to ease construction of the simulation process, especially targeting computing and social simulation (e.g. as in the case of multi-agent based simulation [30], [31], [32], [33], [34]), or they stick to quite foundational computing languages to better tackle performance, mostly used in biology-oriented applications [35], [36], [37], [38].

Alchemist is a discrete-event simulator (DES), since it combines a continuous time base with the description of system dynamics by distinguished state changes [39]. The class of DES more related to our approach are those commonly used to simulate biological-like systems, by which in fact Alchemist was originally inspired. A recent overview of them is available in [38], which takes into account: DEVS [40], Petri Nets [36], State Charts [41], and stochastic π-calculus [35].

VII. Conclusions and future works

In this paper, we introduced `scafi`, a Scala based API and DSL for aggregate programming, equipped with an actor based platform and integrated with the Alchemist simulator. We described the main features of the API/DSL, and demonstrated how `scafi` can be used to realise reusable building blocks that ease the creation of aggregate programs. We presented the integration between `scafi` and Alchemist, a discrete-event simulator targeting pervasive systems, detailing the mapping between the Alchemist meta-model and the concrete `scafi` entities. We were able to push the integration to the point that there exists no difference between the production code and the code required to perform a simulation. We argue that such a deep level of integration will improve the engineering practices when it comes to leveraging aggregate programming for building actual systems, allowing for debugging and testing on a centralized testing platform prior to deployment. We validated the approach by translating complex examples found in literature in `scafi`, using Alchemist as simulation platform.

Further development of this research includes a refinement of the current `scafi` architecture and of its simulator integration. While Alchemist is already publicly available, `scafi` is set to be ready for the general public shortly, as well as the Alchemist incarnation supporting the execution of `scafi` programs in a controlled environment. We expect `scafi` and the related tool chain to boost the adoption of aggregate programming as a new paradigm for the engineering of complex, pervasive systems, such as the typical Internet of Things applications.

References

[1] J. Beal, D. Pianini, and M. Viroli, "Aggregate programming for the Internet of Things," *IEEE Computer*, 2015.

[2] J. Beal and M. Viroli, "Space–time programming," *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 373, no. 2046, 2015. doi: 10.1098/rsta.2014.0220. [Online]. Available: http://rsta.royalsocietypublishing.org/content/373/2046/20140220

[3] F. Damiani, M. Viroli, and J. Beal, "A type-sound calculus of computational fields," *Science of Computer Programming*, vol. 117, pp. 17 – 44, 2016. doi: http://dx.doi.org/10.1016/j.scico.2015.11.005. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167642315003573

[4] M. Viroli, J. Beal, F. Damiani, and D. Pianini, "Efficient engineering of complex self-organising systems by self-stabilising fields," in *IEEE Self-Adaptive and Self-Organizing Systems 2015*. IEEE, Sept 2015. doi: 10.1109/SASO.2015.16 pp. 81–90.

[5] M. Viroli, D. Pianini, A. Ricci, P. Brunetti, and A. Croatti, "Multi-agent systems meet aggregate programming: Towards a notion of aggregate plan," in *PRIMA 2015: Principles and Practice of Multi-Agent Systems*, ser. Lecture Notes in Computer Science, Q. Chen, P. Torroni, S. Villata, J. Hsu, and A. Omicini, Eds. Springer International Publishing, 2015, vol. 9387, pp. 49–64. ISBN 978-3-319-25523-1. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-25524-8_4

[6] D. Pianini, S. Montagna, and M. Viroli, "Chemical-oriented simulation of computational systems with Alchemist," *Journal of Simulation*, 2013. doi: 10.1057/jos.2012.27. [Online]. Available: http://www.palgrave-journals.com/jos/journal/vaop/full/jos201227a.html

[7] R. Casadei and M. Viroli, "Towards aggregate programming in scala," in *First Workshop on Programming Models and Languages for Distributed Computing*. ACM, 2016, p. 5.

[8] M. Mamei and F. Zambonelli, "Programming pervasive and mobile computing applications: The TOTA approach," *ACM Trans. on Software Engineering Methodologies*, vol. 18, no. 4, pp. 1–56, 2009. doi: http://doi.acm.org/10.1145/1538942.1538945

[9] M. Viroli, D. Pianini, S. Montagna, and G. Stevenson, "Pervasive ecosystems: a coordination model based on semantic chemistry," in *27th Annual ACM Symposium on Applied Computing (SAC 2012)*, S. Ossowski, P. Lecca, C.-C. Hung, and J. Hong, Eds. Riva del Garda, TN, Italy: ACM, 26-30 March 2012. ISBN 978-1-4503-0857-1 pp. 295–302.

[10] J. L. Fernandez-Marquez, G. D. M. Serugendo, S. Montagna, M. Viroli, and J. L. Arcos, "Description and composition of bio-inspired design patterns: a complete overview," *Natural Computing*, vol. 12, no. 1, pp. 43–67, 2013. doi: 10.1007/s11047-012-9324-y

[11] M. Odersky, P. Altherr, V. Cremet, B. Emir, S. Maneth, S. Micheloud, N. Mihaylov, M. Schinz, E. Stenman, and M. Zenger, "An overview of the scala programming language," Tech. Rep., 2004.

[12] J. Beal and J. Bachrach, "Infrastructure for engineered emergence in sensor/actuator networks," *IEEE Intelligent Systems*, vol. 21, pp. 10–19, March/April 2006.

[13] M. Viroli and F. Damiani, "A calculus of self-stabilising computational fields," in *Coordination Languages and Models*, ser. LNCS, eva Kühn and R. Pugliese, Eds. Springer-Verlag, Jun. 2014, vol. 8459, pp. 163–178, proceedings of the 16th Conference on Coordination Models and Languages (Coordination 2014), Berlin (Germany), 3-5 June. Best Paper of Discotec 2014 Federated conference.

[14] D. Pianini, M. Viroli, and J. Beal, "Protelis: Practical aggregate programming," in *Proceedings of ACM SAC 2015*. Salamanca, Spain: ACM, 2015, pp. 1846–1853.

[15] F. Damiani, M. Viroli, D. Pianini, and J. Beal, "Code mobility meets self-organisation: A higher-order calculus of computational fields," ser. Lecture Notes in Computer Science. Springer International Publishing, 2015, vol. 9039, pp. 113–128. ISBN 978-3-319-19194-2. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-19195-9_8

[16] J. Fruin, *Pedestrian and Planning Design*. Metropolitan Association of Urban Designers and Environmental Planners, 1971.

[17] B. Anzengruber, D. Pianini, J. Nieminen, and A. Ferscha, "Predicting social density in mass events to prevent crowd disasters," in *Social Informatics*, ser. Lecture Notes in Computer Science, A. Jatowt, E.-P. Lim, Y. Ding, A. Miura, T. Tezuka, G. Dias, K. Tanaka, A. Flanagin, and B. Dai, Eds. Springer International Publishing, 2013, vol. 8238, pp. 206–215. ISBN 978-3-319-03259-7. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-03260-3_18

[18] D. Pianini, M. Viroli, F. Zambonelli, and A. Ferscha, "HPC from a self-organisation perspective: The case of crowd steering at the urban scale," in *High Performance Computing Simulation (HPCS), 2014 International Conference on*, July 2014. doi: 10.1109/HPCSim.2014.6903721 pp. 460–467.

[19] R. Gummadi, O. Gnawali, and R. Govindan, "Macro-programming wireless sensor networks using kairos." in *Distributed Computing in Sensor Systems (DCOSS)*, 2005, pp. 126–140.

[20] M. P. Ashley-Rollman, S. C. Goldstein, P. Lee, T. C. Mowry, and P. Pillai, "Meld: A declarative approach to programming ensembles," in *IEEE International Conference on Intelligent Robots and Systems (IROS '07)*, 2007, pp. 2794–2800.

[21] J. Dean and S. Ghemawat, "MapReduce: simplified data processing on large clusters," *Communications of the ACM*, vol. 51, no. 1, pp. 107–113, 2008.

[22] S. R. Madden, R. Szewczyk, M. J. Franklin, and D. Culler, "Supporting aggregate queries over ad-hoc wireless sensor networks," in *Workshop on Mobile Computing and Systems Applications*, 2002.

[23] R. D. Nicola, M. Loreti, R. Pugliese, and F. Tiezzi, "A formal approach to autonomic systems programming: The SCEL language," *ACM Trans. Auton. Adapt. Syst.*, vol. 9, no. 2, pp. 7:1–7:29, Jul. 2014. doi: 10.1145/2619998. [Online]. Available: http://doi.acm.org/10.1145/2619998

[24] J. Beal, S. Dulman, K. Usbeck, M. Viroli, and N. Correll, "Organizing the aggregate: Languages for spatial computing," in *Formal and Practical Aspects of Domain-Specific Languages: Recent Developments*, M. Mernik, Ed. IGI Global, 2013, ch. 16, pp. 436–501. ISBN 978-1-4666-2092-6 A longer version available at: `http://arxiv.org/abs/1202.5509`.

[25] M. Viroli, A. Omicini, and A. Ricci, "Engineering MAS environment with artifacts," in *2nd International Workshop "Environments for Multi-*

Agent Systems" (E4MAS 2005), D. Weyns, H. V. D. Parunak, and F. Michel, Eds., AAMAS 2005, Utrecht, The Netherlands, 26 Jul. 2005.

[26] A. K. Kalia and M. P. Singh, "Muon: designing multiagent communication protocols from interaction scenarios," *Autonomous Agents and Multi-Agent Systems*, vol. 29, no. 4, pp. 621–657, 2015. doi: 10.1007/s10458-014-9264-2. [Online]. Available: http://dx.doi.org/10.1007/s10458-014-9264-2

[27] A. Artikis, M. J. Sergot, and J. V. Pitt, "Specifying norm-governed computational societies," *ACM Trans. Comput. Log.*, vol. 10, no. 1, 2009. doi: 10.1145/1459010.1459011. [Online]. Available: http://doi.acm.org/10.1145/1459010.1459011

[28] A. U. Mallya and M. P. Singh, "An algebra for commitment protocols," *Autonomous Agents and Multi-Agent Systems*, vol. 14, no. 2, pp. 143–163, 2007. doi: 10.1007/s10458-006-7232-1. [Online]. Available: http://dx.doi.org/10.1007/s10458-006-7232-1

[29] H. V. D. Parunak, S. Brueckner, R. S. Matthews, and J. A. Sauter, "Pheromone learning for self-organizing agents," *IEEE Transactions on Systems, Man, and Cybernetics, Part A*, vol. 35, no. 3, pp. 316–326, 2005. doi: 10.1109/TSMCA.2005.846408. [Online]. Available: http://dx.doi.org/10.1109/TSMCA.2005.846408

[30] S. Bandini, S. Manzoni, and G. Vizzari, "Agent based modeling and simulation: An informatics perspective," *Journal of Artificial Societies and Social Simulation*, vol. 12, p. 4, 2009. [Online]. Available: http://EconPapers.repec.org/RePEc:jas:jasssj:2009-69-1

[31] M. Schumacher, L. Grangier, and R. Jurca, "Governing environments for agent-based traffic simulations," in *Proceedings of the 5th international Central and Eastern European conference on Multi-Agent Systems and Applications V*, ser. CEEMAS '07. Berlin, Heidelberg: Springer-Verlag, 2007. doi: 10.1007/978-3-540-75254-7_17. ISBN 978-3-540-75253-0 pp. 163–172. [Online]. Available: http://dx.doi.org/10.1007/978-3-540-75254-7_17

[32] S. Bandini, S. Manzoni, and G. Vizzari, "Crowd Behavior Modeling: From Cellular Automata to Multi-Agent Systems," in *Multi-Agent Systems: Simulation and Applications*, ser. Computational Analysis, Synthesis, and Design of Dynamic Systems, A. M. Uhrmacher and D. Weyns, Eds. CRC Press, Jun. 2009, ch. 13, pp. 389–418. ISBN 978-1-4200-7023-1. [Online]. Available: http://crcpress.com/product/isbn/9781420070231

[33] M. J. North, T. R. Howe, N. T. Collier, and J. R. Vos, "A declarative model assembly infrastructure for verification and validation," in *Advancing Social Simulation: The First World Congress*, S. Takahashi, D. Sallach, and J. Rouchier, Eds. Springer Japan, 2007, pp. 129–140.

[34] E. Sklar, "Netlogo, a multi-agent simulation environment," *Artificial life*, vol. 13, no. 3, pp. 303–311, 2007.

[35] C. Priami, "Stochastic pi-calculus," *The Computer Journal*, vol. 38, no. 7, pp. 578–589, 1995.

[36] T. Murata, "Petri nets: Properties, analysis and applications," *Proceedings of the IEEE*, vol. 77, no. 4, pp. 541–580, Apr. 1989. doi: 10.1109/5.24143. [Online]. Available: http://dx.doi.org/10.1109/5.24143

[37] A. M. Uhrmacher and C. Priami, "Discrete event systems specification in systems biology - a discussion of stochastic pi calculus and devs," in *Proceedings of the 37th conference on Winter simulation*, ser. WSC '05. Winter Simulation Conference, 2005. ISBN 0-7803-9519-0 pp. 317–326. [Online]. Available: http://dl.acm.org/citation.cfm?id=1162708.1162767

[38] R. Ewald, C. Maus, A. Rolfs, and A. M. Uhrmacher, "Discrete event modeling and simulation in systems biology," *Journal of Simulation*, vol. 1, no. 2, pp. 81–96, 2007. [Online]. Available: http://dx.doi.org/10.1057/palgrave.jos.4250018

[39] B. P. Zeigler, *Theory of Modeling and Simulation*. John Wiley, 1976.

[40] B. Zeigler, *Multifaceted modelling and Discrete Event Simulation*. Academic Press, 1984.

[41] D. Harel, "Statecharts: A visual formalism for complex systems," *Sci. Comput. Program.*, vol. 8, no. 3, pp. 231–274, Jun. 1987. doi: 10.1016/0167-6423(87)90035-9. [Online]. Available: http://dx.doi.org/10.1016/0167-6423(87)90035-9

4th International Workshop on Smart Energy Networks & Multi-Agent Systems

THE emerging smart infrastructure in energy networks represents a major paradigm shift in resource allocation management with the aim to extend the centralised supply management model, towards a decentralised supply-and-demand management that is expected to enable more efficient, reliable and environment-friendly utilisation of primary energy resources.

Together with this vision, there are new and complex tasks to manage, in order to ensure safe, cost-reducing and reliable energy network operations. This includes the integration of various renewable energy systems, like the photovoltaic or the wind energy, which are able to reduce the greenhouse gas emissions but that are working under greater uncertainty; as well as the interaction of transport and storage systems for energy that are envisioned through techniques like 'Power to Gas' and fuel cells, which are using the electrical and the gas transportation network.

Further tasks can be found in the fact that the market participants (e.g. simply households) are becoming more autonomous and intelligent through technologies like smart metering, which requires a coordinated demand side management for millions of producers, consumers or, if this applies, prosumers by means negotiations and agreements.

Information and communication technologies are key enablers of the envisioned efficiencies, both on the demand and the supply sides of the smart energy networks, where the agent-paradigm provides an excellent first modelling approach for the distributed characteristic in energy supply systems. On the demand side they aim at supporting end-users in optimising their individual energy consumption, e.g. through the deployment of smart meters providing real-time usage and cost of the energy and the use of demand-response appliances that can be controlled according to the user preferences, energy cost and carbon footprint. On the supply side they aim at optimising the network load and reliability of the energy provision, e.g. through active monitoring and prediction of the energy usage patterns, and proactive control and management of the reliable energy delivery over the networks. It is also envisaged that they will be able to influence the demand through the dynamic adjustments of the energy price in order to influence the end user behaviour and energy usage patterns throughout and across the energy networks for electricity, gas and heat.

Although a significant effort and investment have been already allocated into the development of smart grids, there are still significant research challenges to be addressed before the promised efficiencies can be realised. This includes distributed, collaborative, autonomous and intelligent software solutions for simulation, monitoring, control and optimization of smart energy networks and interactions between them.

Topics

- Experiences of Smart Grid implementations by using MAS
- Applications of Smart Grid technologies
- Management of distributed generation and storage
- Islands Power Systems, Microgrid Applications
- Real time configurations of energy networks
- Distributed planning process for energy networks by using MAS
- Self-configuring or self-healing energy systems
- Load modelling and control with MAS
- Simulations of Smart Energy Networks
- Software Tools for Smart Energy Networks
- Energy Storage
- Electrical Vehicles
- Interactions and exchange between networks for electricity, gas and heat
- Stability in Energy Networks
- Distributed Optimization in Energy Networks

Event Chairs

- **Derksen, Christian,** University Duisburg-Essen, Germany
- **Kowalczyk, Ryszard,** Swinburne University of Technology, Melbourne, Victoria, Australia

Steering Committee

- **Derksen, Christian,** University Duisburg-Essen, Germany
- **Lehnhoff, Sebastian,** OFFIS - Institute for Information Technology, Germany
- **Kowalczyk, Ryszard,** Swinburne University of Technology, Melbourne, Victoria, Australia
- **Nahorski, Zbigniew,** Systems Research Institute - Polish Academy of Science, Poland

Program Committee

- **Guttmann, Christian,** Monash University, Australia
- **Ketter, Wolf**
- **Moench, Lars,** FernUniversität Hagen, Germany
- **Ossowski, Sascha,** University Rey Juan Carlos, Spain
- **Özdemir, Serkan**
- **Redder, Mareike**

- **Renz, Wolfgang,** HAW Hamburg
- **Sonnenschein, Michael,** University of Oldenburg, Germany
- **Sudeikat, Jan,** Hamburg Energie GmbH, Germany

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1507–1515
DOI: 10.15439/2016F392
ACSIS, Vol. 8. ISSN 2300-5963

The EOM: An Adaptive Energy Option, State and Assessment Model for Open Hybrid Energy Systems

Christian Derksen, Rainer Unland
DAWIS - University of Duisburg-Essen
Schützenbahn 70, 45127, Essen, Germany
Email: {christian.derksen, rainer.unland@icb.uni-due.de}@icb.uni-due.de

***Abstract*—The current transformation process of how energy is supplied attracts great interest from many different market players. As a consequence, many proprietary solutions for "smart" energy applications are flooding the market. This turns out to be rather a problem than part of the solution for the systematic development of future energy grids. Additionally, the absence of necessary standards blocks further developments that enable the creation of novel, market-driven and hybrid control solutions. To overcome these problems, we suggest a standardized control approach for hybrid energy systems by means of a so called Energy Option Model (EOM). This unifying model and the therewith developed decision support system provides the necessary technical understanding and the economic assessment options for network-connected energy conversion systems. Thus, it can be used for single on-site systems as well as for aggregated systems that are controlled in centralized or decentralized manner. This paper presents and discusses exemplary use cases for our EOM that illustrate the centralized as well as the decentralized use of our approach within hybrid energy systems. Overall, we believe that the EOM represents the key approach for a further systematic development of an open hybrid energy grid.**

I. Introduction

THE tendency towards decentralized controlled energy conversion systems and the increasing number of IT-enriched smart systems in general leads towards an energy landscape that consists of complex, globally connected and mainly software driven systems. In order to reach climate targets or just to maximize organizational profit, smart markets need to be provided on top of the underlying technical systems with their inherent flexibility. Concurrently, a stable volatile energy production must be guaranteed. However, global goals, such as the stabilization of a distribution network, require a minimum of adaptive interoperability that has to be expressed in one standard. We believe that the developments in smart grids and related areas are now at a point, were it has to be asked, if we want to build control systems that will create new monopolies, caused by proprietary software solutions, or if we want to support an unbundled and open energy supply that, on the one hand, offers the needed intelligent flexibility and, on the other hand, supports further developments over the next decades? Assuming that the latter is the case, it is obvious that software standards are required that prevent our already highly complex energy supply from becoming more complex and possibly uncontrollable with respect to the technical foundations and market regulations [1].

This paper presents the approach and the framework of our system-centered Energy Option Model (EOM), that permits to comprehensively describe the energetic and the economic behavior of any type of energy conversion system. Additionally, based on this unifying concept, the dynamic aggregation of hybrid energy systems is supported so that coalitions of systems, like virtual power plants, prosumers in a distribution network or even the devices in a Smart House can be represented, observed and optimally controlled. As a consequence of this unifying approach, the EOM can be used for several purposes and in different scenarios. By describing energy systems and their abilities in a comprehensive manner, the EOM can first be used as foundation for complex system simulations. Moreover, connected to real on-site systems, it permits the construction of centralized, partially decentralized or completely decentralized control approaches within Smart Grid and Smart Market scenarios. To demonstrate this capability, we demonstrate here two simple application cases by means of single and aggregated electrical vehicles. Further, we will discuss the consequences for our energy agent approach [2]. For this the paper is structured as follows: The next section provides some background information and motivates our work. Section III will outline the energy agent approach, while IV introduces the basic structure of the EOM. In section V the two mentioned use cases of the EOM will be shown. Subsequently, our solution will be discussed and compared to other approaches, known from literature. The paper end with a conclusion in section VI.

II. Backgroung & Motivation

The ongoing discussions of whether Smart Grid devices or even more complex system aggregations are controlled in a centralized or decentralized manner and under which

market rules they operate, is still an open scientific, organizational and social question. Here, the requirements or goals for such control approaches differ and may address stability issues in distribution networks, the efficient usage of energy resources, a cost and revenue optimization with respect to the actual part of the energy market or simply aspects of a customer's comfort. Based on these different requirements, diverse control solutions were developed. Accordingly, there is a lack of standardized control approaches that provide both, first an investment protection for developer and customer of smart devices and secondly an adaptive flexibility with respect to an organizational affiliation and thus to the actual control solution used with such devices. We will discuss these two aspects in the following.

In the context of control approaches, literature of the recent years presents a significant number of publications, describing the successful applications of agents and Multi-Agent systems (MAS) in specific Smart Grid scenarios, as for example in virtual power plants [3], in Demand-Side-Management systems [4] or within price-based, indirect controlled approaches that are known as Demand Response [5]. Beyond, a few decentralized control approaches were introduced, as for example with [6] and other. Since most of these developments were motivated by the problems that occurred with the increasing number of regenerative and thus volatile energy production systems, they basically focus on aspects around electrical grids and markets. However, such a one-sided focus neglects the flexibility-potentials that could possibly be utilized in order to close the storage gap electrical networks. Research has already introduced a couple of approaches here that are characterized as so called Power-to-X applications [7]. In such applications, electrical excess energy will either be stored, as for example in an electrical vehicles accumulator (Power-to-Vehicle), or it is converted to a different form of energy, such as hydrogen or heat. In this context, [8] provides an overview of different approaches, technologies and strategies that focuses on the management of large-scale schemes of variable renewable electricity. Considerations about such hybrid energy systems, however, are fairly new in a broader range and thus even more far away from any standardization for an open and adaptive control.

The lessons that could be learned with the current state of the art approaches, clearly indicate that any further diversification of control approaches has to be prevented in order to avoid uncontrollable situations that may lead to a chaotic overall system behavior of our energy supply. Additionally, it can be expected that an increasing diversification of self-containing control approaches will also increase the customer's uncertainty about the expectable functionalities, its benefits, the customers privacy and will thus probably prevent the needed investment decisions that would bring the "smart" market into motion [9]. This similarly applies also for the developer and provider of such control solutions, since they have to live with the uncertainty of non-existing or suddenly appearing policies that may destroy their control approach or even their business models.

In contrast to an adaptive control that is know from automation [10], we are using the notion of an adaptive control as a synonym for an open overall system architecture that dynamically allows to adapt and thus to 'understand' any kind of energy conversion process and its inherent flexibility. Here we go by the claim that individual systems, or energy conversion processes respectively, should be arbitrarily integrated into any larger overall system. The reasons for this requirement are versatile. The simplest argument for that is the avoidance of new regional or local monopolies that occur, if proprietary and self-containing on-site solutions are used, as already mentioned in the introduction of this paper. Following the unbundling principle, it is our opinion that a customer should be able to freely select and change its energy supplier, regardless of what specific smart grid device or what actual control approach is used; this freedom must not be prevented by any smart device or software system.

The described bilateral relation between a customer and a supplier, however, does not go far enough, since the number of parties involved in this context is much larger. For example, it is conceivable that the system immanent flexibility of an energy conversion process can also be used by a distribution network operator (DNO) in order to stabilize critical network situations. But this would require that also a DNO is able to utilize a systems flexibility, which is not possible if no unified description of an energy conversion process is used.

In turn, if one sees a unified system description not only as a technical or regulatory requirement, such description would have the potential to be the key enabler for a real sustainable development of a market driven future energy grid. Starting from the scientific point of view, unified system descriptions could be used in order to systematically explore the complexity of large-scale system aggregations and their cross associations, as they can already be found in a single distribution network that is organizationally also connected to different energy suppliers. Using the same base control model for single systems, researchers could get the chance to similarly compute and compare their results and thus their different control approaches. Based on that, the necessary market rules could be derived and thus hopefully help to resolve the current chicken and egg problem in sustainably designing a future energy grid.

Based on that, also the energy market could participate in various ways. Providing an adaptive operation and flexibility model for any energy conversion process involved, could theoretically enable new market mechanisms that would for example allow to buy the needed, task-dependent energy for a single system on the fly. Assuming a suitable accuracy of such model, it could

also be used as a predictive model that provides the information for the needed energy amount. Beyond that, also the change of supplier could be made easier, since the inclusion of a specific system would be based on the same operation and flexibility model.

With above described requirements or even visions, the here introduced EOM was firstly developed as a modular approach, focusing on the abilities and thus also on the flexibility of single energy conversion processes. This modularity concept goes along with our energy agent approach that will shortly be outlined in the next section. Since these energy agents may need to consider multiple systems simultaneously, the EOM and its framework were designed also to manage system aggregations. These can be understood as 'system of systems' and will be explained in more detail in section IV.

III. ENERGY AGENTS

According to our previous publication [2], an energy agent can be understood as the computing entity that manages the concerns between an on-site system and its stakeholder. For this, we see an energy agent as a mediator software system that for the most part is located in between of the local system controller and the outside, possibly "smart" world. An exception is the case, if the actual system being guided through an agent has no own controller; in this case an energy agent may also take the control tasks.

For the interaction with the local system, energy agents need to incorporate a multitude of possible connectivity protocols, as for example serial protocols like *Modbus RTU* or more sophisticated protocols like *OPC UA*, *IEC 61850* and other [11], [12]. By using this connectivity, the energy agent should be enabled to receive or set all relevant information that help to comprehensively monitor or guide the actual system. In the context of the EOM, those information are part of the so called 'system variables' and include measurement values, system set points that can be controlled by the agent and system set points that were configured by an end user, as for example the desired room temperature for a heating system. To cooperate with further agents and software systems, as for example with a web service that provides weather forecasts, the energy agent should be able to use a network connection that allows these interactions. Fig. 1 below visualizes the described system environment and shows the full list of system variable types that are used within the EOM, described in the next section.

With the static information, system specific data models are meant that describe the system abilities in a mathematical, empirical or theoretical manner. Examples can be found with a consumption map of a combustion engine or with a turbine or compressor map of rotating equipment.

The remit of an energy agent depends on the actual scenario definition for which it will be designed. Here we introduced the notion of 'integration level' (IL) that describes the level of sophistication with respect to a scenario and thus also to the inherent complexity and the abilities of an energy agent [13]. In a rough classification, we differentiated between an IL0 that describes the initial construction state of the energy infrastructure for a time, where no decentralized computational entities could be found. Based on that and with increasing and coherent integration level, we assumed a continuous change from centralized controlled system to more and more decentralized controlled systems. Finally, we assumed that a completely decentralized control approach that was described with IL5 is only a theoretical consideration that can't be realized.

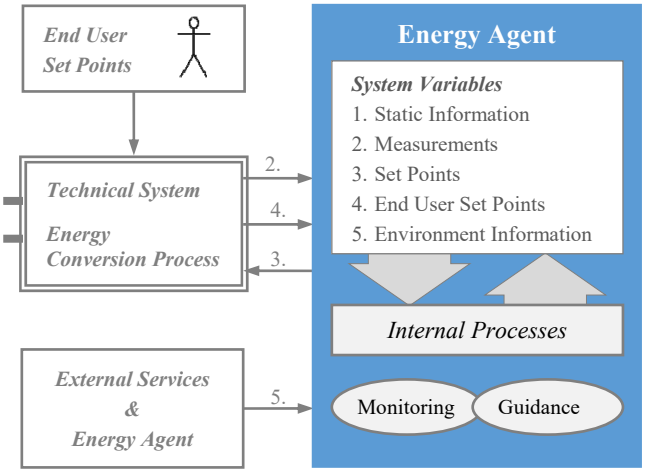


Fig. 1 Energy Agent Connectivity and its System Variables

With the coherent set of integration levels, we also wanted to reflect the fact that local software components as the energy agents will be subject to changing conditions over time. For example, it can be assumed that the regulatory policies will change successively over the next decades. Accordingly, the on-site software has to be further developed and adjusted in order to meet the changing market rules and policies. This, in fact, corresponds to another type of hybrid consideration, since it can be expected that not every system involved will always be updated.

It is our goal to use energy agent in simulations, testbed applications and in real on-site systems across all integration levels. Therefore, we are using the characteristics of software agents that allow to further enrich or exchange specific behaviors. Especially the ability to interact with the underlying technical system and thus to monitor or guide it, represents here the main difference between simulation, testbed or real application. Here it is planned, to either use a simulated or a real input and output behavior for the (virtual) connection to the technical system. However, while in a real system real data can be gathered, in simulations a mechanism is required that allows to simulate the acquisition data that would originally come from the underlying system. Additionally, these data have to dynamically change in case

that an energy agent meets a control decision. That in turn requires a suitable model that allows to describe the dynamic system behavior.

It is obvious that an energy agent requires a suitable model and thus an understanding about the associated technical system for both, for simulations as well as for real applications. This applies not only if a decentralized control solution is to be used or validated, as for example in case of a price-based Demand Response approach. This already applies for the lowest integration level 0, where basically the physical behavior of one or more energy conversion systems is described. Beyond, using a larger number of such dynamic models will further enable to simulate complex infrastructures, such as distribution or transportation networks and their dynamics.

Independent of the here introduced idea of an energy agent, we would like to state that an open description of the capabilities and hence the flexibilities of individual energy conversion systems is essential for a sustainable development of our energy infrastructure. Our approach is presented in the next section.

IV. THE ENERGY OPTION MODEL (EOM)

The fundament for the EOM is a system delimitation that separates the considered energy conversion process (ECP) and its environment. Based on that, a customizable and scalable system definition can individually be modeled. That means that on the one hand small single systems can be described, while on the other hand also bigger systems or multiple simultaneous systems might be modeled at once, in case that the systems effect on the environment (or on the network connections respectively) can reliably be described. Thus, a system definition described by the EOM, can also be the model of a complex plant, even if it consists of several conversion processes in detail.

Based on the first law of thermodynamics [14], that describes the conservation of energy, the EOM allows to capture all network connections and usage types of an ECP. Therefore, the concept of a so called *TechnicalInterface-Configuration* (TIC) was introduced as an anchor for the further modelling that enables to differentiate those cases. Thus the EOM allows to capture different connectivity types (as for electrical vehicles) or to record different execution modes for single systems (as for example the different programs of a washing machine). Each TIC can be further modeled as described in the following.

For the further inventory of an ECP, the EOM allows to capture the actual network connections of the system. For a comprehensive description, the EOM requires information about the used energy carrier with each connection (e.g. electricity, natural gas or heat) and the direction of the energy flow, if relevant. In case that the described system contains an internal energy storage, additionally the storage capacity can be captured with a specific network connection.

To capture the regular operating states and thus the usual system behavior, the EOM allows their modelling with the help of a directed graph *G (V, E)*. Here, the vertices *V* represent the operating states, while the directed edges *E* describe the subsequent order in which the operating states occur. Through a reference to itself, an operating state can be repeated as many times as needed. Each vertex contains further information about the duration and the energy revenues that can be generated in an operating state. Defining the duration of an operating state and repeating it, allows thus a flexible modelling that corresponds to any desired discretization for the behavior of the considered ECP.

The description of an energy flow, in a specific operating state, for a specific network connection and for the chosen duration is a further part of the operating state description that corresponds to one of the above described vertices. The EOM offers various possibilities to describe energy flows. For that purpose, it differentiates between constant, empirical or calculated energy flows that can be used as description for each single network interface. Here, measurements that were statistically evaluated are seen as empirical information. Those data can especially be used in case that an energy flow is not simply constant or if the energy flow can't be calculated by mathematical equations. Thus unsteady or transient operating states can be described by datasets and help to better 'understand' or predict a systems behavior.

For the calculation of energy flows, the framework of the EOM offers the extension of specific classes which then contain the required calculation methods. Based on an operating state and the system variables that were already described in the previous section, any kind of calculation can be executed in order to determine an energy flow. Thus, this adaptive approach enables to use the theoretical knowledge that is especially present in the engineering disciplines.

Up to here, we described the first part of the EOM and its framework, which can be considered as the basic model and thus the fundamental for all further steps. The corresponding data structure of the above described model is also available as an XML scheme [2]. Based on that, the EOM-framework allows the persistence and thus the exchange of ECP descriptions.

The second part of EOM and its framework is aimed to comprehensively generate schedules for the system considered. Comprehensively means here that with the help of the EOM a full description of the system behavior over time can be generated that consist of all time-dependent values of the system variables and all applied energy flows. Furthermore, the converted energy amounts, the energy losses and an individual utility function can be taken into account, which enables an assessment of single system states and thus the assessment of the overall system usage.

To get there, the EOM framework enables to configure the needed evaluation parameters that are first a starting, time-dependent system state and an end time, wherein the evaluation may be terminated. If required, the end time for an evaluation can also be connected to a target system state. This could be for example a fully charged battery of an electric vehicles battery at the end of the evaluation period.

Further information for the time-dependent evaluation of the system can be predictions of specific system variables that are required to calculate energy flows. Such variables can be, for example, temperature or insolation information that are derived from weather forecasts and that are used in order to pre-calculate the heat demand of a house or to predict the energy production of a photovoltaic plant. Here, the framework of the EOM permits again the extension and usage of individual implementations that may provide the needed data by using any type of data source.

For the definition of an individual utility function, the EOM separates between two assessment paths. The first path is designed to enable the assessment of the energy flows and amounts that are transferred over each network interface. Therefore, a corresponding function can be assigned for each energy carrier in combination to a flow direction. Thus, a power feed or consumption can be considered differently. With the second path of the utility function, the EOM enables once more the extension and usage of individual calculation classes. Within such a class, any type of calculation can freely be implemented; its result will finally be added to the utility results of the energy flow assessment. In this way, e.g. depreciation costs for the operation of an ECP can additionally be added or further, more complex relationships.

The settings described above, finalize the preparation of an evaluation process within the EOM framework and an actual evaluation process can be executed. Since the range of goals that have to be considered here may differ, the EOM framework permits once more to define individual so called ‘evaluation strategies’.

Based on the information of the basic model and the evaluations settings, the goal of such a strategy is to produce a system schedule in a predefined data structure. Therefore, the EOM does not restrict the way, nor limits it to a specific algorithm that has to be used. Rather, it leaves the actual approach for the creation of a system schedule deliberately open in order to enable the application of different competitive approaches. That means that a developer is free to decide which algorithm is to be used and how the actual evaluation strategy is designed. This especially enables to reuse already proven and reliable approaches for the generation of system dependent schedules.

Beside the above described open architecture, the EOM framework additionally provides a comprehensive assistance for generating system schedules through an own evaluation strategy. Therefore, a graph based methodology was developed that theoretically can be applied to any kind of ECP. The foundation for this approach is the unique identification of systems states. Since the identification of these states may differ depending on the actual system, the EOM permits to specify the parameters that have to be considered here. By default, the evaluation time, the chosen interface configuration, the operating state, the set points as well as controlled measurements are used in order to specify this identifier. Additionally, also storage loads can be taken into account, if required. For the discretization of unique system states, an increment can be set for each parameter of the state identifier.

Based on those clearly differentiable and discretized unique system states, a graph can be drawn that represents the states and their subsequent states over time. Figure 2 below shows the so called Differences Graph of this evaluation approach.

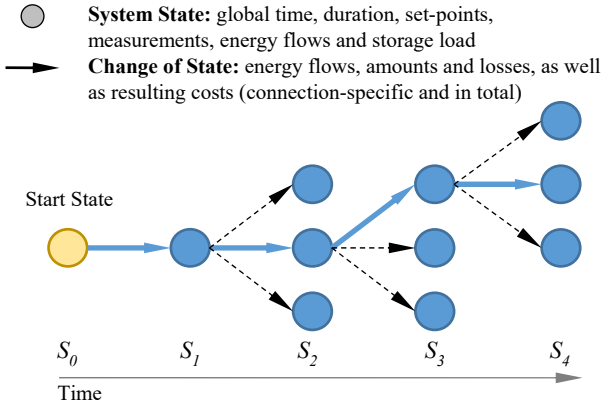


Fig.2: Differences Graph for an evaluation strategy

While the vertices of the graph represent the unique system states, the edges correspond to the actual changes that occur with the transition from one state to another. Thus, a change in a storage level can be determined as well as the changes that occur if an individual utility function is applied. Based on this differences approach, a number of well know graph-based optimization algorithms may be applied (e.g. Dijkstra and many more) in order to optimize the system usage, taking into account certain objectives. Additionally, the EOM framework provides a general algorithm structure that can be used within an evaluation strategy and that proceeds along the graph structure. For reasons of space we skip a detailed explanation here, but we want to point out that an evaluation becomes a sequence of time dependent decision processes that are based on qualified system information, represented by the Differences Graph. Thus, the EOM and its framework may also be considered as a decision support system [15].

The aggregation method for several technical systems that is provided by the EOM is realized in the sense of a ‘system of systems’. Analogously to a single technical system, the aggregator will also be considered as a technical system, summarizing the sub systems in regard to interfaces by energy carrier and by add up energy carrier-dependent

storage capacities. In more complex scenarios, where wider areas are to be considered, an aggregation may additionally require a network calculation, which is also supported by the EOM. Similar to the evaluation strategy for single technical systems, a strategy for aggregated technical systems can individually be designed. In contrast to single systems, decisions must be made for each subordinate system, so that several decision graphs will concurrently be used during an evaluation of aggregated ECP's.

The EOM is available as an end user application and laboratory tool, but can also be used "head-less" within an energy agent. In this case it is the task of an agent to get and provide the needed base information for a single technical system or an aggregation of technical systems (e.g. predictions, cost information and other) and start an appropriate evaluation process, if required.

V. APPLICATION OF THE EOM

To demonstrate the applicability of the EOM, this section describes the modelling and algorithm approach for planning processes of one and more electrical vehicle's (EV) battery. For this, first, the actual system will be described in the next sub-section, while subsequently the used algorithms will be outlined.

A. Single Electrical Vehicle

The chosen EV has a battery with a storage capacity of 24 kWh. According to IEC 62196, it can differently be connected to an electrical network. For the experiments we have chosen a Mode 1 connection, which corresponds to a slow charging from a household-type socket-outlet (230V, 3.5 kW charging or discharging). With this information, the base graph of the operating states was modelled as shown in Figure 3.

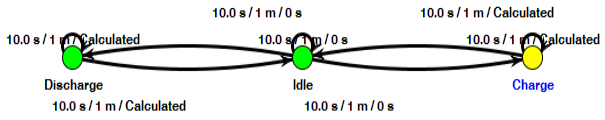


Fig.3: Graph for the operating states of an EV battery

It can be seen that the operation states *Charge* and *Discharge* are modelled in a way that each of them apply for at least 10 seconds. Both times are followed by a minimum and a maximum time in which the system can be kept in the respective operating state. While here the minimum time was defined in order to avoid a too short stay and thus a too quick change of the operating state, the maximum time can only be calculated depending on the current storage load (e.g. a full battery cannot be charged anymore). Further, for the operating state *Idle* no maximum must be defined, since this system neutral state has no time limit in principal.

For the evaluation of the system, we assumed a planning scenario, where the EV has a time frame between 8 p.m. and 6 a.m. for charging. Further, we assumed an initial storage load of 5 kWh, while the car has to be fully charged in the morning. These assumptions correspond to the definition of two system states (an initial and an end system state) with corresponding timestamps. This limits the evaluation period for the developed evaluation strategies. For the assessment of the time dependent system behavior, we used the EOM's cost function approach by means a time variable pricing for the usage of electrical energy.

To create a base for later comparisons, we first developed a simple strategy that's behaves in a regular or Greedy-styled manner. That means, like a non-controlled or scheduled charging process, the EV's battery will directly be charged until it reaches its storage capacity.

Thereon based, another strategy was developed that considers the specified procurement costs for electrical energy over time. Simply by selecting time ranges that have the best prices and put them in an ascending order, the prioritized charging times were determined. Consequently, for the used evaluation strategy, the task of cost optimization results to a fairly simple observation of time and a selection of the desired operating state until the battery reaches it maximum capacity. Figure 4 below shows the used electricity costs and compares the results of the two evaluation strategies.

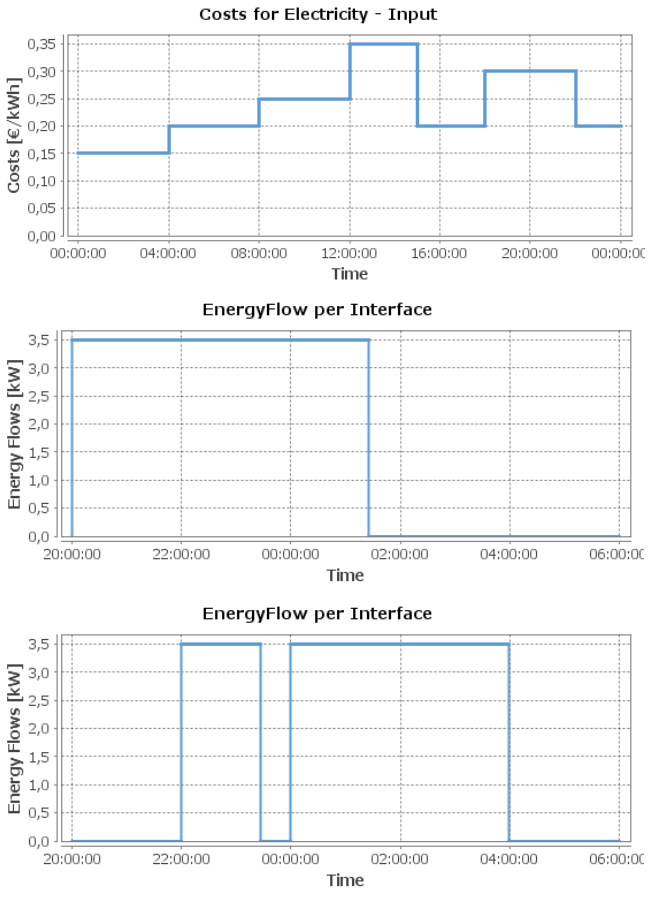


Fig.4: Cost-Optimal Charging of an EV's battery, based on price signals

For the example, a saving of 27% of the procurement costs could already be realized, simply by the described more sophisticated selection process of charging times.

Even the complexity of the used algorithms, or evaluation strategies respectively, were very manageable here, the learning curve for the development of the EOM and its usage as a reasoning model were remarkable. Here we realized that, as first, the development of an evaluation strategy is required that is able to reproduce the regular behavior of an ECP. Only on this basis, the improvements of 'intelligent' algorithms can be found and compared.

Despite the simplicity of the presented example, a very important capability of the EOM is presented here. This is the capability to concurrently apply different evaluation or plan generation strategies in a parallel manner. Since competitive evaluation strategies can be executed concurrently, single energy agents can be empowered to become intelligent agents. Based on the current system state that can be acquired from the system (by means of measured values, set-points, energy flows etc.) and a defined goal, like a cost optimal charging of an EV's battery, the energy agent can execute different strategies, where each results to a different system usage of time. Based on a comparison or assessment of these resulting execution schedules, the energy agent is able to select the best plan and consequently to optimize the systems behavior. With this general approach, the EOM is closely related to the well-known BDI-concept [16], but its focusses on the special needs of the energy domain.

B. Multiple Electrical Vehicles

For the second application case of the EOM, we used the capability to consider several systems within an EOM-aggregation. For this, the EOM supports the dynamic configuration of aggregations by means of a tree organized structure, where the root node represents the overall aggregation and the sub nodes the aggregated systems. Analogously to single systems, the aggregation can be modeled as a system, while the considered sub systems can either be defined as static load curves, as dynamic single systems or as dynamic aggregations again. The base model of the aggregation was defined as shown in Figure 5.

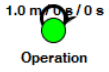


Fig.5: Graph for the operating state of an EV Fleet Aggregation

It can be recognized that the superior system of the aggregation is specified as system with a single operation state *Operation*, where the discrete time step is defined with 1 minute. Further, neither a minimum nor a maximum time is defined.

For the experiments, we used the above described EV model and varied the number of systems between 5, 10, 20, 30, 40, 50 75, 100, 150 and 200 EV's. Since a price optimal charging for the whole EV fleet results to a similar curve in terms of the time course (e.g. 10 EV's, charging by 3,5 kW = 35 kW), we slightly changed the objective of the evaluation strategy. In addition to a possible price optimal charging, an upper bound for the overall energy consumption should not be exceed now.

To archive this goal, the developed strategy first determined the demand of energy for the evaluation period. Further, it again sorted the consumption costs in an ascending manner. Subsequently, and with respect to the maximal available power that was defined by the upper boundary, the needed energy amounts could successively be assigned to the sub systems. Figure 6 below shows the resulting energy consumption for an EV fleet, consisting of 10 cars, where the upper bound was defined with 21 kW.

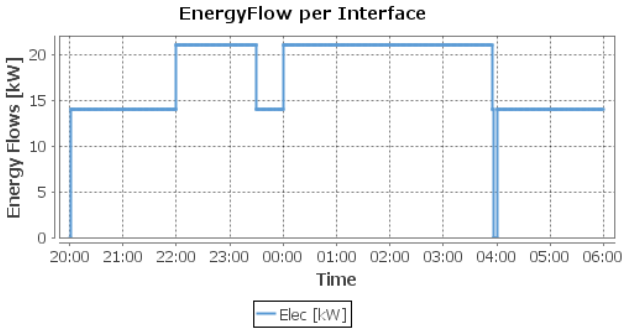


Fig.6: Cost-optimal charging with upper boundary for an EV Fleet

Compared to a simple price-optimal charging without upper boundary, the resulting costs were naturally slightly higher, since energy had to be consumed during periods with higher prices.

The time for the execution of the evaluation strategy and the creation of the execution schedules for each sub system ranged between 0.13 seconds for a single and 37.7 seconds for 200 systems. Figure 7 below shows the determined relationship between the number of systems involved and the thereon depending time for the evaluation process.

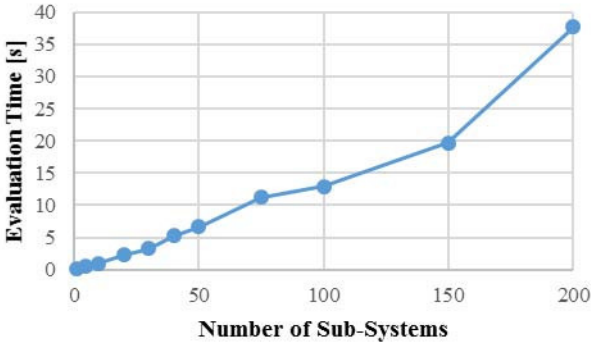


Fig.7: Time consumption for different numbers of sub systems

Since we considered the usage of the EOM within a planning process here, the required times for the evaluations seem to be very promising for us.

In addition to the experiments shown here, further application scenarios for the EOM were already investigated. Here we tested and improved the EOM for the usage with different energy carrier, but also for the real time

application within an electrical distribution network. For reasons of space, unfortunately, these experiments and results can not be explained or shown in this paper and will be part of future publications.

VI. DISCUSSION & COMPARISONS

Starting from the modular perspective, we already mentioned the use of the EOM within an energy agent that is located within or beside an on-site ECP. Here it is the task of the agent to configure the setup for an evaluation as described in the previous sections. Since the EOM also enables the parallel execution of several evaluation strategies, the final result may consist of a number of alternative schedules for the underlying system. As a next step, one plan has to be selected and passed into an execution process. This process then has to care about the system execution according to the selected plan. With the ability to deliberatively provide several plans for an agent, the EOM can also be considered as a kind of BDI concept that is well known in the agent community. In contrast to the general BDI concept the EOM, however, is concretely designed to meet the needs for comprehensively consider hybrid ECP's.

Furthermore, the EOM concurrently provides the ability to consider several ECP's within a system aggregation. This additionally enables to create central control approaches as they are realized in a vast majority today, e.g. in virtual power plants. Assuming the availability of suitable system descriptions, with the help of the EOM, adaptive mechanisms could be realized that enable a flexible integration of any type of ECP. Here the only question is, who initially owns the EOM; in case that this knowledge is located on-site, the model has to be transferred to the aggregator. Analogously to an on-site located energy agent, it is the task of the centralized aggregator process to collect all needed information for the Sub-ECP's and execute one or more evaluations strategies for the aggregation, in order to determine and select the overall and the sub-schedules.

But also for completely decentralized control approaches, the EOM can be applied. Therefore, a sequential turn-based message process can be considered. Instead of the use of the EOM for single systems, however, each system uses the aggregation part of the EOM. Here, the aggregation basically consists of the single local system and a summarizing schedule of all previous systems. The task of the local evaluation process is thus to generate a new local and a summarizing schedule. The latter has to be forwarded to the next ECP or energy agent respectively.

In addition to the above described three application types (local, centralized and decentralized), we believe that further types can be derived combinatorically. For example, several decentralized control processes could be controlled centrally and so on. In our view the EOM can be used for planning as well as for a real time control purposes. A crucial point, however, is the temporal resolution with which processes are modeled. While a fine granular resolution is helpful for real time applications, it will make planning process more expensive in the sense of needed computing time: every discrete step will require to resolve all possible subsequent states and thus to meet a decision for it. We assume that the modelling also requires to find a trade-off for that, but we can't answer this question at this point. Therefore, we will investigate the application of the EOM for further real time process in the near future.

Comparing the EOM with other approaches, requires first to highlight its unique position. Those highlights can be found with several facets that are: a) the comprehensive, system centric consideration of an energy conversion process that enables local reasoning processes for intelligent energy agents, b) the non-exclusive consideration of electrical energy and c) the adaptive characteristic that is offered by the EOM and that enables to realize centralized as well as decentralized control approaches.

Starting with the latter aspect and focusing on centralized control approaches, it is clear that such control approaches are in the main focus of most energy provider currently. Concrete solutions like this are known as energy management systems or can be found with the approach of the demand side management that was already mentioned in the background section of this paper. Here, one well known solution can be found e.g. with the *PowerMatcher* [17] and other. Beyond, also examples for decentralized control approaches were referred in the background section, but compared to the EOM, no solution known to us provides such flexibility for the actual implementation of a concrete control approach (e.g. centralized and/or decentralized).

Further, the consideration of energy conversion systems in general is a quite new aspect in the current Smart Grid research. Thus, similar adaptive and hybrid approaches as they are provided by the EOM are new and could not be found yet. Several general structures for an IT-based, interoperable management of distributed system are provided by the *Common Information Model* (CIM). Since the organizing *Distributed Management Task Force*[1] basically combines the interests of IT organizations and companies, energy specific model definitions like the EOM can't be found here.

Last but not least, the reasoning capability that is provided by the possible parallel execution of different evaluation strategies is a very unique property of the EOM. Even this idea is similar to the general BDI concept, the EOM is the only known, energy specific approach that provides the needed description cardinality for hybrid energy conversion processes. Thus, it has the capability to meet the broad range of requirements and applications in the energy domain.

[1] http://www.dmtf.org/

VII. Conclusion & Outlook

In this paper we presented our concept and the control approach for hybrid energy conversion processes by means a unifying Energy Option Model (EOM). Since this model has the capability to describe all relevant types of energy conversion processes, it provides the foundation for the design of new adaptive control approaches that permit a dynamic aggregation and optimization of aggregations of energy conversion processes.

In the next years, the model will be improved and further developed under the project *Agent.HyGrid*. Here it is planned to close the gap between agent-based simulations and real-world applications in order to produce comparable results for both cases. Thus, a realistic laboratory and test-bed environment will be created for control solution of Future Energy Grids.

Overall, we believe that standards that enable a homogeneous and in particular open development of "smart" energy systems are urgently required; for science as well as for systems used in real applications on-site. We further believe that therefore the concept of a generally accepted energy agent with a unified description of the underlying technical system that we call Energy Option Model is the necessary foundation.

The EOM introduced was registered for a patent. The grant is pending.

References

[1] V. Vyatkin, G. Zhabelova, N. Higgins, M. Ulieru, K. Schwarz, and N. C. Nair, "Standards-enabled Smart Grid for the future Energy Web," in Innovative Smart Grid Technologies (ISGT), 2010, 2010, pp. 1–9.

[2] C. Derksen, T. Linnenberg, R. Unland, and A. Fay, "Structure and Classification of Unified Energy Agents as a base for the systematic development of Future Energy Grids," EAAI—Engineering Applications of Artificial Intelligence, 2014.

[3] D. Nestle and J. Ringelstein, "Integration of DER into Distribution Grid Operation and Decentralized Energy Management," Smart Grids Europe, vol. 19, 2009.

[4] [Online] Available at: http://www.e-energy.de/de/modellregionen.php

[5] A. Faruqui and S. Sergici, "Household response to dynamic pricing of electricity: a survey of 15 experiments," Journal of Regulatory Economics, vol. 38, no. 2, pp. 193–225, 2010.

[6] S. Vandael, N. Boucké, T. Holvoet, K. De Craemer, and G. Deconinck, "Decentralized Coordination of Plug-in Hybrid Vehicles for Imbalance Reduction in a Smart Grid," in The 10th International Conference on Autonomous Agents and Multiagent Systems - Volume 2, 2011, pp. 803–810.

[7] A. Sternberg and A. Bardow, "Power-to-What? - Environmental assessment of energy storage systems," Energy Environ. Sci., vol. 8, no. 2, pp. 389–400, 2015.

[8] P. D. Lund, J. Lindgren, J. Mikkola, and J. Salpakari, "Review of energy system flexibility measures to enable high levels of variable renewable electricity," Renewable and Sustainable Energy Reviews, vol. 45, pp. 785–807, 2015.

[9] T. Krishnamurti, D. Schwartz, A. Davis, B. Fischhoff, W. B. de Bruin, L. Lave, and J. Wang, "Preparing for smart grid technologies: A behavioral decision research approach to understanding consumer expectations about smart meters," Energy Policy, vol. 41, pp. 790–797, 2012.

[10] P. A. Ioannou and J. Sun, Robust adaptive control. Courier Corporation, 2012.

[11] [Online] Available at: http://www.modbus.org.

[12] R. E. Mackiewicz, "Overview of IEC 61850 and Benefits," in Transmission and Distribution Conference and Exhibition, 2005/2006 IEEE PES, 2006, pp. 376–383.

[13] C. Derksen, T. Linnenberg, R. Unland, and A. Fay, "Unified Energy Agents as a Base for the Systematic Development of Future Energy Grids.," in MATES, 2013, vol. 8076, pp. 236–249.

[14] G. J. Van Wylen, R. E. Sonntag, and C. Borgnakke, Fundamentals of Classical Thermodynamics, no. Bd. 1. Wiley, 1994.

[15] D. J. Power, R. Sharda, and F. Burstein, Decision support systems. Wiley Online Library, 2002.

[16] J. Sudeikat, L. Braubach, A. Pokahr, W. Lamersdorf, and W. Renz, "Validation of BDI Agents," in PROMAS, 2006, pp. 185–200.

[17] J. K. Kok, C. J. Warmer, and I. G. Kamphuis, "PowerMatcher: Multiagent Control in the Electricity Infrastructure," in Proceedings of the Fourth International Joint Conference on Autonomous Agents and Multiagent Systems, 2005, pp. 75–82.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1517–1525
DOI: 10.15439/2016F76
ACSIS, Vol. 8. ISSN 2300-5963

Local Soft Constraints in Distributed Energy Scheduling

Astrid Nießе, Michael Sonnenschein
R&D Division Energy
OFFIS – Institute for Information Technology
26121 Oldenburg, Germany
{niesse, sonnenschein}@offis.de

Christian Hinrichs, Jörg Bremer
Environmental Informatics
University of Oldenburg
26129 Oldenburg, Germany
{christian.hinrichs, joerg.bremer}@uni-oldenburg.de

***Abstract*—In this contribution we present an approach on how to include local soft constraints in the fully distributed algorithm COHDA for the task of energy units scheduling in virtual power plants (VPP). We show how a flexibility representation based on surrogate models is extended and trained using soft constraints like avoiding frequent cold starts of combined heat and power plants. During the task of energy scheduling, the agents representing these machines include indicators in their choice for a new operation schedule. Using an example VPP we show that our approach enables the agents to reflect local soft constraints without sacrificing the global result quality.**

I. Introduction

IN DECENTRALIZED energy systems, small combined heat and power (CHP) plants, electrical storages and renewable energy units are aggregated both for an integration of these energy units into the energy markets and for the provision of ancillary services for a stable grid operation. Both applications are widely known as virtual power plants (VPP) and expected to be one of the core concepts for distributed energy systems [1]. One of the core challenges during operation of such a VPP arises from the complexity of the scheduling task due to the large amount of (small) energy units in the distribution grid [2]. To this end, multiple scalable scheduling algorithms have been proposed for distribution grid energy unit scheduling for VPP, with many of them using software agents technology and distributed algorithms [3], [4]. During scheduling, both global constraints (i. e. concerning the VPP as a whole) and local constraints (i. e. restricted to a single energy unit) have to be handled in an appropriate way. Both types of constraints may be either hard or soft constraints (cf. Table I). Local hard constraints set defined limits to the operational flexibility of an energy unit, thus defining feasible operation schedules. For the example of a CHP installation including thermal storage, the thermal capacity of the storage sets a hard constraint to the CHP's operation in combination with the current thermal load. Local soft constraints comprise technical or economical preferences, e. g. preferred operation times or the avoidance of technically unfavorable frequent cold starts. Global hard constraints can be market driven, like obligations from existing contracts. The reflection of global soft constraints is out of scope of the work presented here.

This work was partly supported by the Lower Saxony Ministry of Science and Culture through the "Niedersächsisches Vorab" grant program (grant ZN 2764 'Smart Nord'

TABLE I
Constraint types and examples for a VPP comprising CHP including thermal storage.
HC: hard constraints. SC: soft constraints.

	local: unit level	**global: VPP level**
HC	operational limits of thermal storage	energy amount contracted at the market
SC	avoidance of cold starts	*(out of scope of this work)*

VPP scheduling is the task of identifying distinct operation schedules for all components of a VPP that are within the flexibility range of the respective components. Thus, the process can be split in two parts: First, the flexibility of the components has to be assessed and modelled, and then a scheduling solution is identified within an optimization process. The overall process thus constitutes a multi-objective optimization problem, with the different types of constraints either being reflected during modelling or during optimization.

In this contribution, we extend a scalable VPP scheduling approach from previous work that takes into account local and global hard constraints with the ability to additionally reflect local soft constraints of individual energy units, thus solving the given multi-objective optimization problem. Individual preferences (i. e. preferences for single energy units and their operation) do not have to be disclosed within the VPP during the scheduling process.

The rest of this contribution is structured as follows: In section II we present relevant approaches for the tasks of modelling and optimization in distributed energy scheduling and identify basic algorithms used for the work presented here. We then elaborate on the chosen approach to model feasible operation schedules using a support vector data description (SVDD) based method and extend this model to include local soft constraints (section III). In section IV the COHDA heuristic is presented. In former work, this distributed algorithm has been used to generate VPP schedules satisfying local and global hard constraints [5]. We show how the modelled local soft constraints can be included in the optimization process

using an apriori-multi-objective optimization approach. Results from a case study evaluating the presented general multi-objective optimization approach are presented in section IV. We summarize and discuss open issues in section V.

II. Choosing the basic algorithms

In recent years, a large body of research has emerged in the field of distributed energy scheduling. For the tasks of constraint modelling and optimized scheduling relevant for the contribution at hand, in the following some basic approaches are presented, along with a discussion on the chosen algorithms.

Flexibility modelling can be understood as the task of modelling constraints. Apart from global VPP constraints, constraints often appear within single energy components; affecting the local decision making. Since these constraints are not of a distributed nature, they can be solved locally using central approaches. A widely used approach is the introduction of a penalty into the objective function that devalues a solution that violates some constraint [6]. In this way, the problem is transferred into an unconstrained one by treating fulfillment of constraints as additional objective. Alternatively, some combinatorial optimization problems allow for an easy repair of infeasible solutions. In this case, it has been shown that repairing infeasible solutions often outperforms other approaches [7]. Another popular method treats constraints or aggregations of constraints as separate objectives, also leading to a transformation into a (unconstrained) multi-objective problem [8]. A hierarchical approach that combines both hard and soft constraints in an explicit model formulation and weighted objective functions has been introduced in [9]. For optimization approaches in smart grid scenarios however, black-box models capable of abstracting from the intrinsic model have proven useful [10], [11]. They do not need to be known at compile time. A powerful, yet flexible way of constraint-handling is the use of a decoder that gives a search algorithm hints on where to look for schedules satisfying local hard constraints (*feasible schedules*) [11], [12]. This approach has been chosen in the work presented here. In section III-A an introduction to the decoder approach is given.

The chosen flexibility representation is the foundation for scheduling algorithms. The work presented by Akkermans, Ygge and Gustavsson in 1996 has been one of the first applications of distributed agent-based control in the electrical energy system [13]. The so-called HomeBots approach was motivated by an expected need for scalability, flexibility, adaptivity and broad applicability for future distributed energy systems [14]. Since this work, many distributed agent-based approaches have been developed in the disciplines of electrical engineering, control and system theory, information technology and information systems. The understanding of what constitutes a distributed system differs a lot, from software agents as gateways to the energy units and hierarchical systems [15] up to fully distributed algorithms [16]. Prior to the work presented here, a requirement based analysis has been done to identify appropriate algorithms for the task of distributed energy scheduling [17]. In the following a short overview on those algorithms already evaluated for energy scheduling tasks is presented.

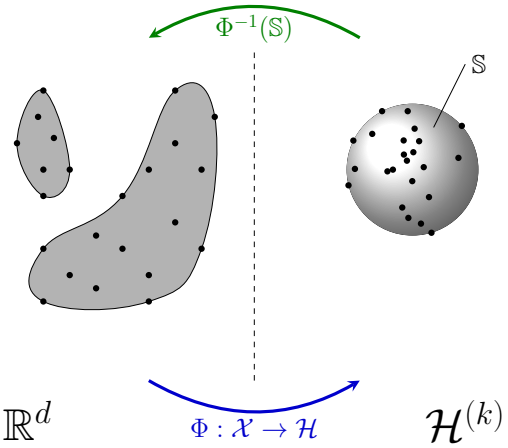


Fig. 1. General support vector model scheme for individual search spaces.

The Holonic Virtual Power Plant (Hol. VPP) presented in [18] was introduced for the reactive rescheduling process in VPPs. In this concept, a dedicated agent performs the task of evaluating a VPP schedule regarding global constraints. The agents presented in the concept are not capable of evaluating the quality of a new VPP schedule but decide on their contributions based on local constraints. The same approach has been chosen for Autonomous Virtual Power Plants [19]. ALMA [20] is a fully distributed and highly dynamic approach implemented for a dynamic supply-demand-matching task. The scheduling task at hand is not solved using a set of feasible schedules but based on a different modeling approach: The energy units communicate comfort levels, thus allowing to operate them flexibly within the defined levels. With COHDA, a fully distributed heuristic has been presented for energy scheduling [16]. The operational limits are modeled using the concept of a set of feasible schedules. Although possible, soft constraints have not yet been integrated in the process.

As COHDA satisfies the requirements regarding our motivating use case, the approach has been chosen as basic algorithm in the work presented here. For the full analysis, cf. [17]. In section IV-A we will introduce COHDA before discussing the concept of soft constraint integration.

III. Flexibility and Local Constraint Modelling

A. Decoder for Local Constraint Handling

In this section we briefly recap the technique for local hard constraint handling used in this work. For handling individual local constraints from different types of energy units, we use a decoder technique. An in-depth discussion of the technique can for example be found in [11], [12].

In general, a decoder is a technique that gives algorithms hints on where to look for feasible solutions and thus allows for a targeted search. It imposes a relationship between a decoder solution and a feasible solution and gives instructions

on how to construct a feasible solution [12]. A simple version without a need for machine learning techniques to deduce a meta-model, a response surface or similar uses directly a given set $\mathcal{X}$ of feasible schedules derived from a simulation model [16]. This approach has the limitation of supporting only discrete combinatorial problems. In [21] a homomorphous mapping between an n-dimensional hyper cube and the feasible region has been proposed in order to transform the problem into a topological equivalent one that is easier to handle. This approach has the problem of introducing additional parameters that have to be tuned and adapted to the problem instance at hand. In order to be able to derive a decoder automatically from any given energy unit model, [22] developed an approach based on a support vector model [23].

Fig. 2 shows the idea of using a so called support vector decoder. The basic idea is to start with a set of feasible example schedules derived from a simulation model of the respective energy unit and use this sample as a stencil for the region (the sub-space in the space of all schedules) that contains just feasible schedules.

We regard a schedule of an energy unit as a vector $\boldsymbol{s} = (s_0, \ldots, s_d) \in \mathcal{S} \subset \mathbb{R}^d$ with each element s_i denoting mean power generated (or consumed) during the ith time interval. As has been shown in [24], it is advantageous from a machine learning point of view to use scaled schedules for learning the feasible region. Thus, we construct the training set $\mathcal{X}$ by a normalization with

$$\begin{aligned} &\mathcal{N} : \mathcal{S} \to \mathcal{X} \subset [0,1]^d \\ &\qquad \boldsymbol{s} \mapsto \boldsymbol{x} = \mathcal{N}(\boldsymbol{s}), \text{with } x_i = \frac{s_i - p_{min}}{p_{max} - p_{min}}; \end{aligned} \tag{1}$$

p_{min} and p_{max} denoting minimum and maximum power respectively. The scaled sample $\mathcal{X}$ is then used as a training set for a support vector data description (SVDD) approach [25] that derives a geometrical description of the sub-space that contains the given data (in our case: the set of feasible schedules). Given a set of data samples, the inherent structure of the scope of action of the respective energy unit can be derived as follows: After mapping the data to a high dimensional feature space by means of an appropriate kernel, the smallest enclosing ball in this feature space is determined. When mapping back this ball to data space, it forms a set of contours enclosing the given data sample.

This task is achieved by determining a mapping

$$\Phi : \mathcal{X} \subset \mathbb{R}^d \to \mathcal{H};\ x \mapsto \Phi(x) \tag{2}$$

such that all data from a sample $\mathcal{X}$ from the feasible region $\mathcal{F}$ is mapped to a minimal hypersphere in some high-dimensional space $\mathcal{H}$. The minimal sphere with radius R and center a in $\mathcal{H}$ that encloses $\{\Phi(\boldsymbol{x}_i)\}_N$ can be derived from minimizing $\|\Phi(\boldsymbol{x}_i) - a\|^2 \leq R^2 + \xi_i$ with $\|\cdot\|$ as the Euclidean norm and slack variables $\xi_i \geq 0$ for soft constraints (here for getting a smoother ball).

After introducing Lagrangian multipliers and further relaxing to the Wolfe dual form, the well-known Mercer's theorem (cf. e.g. [26]) may be used for calculating dot products in $\mathcal{H}$ by means of a kernel in data space: $\Phi(\boldsymbol{x}_i) \cdot \Phi(\boldsymbol{x}_j) = k(\boldsymbol{x}_i, \boldsymbol{x}_j)$. In order to gain a more smooth adaption, it is known to be advantageous to use a Gaussian kernel: $k_G(\boldsymbol{x}_i, \boldsymbol{x}_j) = e^{-\frac{1}{2\sigma^2}\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2}$ [27]. Putting it all together, the equation that has to be maximized in order to determine the desired sphere is:

$$W(\beta) = \sum_i k(\boldsymbol{x}_i, \boldsymbol{x}_i)\beta_i - \sum_{i,j} \beta_i \beta_j k(\boldsymbol{x}_i, \boldsymbol{x}_j). \tag{3}$$

With $k = k_G$ one gets two main outcomes from the training procedure: the center $a = \sum_i \beta_i \Phi(\boldsymbol{x}_i)$ of the sphere in terms of an expansion into $\mathcal{H}$ and a function $R : \mathbb{R}^d \to \mathbb{R}$ that allows to determine the distance of the image of an arbitrary point from $a \in \mathcal{H}$, calculated in $\mathbb{R}^d$ by:

$$R^2(\boldsymbol{x}) = \ 1 - 2\sum_i \beta_i k_G(\boldsymbol{x}_i, \boldsymbol{x}) + \sum_{i,j} \beta_i \beta_j k_G(\boldsymbol{x}_i, \boldsymbol{x}_j). \tag{4}$$

Because all support vectors show the characteristics of being mapped onto the surface of the sphere, the sphere radius $R_{\mathcal{S}}$ can be easily determined by the distance of an arbitrary support vector to the center a. Thus the feasible region can now be modeled as $\mathcal{F} = \{\boldsymbol{x} \in \mathbb{R}^d | R(\boldsymbol{x}) \leq R_{\mathcal{S}}\} \approx \mathcal{X}$.

The comparably small set of support vectors together with a reduced version of vector $\boldsymbol{\beta}$ that contains non zero weight values (denoted $\boldsymbol{w}$) for the support vectors is sufficient for building the model. The model might then be used as a black-box that abstracts from any explicitly given form of constraints and allows for an easy and efficient decision on whether a given solution is feasible or not. In this way, the model allows for an easy check whether a given schedule is operable or not by using decision function (4).

So far, this surrogate model is just capable of checking feasibility when already given a schedule. In this way, the surrogate may tell feasible and infeasible schedules apart on behalf of the specific simulation model of the energy unit and thus already allows for an abstraction from any model specific implementation. On the other hand, it is not yet a sufficient constraint-handling technique as it still needs externally (e. g. by any optimization algorithm) generated schedules which can merely be checked. But, due to the tiny share of the search space that is actually feasible, it is quite unlikely that a feasible schedule is generated by an algorithm just by chance [28].

Hence, a way is needed to guide an algorithm where to look for feasible schedules. To achieve such systematic search for a good and still feasible solution, a decoder can be derived automatically from the support vector surrogate. The set of feasible schedules is represented as pre-image of a high-dimensional ball $\mathbb{S}$. Fig. 1 shows the geometric situation. This representation has some advantageous properties. Although the pre-image might be some arbitrary shaped non-continuous blob in $\mathbb{R}^d$, the high-dimensional representation is still a ball and thus geometrically easier to handle.

The relation is as follows: If a schedule is feasible, i.e. can be operated by the unit without violating any technical constraint, it lies inside the feasible region (grey area on the left hand side in Fig. 2). Thus, the schedule is inside the pre-image (that represents the feasible region) of the ball and thus

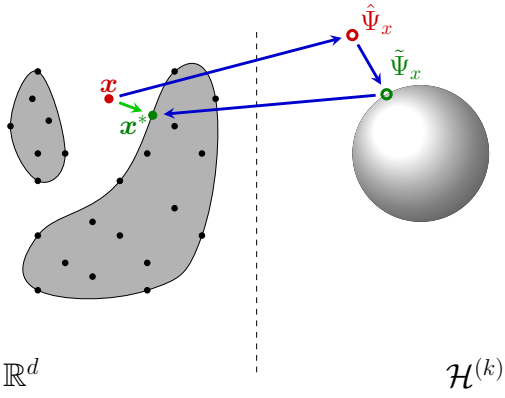


Fig. 2. General support vector decoder scheme for solution repair and constraint handling.

its image in the high-dimensional representation lies inside the ball. An infeasible schedule (e. g. $\boldsymbol{x}$ in Fig. 2) lies outside the feasible region and thus its image $\hat{\Psi}_x$ (generated by the empirical kernel mapping $\hat{\Psi}_x$) lies outside the ball. But we know some relations: the center of the ball, the distance of the image from the center and the radius of the ball. Hence, we can move the image of an infeasible schedule along the difference vector towards the center until it touches the ball. Finally, we calculate the pre-image of the moved image $\tilde{\Psi}_x$ (mover by translation function Γ_a) and get a schedule at the boundary of the feasible region: a repaired schedule $\boldsymbol{x}^*$ that is now feasible. We do not need a mathematical description of the original feasible region or of the constraints to do this. More sophisticated variants of transformation are e. g. given in [22].

Formally, we want to derive a mapping function (the so called decoder Θ)

$$\begin{aligned} \Theta : [0,1]^d &\rightarrow \mathcal{F}_{[0,1]} \subseteq [0,1]^d \\ x &\mapsto \Theta(x) \end{aligned} \tag{5}$$

that transforms any given (maybe in-feasible) schedule into a feasible one. This decoder mapping Θ is derived automatically from the trained SVDD repesentation of the search space using three steps:

$$\begin{array}{ccc} \boldsymbol{x} \in [0,1]^d & \xrightarrow{\hat{\Phi}_\ell} & \hat{\Psi}_x \in \mathcal{H}^{(\ell)} \\ \Theta \downarrow & & \downarrow \Gamma_a \\ \boldsymbol{x}^* \in \mathcal{F}_{[0,1]} \subseteq [0,1]^d & \xleftarrow[\tilde{\Phi}_\ell^{-1}]{} & \tilde{\Psi}_x \in \mathcal{H}^{(\ell)} \end{array} \tag{6}$$

Applying such decoder to some internal solution representation $\boldsymbol{x} \in [0,1]^d$ transforms the solution to some feasible solution $\Theta(\boldsymbol{x}) \in \mathcal{F}_{[0,1]}$.

Thus, we are able to transform any global scheduling problem into a formulation that is unconstrained regarding local hard constraints. Apart from finding a combination of schedules whose sum resembles a given target power profile best, further objectives are usually integrated due to the many-objective nature of energy scheduling.

This procedure of training a decoder has to be done only once prior to the scheduling process. During scheduling the all decoders are merely used for generating feasible schedules. For our experiments, training was usually done within milliseconds. For an in-depth discussion of computational issues of the decoder we refer to [29].

For distributed problem solving, the decoder can serve as a substitute for an often (particularly with regard to a fully automated generation) hardly derivable mathematical model of feasibility. Like in well studied industrial approaches for model predictive control [30] only a simulation model as a source for learning the model is needed. Due to the abstraction from the underlying simulation model (or real unit), no information on which operations are possible and no information on limiting restrictions, cost considerations or soft constraints are needed at runtime. Thus, the ad-hoc integration of arbitrary (even of at compile-time unknown units) becomes easily possible and hence eases the implementation of many control algorithms for the smart grid.

B. Modeling local soft constraints

In general, the satisfaction of a local soft constraint by a defined operation schedule s is modeled as a DER specific function $\mathcal{I}_i$ that assigns a value between 0 and 1 to the respective operation schedule, with i denoting the respective DER within the set of DER regarded in the scheduling task, i. e. the current VPP setup.

$$\begin{aligned} \mathcal{I}_i &: S \rightarrow [0,1] \\ \mathcal{I}_i(s) &= \sigma, \forall s \in S \end{aligned} \tag{7}$$

In the following, we omit the subscript i for reasons of brevity. As the soft constraint evaluation is a component-specific task it is performed locally in all upcoming definitions.

We now define an extended search space S' that integrates the soft constraints into the search space S:

$$S' : \{(\ s,\ \sigma^{(1)}, ..., \sigma^{(m)}\ ) \mid s \in S\} \tag{8}$$

With this definition, the extended search space S' is the set of all tuples of feasible operation schedules s and their respective soft constraint values, with m being the number of modelled soft constraints.

But how to model this extended search space? For the decoder concept presented in Section III-A, a possible integration of indicators has been shown: Data vectors containing the mean power levels for the respective time intervals are extended by one element per indicator to mixed feature vectors. This approach has been proposed for environmental performance indicators [23], but in general, arbitrary indicators can be added as long as a functional relationship exists between the power part and the indicator. In this way, we can build a modified sample $\boldsymbol{x}'$ as

$$\boldsymbol{x}' = (x_1, \ldots, x_d, \mathcal{I}^{(1)}_{[0,1]}(\boldsymbol{x}), \ldots, \mathcal{I}^{(m)}_{[0,1]}(\boldsymbol{x})),\ \boldsymbol{x}' \in [0,1]^{d+m}, \tag{9}$$

with the first d elements denoting real power and m trailing elements denoting indicator values. Whereas $\mathcal{I}$ takes a schedule, $\mathcal{I}_{[0,1]}$ maps an already scaled training vector $\boldsymbol{x}$ instead. This sample is fed into exactly the same support vector training process to build the model. The decoder is derived in exactly the same way. The decoder mapping Θ then likewise maps feature vectors $\boldsymbol{x}' \in \mathcal{X}'$

$$\begin{aligned}[0,1]^{d+m} &\rightarrow \mathcal{F}_{[0,1]} \times [\mathcal{I}_{[0,1]}]^m \\ \Theta(\boldsymbol{x}') &\mapsto (\boldsymbol{x}, \mathcal{I}^{(1)}_{[0,1]}(\boldsymbol{x}), \ldots, \mathcal{I}^{(m)}_{[0,1]}(\boldsymbol{x})).\end{aligned} \tag{10}$$

If $\boldsymbol{x}'$ is given with arbitrary values then $\Theta(\boldsymbol{x}')$ contains a feasible active power schedule in the first d elements as well as m elements evaluating this schedule correctly (slight inaccuracies are possible) with regard to the secondary optimization objectives.

Using this concept of the decoder approach including indicator reflection, we can set up an extended search space from a set of samples (i. e. normalized schedules) as shown in equation 9. For this purpose, we have to add the indicator value $\mathcal{I}$ for the chosen soft constraint to the sample schedule prior to the support vector training phase. In the example chosen here, we want to reduce the amount of cold starts within an operation schedule to minimize motor deterioration. We can infer the number of cold starts directly from the schedules within a preprocessing step. We omit the precise definition of cold starts for reasons of brevity, but usually it is a defined change of switching an engine off and on within a given time span. The indicator value thus matches the soft constraint value as given in equation 7. We now have to define the DER specific soft constraint function as an indicator for the amount of cold starts $\mathcal{I}^{cs}$ precisely to feed it into the SVDD training:

$$\mathcal{I}^{cs}(\mathsf{s}) = \left(1 - \frac{cs_{\mathsf{s}}}{cs_{max}}\right)^2 \tag{11}$$

with cs_{s} as the amount of cold starts in the given schedule s and cs_{max} as maximum amount of cold starts in the given schedule set. Using the squared value, a rising amount of cold starts is punished disproportionally high. For schedule sets without cold starts, the value is undefined – these cannot be used to distinguish schedules using this characteristic.[1]

It can be seen that there is a functional relationship between the schedule and the indicator, thus allowing to use the modified sample definition as given in equation 9 and using equation 1 to map schedule s to its normalized sample x. For each sample x in the given sample set we append the indicator as defined in equation 11 and use it for SVDD training.

As a result of these steps, we yield the extended search space S' and can now integrate this in the scheduling process.

IV. Distributed Energy Scheduling

A. Introducing COHDA

The Combinatorial Optimization Heuristic for Distributed Agents (COHDA, originally introduced in [16]) can be used to solve scheduling problems in VPPs. In the present contribution, we consider predictive scheduling: the goal is to select a schedule for each energy unit—from a given search space of feasible schedules with respect to a future planning horizon—such that a global objective function (e. g. a target power profile for the VPP) is optimized. This target profile is understood as global hard constraint within the scheduling process for the rest of this contribution. We will recap the approach briefly, based on the description in [5].

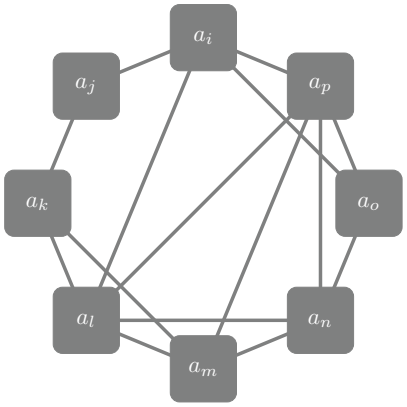


Fig. 3. Exemplary communication topology in the form of a small world topology for a system comprising eight agents.

The key concept of COHDA is an asynchronous iterative approximate best-response behavior, where each agent—representing a decentralized energy unit—reacts to updated information from other agents by adapting its own selected schedule with respect to the global objective. All agents $a_i \in A$ initially only know their own respective set of schedules S_i, so from an algorithmic point of view, the difficulty of the problem is given by the distributed nature of the system in contrast to the task of finding a common allocation of schedules for a global target power profile.

Thus, the agents coordinate by updating and exchanging information about each other. But, in order to preserve privacy, the amount of information that is exchanged is restricted. In particular, the set of feasible schedules S_i is not communicated as a whole by an agent a_i. Instead, the agents try to publish as little information as possible. How these possibly conflicting goals are handled, and how the system is able to converge to sound and satisfying solutions, will be explained in the following.

First of all, the agents are placed in an artificial communication topology (e. g. a *small world* topology, see Fig. 3), such that each agent is connected to a non-empty subset of other agents. To compensate for the resulting non-global view on the system, each agent a_i collects two distinct sets of information: on the one hand the believed current configuration γ_i of the system (that is, the most up to date information a_i has about currently selected schedules of all agents), and on the other hand the best known combination γ_i^* of schedules with respect to the global objective function it has encountered so far.

Beginning with an arbitrarily chosen agent by passing it a

[1]Please note that in the implementation presented here, $\mathcal{I}$ is defined identically for all DER, without limiting the applicability of the presented approach to DER specific soft constraint functions $\mathcal{I}_i$.

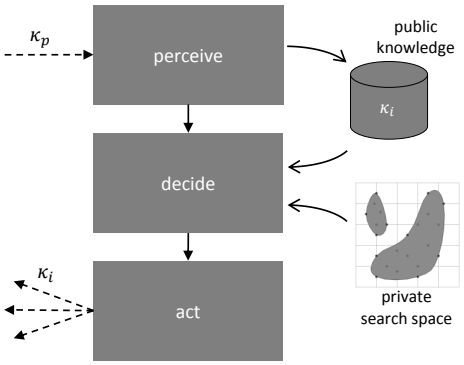


Fig. 4. The *perceive–decide–act* behavioral pattern in COHDA from the point of view of an agent a_i.

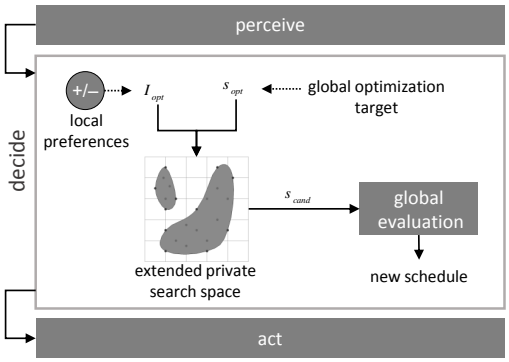


Fig. 5. The *decide* behavioral pattern of COHDA including the reflection of soft constraints.

message containing only the global objective (i. e. the target power profile), each agent repeatedly executes the three steps *perceive*, *decide*, *act* (cf. [5]) as visualized in Fig. 4:

1) **`perceive:`** When an agent a_i receives a message κ_p from one of its neighbors (say, a_p), it imports the contents of this message into its own memory.
2) **`decide:`** The agent then searches S_i for the best schedule regarding the updated system state γ_i and the global objective function, thus respecting the global hard constraints. The reflection of local hard and soft constraints depends on the chosen approach to model the energy unit's flexibility and will be discussed in a later part of this contribution. If a schedule can be found that satisfies both the global and the local objectives, a new schedule selection is created. For the following comparison, only the global objective function must be taken into account: If the resulting modified system state γ_i yields a better rating than the current solution candidate γ_i^*, a new solution candidate is created based on γ_i. Otherwise the old solution candidate still reflects the best schedule combination regarding the global objective the agent is aware of, so the just created schedule selection is discarded and the agent reverts to its schedule selection stored in γ_i^*.
3) **`act:`** If γ_i or γ_i^* has been modified in one of the previous steps, the agent finally broadcasts these to its neighbors in the communication topology.

Following this behavior, only small subsets of the sets of feasible schedules S_i are communicated by the agents. During this process, for each agent a_i, its observed system configuration γ_i as well as solution candidate γ_i^* are empty at the beginning, will be filled successively with the ongoing message exchange and will some time later represent valid solutions for the given optimization problem. After producing some intermediate solutions, the heuristic eventually terminates in a state where for all agents γ_i as well as γ_i^* are identical, and no more messages are produced by the agents. At this point, γ^* (which is the same for all agents then, so the index can be dropped) is the final solution of the heuristic and contains exactly one schedule selection for each agent.

B. Reflecting soft constraints during scheduling

In Fig. 4 an overview on the heuristic COHDA has been given from the perspective of a single agent representing a DER within a VPP. The algorithm is designed originally to optimize for a global objective only: In a two-stepped procedure, first a new schedule is chosen from the agent's a_i local search space S_i. After that, the resulting global result quality is evaluated. However, each agent must be permitted to decide itself which schedule it contributes. This way, technically, economically or ecologically rooted local soft constraints can be taken into account as secondary optimization goals. Moreover, in order to preserve privacy and autonomy of the participating entities, these individual secondary objectives must be treated as private to the corresponding agent, i. e. similar to the set of feasible schedules S_i, the local objectives are not part of the communicated information.

To integrate such local soft constraints into the decision process without compromising the convergence of the distributed algorithm, we modified the first step by replacing the search space S by the extended search space S' as defined in equation 8. During the *decide* phase, the agent thus has surplus information regarding the indicator (see Fig. 5). With this modelling approach, an additional constraint is integrated in the search space. Using the decoder approach as presented in section III-A, the agent can now try to identify a candidate schedule s that enhances the performance regarding the global objective (e. g. energy amount) and additionally enhance the performance regarding local quality as defined by the local soft constraint function $\mathcal{I}_i(s)$ (cf. equation 7). To enable this multi-objective optimization, the *decide* phase of COHDA is extended (cf. Fig. 5): In the first step, the needed schedule s_{opt} to best reach the global optimization target is calculated. The optimal indicator value $\mathcal{I}_{opt}$ (i. e. the best possible soft constraint performance) is added. For multiple soft constraints, additional indicator values would be added. In the next step this combination of needed schedule and indicator(s) is passed to the decoder for a targeted search. While without soft constraint modelling only the schedule best fitting the global

target would be identified, the decoder now returns a schedule depending on both the needed schedule s_{opt} and the optimal local soft constraint performance $\mathcal{I}_{opt}$. Using the decoder concept with an extended search space modelling approach thus leads to an apriori-multi-objective optimization with an implicit weighting of the different constraints. The schedule identified by the decoder is passed as candidate schedule s_{cand} to the evaluation of the global result quality, i. e. the global hard constraint.

In the example chosen here, the indicator gives information regarding the amount of cold starts contained in the schedule. The extended search space S' therefore is built using the indicator $\mathcal{I}^{cs}$ as defined in equation 11. The agent now can choose a candidate schedule that might enhance the performance regarding the global objective and minimize the number of cold starts simultaneously.

In summary, the outlined modelling and decision process yields a reasonable hierarchy of constraint handling in the domain of distributed energy scheduling (cf. Table I in section I): Using the decoder concept as depicted in section III-A, local hard constraints are modelled in such a way that only feasible schedules are returned by the decoder. Local soft constraints are used for SVDD training (see section III-B and guide the decoder targeted search (see section IV-B. Therefore, schedules satisfying the soft constraints are returned preferentially by the decoder. In the last step, the global evaluation is performed, reflecting global hard constraints. With this concept, local hard constraints are prioritized over global objectives, while local soft constraints are being taken into account with least importance.

V. Results

To evaluate the presented approach for the integration of soft constraints in distributed energy scheduling, two hypotheses have been chosen:

- H1 The integration of local soft constraints in the distributed scheduling enhances the performance of the chosen schedules regarding the modeled soft constraint, i. e. the local quality.
- H2 The integration of local soft constraints does not reduce the quality of the chosen solution regarding the global objective, i. e. producing the target power profile.

In the following, we will first introduce the evaluation setup, then discuss the evaluation results using these hypothesis. In all experiments, regarding the local soft constraints, we go with the example of reducing the amount of cold starts within an operation schedule to minimize motor deterioration, as introduced in section III-B.

A. Evaluation Setup

To evaluate the effect of an integration of local soft constraints in the distributed scheduling process, a setup is needed where agents have the choice to either reflect or ignore the performance indicator regarding the amount of cold starts $\mathcal{I}^{cs}(s)$. As the global objective is to fulfill a defined energy profile, enough flexibility within the aggregation of agents is needed to fulfill this objective either using schedules with high or low performance values regarding cold starts. In the experimental setup we therefore choose a mixed set of agents representing small CHP plants ($4.7\ \mathrm{kW_{el}}$): For 15 CHP plants, the conventional search space is used, without adding the indicator value. For additional 15 CHP plants, the search space is extended, thus allowing these agents to reflect the indicator during scheduling. For the support vector training phase, we need a set of schedules that can be distinguished regarding cold starts: On a winter day, CHP plants are expected to run nearly the whole day. Therefore, schedules of a CHP for a winter day are not suitable for the evaluation task. The opposite holds for a summer day. We chose a spring day and generated schedules from a CHP simulation for this task. The largest number of cold starts (cs_{max}) within this set has been 9. Thus, a schedule with 9 cold starts within one day might have been chosen without reflecting this constraint for each CHP.

For this aggregation of 30 CHP plants, different target profiles have been defined manually in such a way that fulfilling the profile is possible with more than 90 % accuracy, thus covering a range of possible target profiles. Each target setting has been simulated 100 times with different random seeds for generating the training set for reasons of statistical soundness.

B. Local Quality (Hypothesis H1)

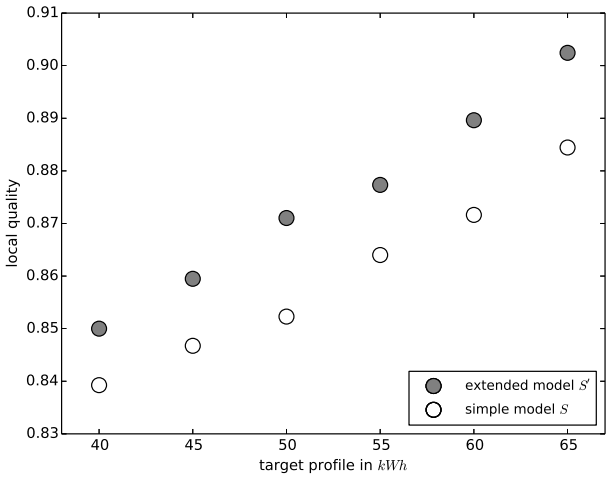


Fig. 6. Simulation results for a VPP with 30 CHPs and different target profiles. Mean values from 100 simulation runs are shown.

In Fig. 6 the results for the simulative experiments are shown regarding local quality: The horizontal axis denotes the different target profiles. On the vertical axis the mean local quality is shown. With only one soft constraint given in our scenario, we define σ as the local quality of a schedule under evaluation (cf. equation 7). The filled circles depict the mean value of 15 CHP plants over 100 simulations for the agents reflecting the amount of cold starts in their search space, whereas the unfilled circles show the mean local quality of

those 15 CHP plants over 100 simulations for the agents that do not consider this local soft constraint.

It can be seen that the local quality is higher for all target profiles, if soft constraints are integrated in the scheduling process. Additionally a trend can be seen: The higher the energy amount of the target profile, the better is the local quality. This can be explained as follows: If more electrical energy has to be produced, the agents choose schedules with longer runtimes. Thus, less cold starts are expected.

For one of the simulation setups (target profile 50 kWh), the raw data are depicted in Fig. 7. It can be seen that using the extended search space model (S'), the number of schedules with 2 cold starts is reduced, whereas the number of schedules without cold starts is increased. There is a slight rise in the number of schedules with 1 and 3 cold starts. This rise is considered to be non-significant compared to the effect regarding the reduction of schedules without cold starts. In general, a shift to schedules with less cold starts can be observed. The results are similar for the other simulation setups but not displayed here.

With the given results, we consider hypothesis H1 strengthened: The integration of local soft constraints in the distributed scheduling enhances the performance regarding the modeled soft constraint.

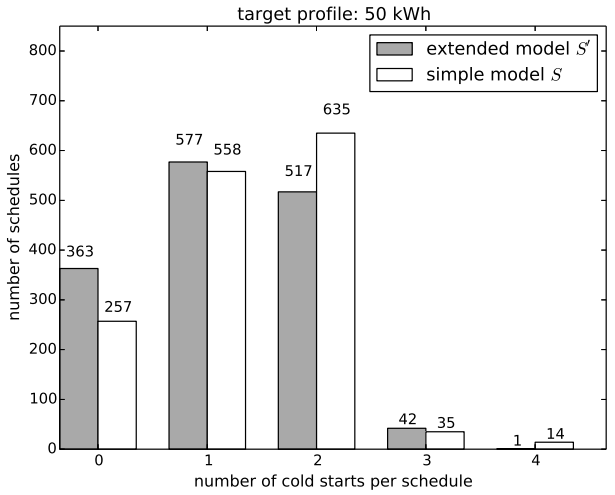


Fig. 7. Detailed simulation results for a VPP with 30 CHPs and target profile 50 kWh. Regarding 100 simulation runs, the distribution of schedules with respect to the amount of cold starts per schedule is shown.

C. Global Quality (Hypothesis H2)

We now focus on the effects of the integration of soft constraints in the scheduling process on the global quality. In Table II the normalized global quality regarding the target profile fulfillment is summarized using the mean values over 100 simulation runs for different target profiles. In general, the energy unit aggregations perform best for a profile with either 50 or 55 kWh, although a high quality could be reached for all simulation settings. This depends on the chosen aggregation of energy units and their feasible schedules: Within the defined range from 40 to 65 kWh, enough operational flexibility is given to adapt to a defined target profile. We now compare the global quality within one target profile simulation setting with and without the reflection of soft constraints: It can be seen that for all profiles, the values differ only very slightly. This effect can be understood from the chosen concept of soft constraint integration: If an agent manages to identify an operation schedule that will outperform the current target fulfillment, this would be chosen although this schedule might decrease local quality. With the chosen concept of guiding the search within the schedule search space along the soft constraint performance though, the heuristic COHDA tends to find those local optima that not only increase global quality but additionally show better local quality.

With the given results, we consider hypothesis H2 strengthened: The integration of local soft constraints does not reduce the quality of the chosen solution regarding the target profile delivery significantly for the chosen setting.

TABLE II
COMPARISON OF TARGET PROFILE FULFILLMENT WITH AND WITHOUT REFLECTION OF SOFT CONSTRAINTS. MEAN VALUES FROM 100 SIMULATIONS RUNS PER TARGET PROFILE ARE GIVEN.

40kWh	**45kWh**	**50kWh**	**55kWh**	**60kWh**	**65kWh**
with reflection of soft constraints					
0.9862	0.9886	0.9921	0.9939	0.9874	0.9590
without reflection of soft constraints					
0.9850	0.9907	0.9931	0.9937	0.9830	0.9605

VI. CONCLUSION AND OUTLOOK

In this contribution we presented an approach on how to include local soft constraints in the fully distributed algorithm COHDA for the task of energy units scheduling in virtual power plants (VPP). For this task, we extended a flexibility representation based on SVDD using indicator values and used these indicators to guide the search for a new schedule. Using the example of preventing frequent cold starts for CHP plants, we could show that the presented approach enables the agents to reflect the modeled local soft constraint without sacrificing the global result quality. As the information on local soft constraints is not communicated within the system and considered only locally, the presented approach reveals benefits regarding privacy without sacrificing global result quality.

With these results, further work should be done on the integration of extended search spaces in the presented distributed scheduling heuristic COHDA. With the extension of the search space using indicators, an implicit weighting is given by the length of the schedules and the number of indicators. Additional evaluation effort is needed to yield an appropriate weighting depending on the specific soft constraint. A straightforward approach would be to adapt the weights of an indicator by multiplying its value during the SVDD training phase, thus

yielding an explicit weighting. The formalization given in the contribution at hand is compatible with this concept.

Additionally, the boundaries of effectiveness for energy unit aggregations with less flexibility should be evaluated, especially compared to other multi-objective optimization concepts: With the presented approach, soft constraints guide the search in the search space. Therefore it has to be evaluated, if for some types of DER the global quality would be reduced to an unaccepted extent. Straight-forward extensions of the presented approach like time-dependent soft constraint relaxation could overcome such problems.

REFERENCES

[1] H.-J. Appelrath, H. Kagermann, and C. Mayer, Eds., *Future Energy Grid. (acatech STUDY)*. acatech, Munich, 2012.

[2] S. D. J. McArthur, E. M. Davidson, V. M. Catterson, A. L. Dimeas, N. D. Hatziargyriou, F. Ponci, and T. Funabashi, "Multi-Agent Systems for Power Engineering Applications—Part 1: Concepts, Approaches, and Technical Challenges," *IEEE Transactions on Power Systems*, vol. 22, pp. 1743–1752, 2007.

[3] ——, "Multi-Agent Systems for Power Engineering Applications—Part 2: Technologies, Standards, and Tools for Building Multi-Agent Systems," *IEEE Transactions on Power Systems*, vol. 22, pp. 1753–1759, 2007.

[4] P. Vrba, V. Mařík, P. Siano, P. Leitão, G. Zhabelova, V. Vyatkin, and T. Strasser, "A Review of Agent and Service-Oriented Concepts Applied to Intelligent Energy Systems," *IEEE Transactions of Industrial Informatics*, vol. 10, no. 3, pp. 1890–1903, 2014.

[5] A. Nieße, S. Beer, J. Bremer, C. Hinrichs, O. Lünsdorf, and M. Sonnenschein, "Conjoint Dynamic Aggregation and Scheduling Methods for Dynamic Virtual Power Plants," in *Proceedings of the 2014 Federated Conference on Computer Science and Information Systems*, ser. Annals of Computer Science and Information Systems, M. Ganzha, L. A. Maciaszek, and M. Paprzycki, Eds., vol. 2. IEEE, 2014. doi: 10.15439/2014F76. ISBN 978-83-60810-58-3 pp. 1505–1514. [Online]. Available: http://dx.doi.org/10.15439/2014F76

[6] A. Smith and D. Coit, *Handbook of Evolutionary Computation*. Department of Industrial Engineering, University of Pittsburgh, USA: Oxford University Press and IOP Publishing, 1997, ch. Penalty Functions, p. Section C5.2.

[7] G. E. Liepins and M. D. Vose, "Representational issues in genetic optimization," *Journal of Experimental and Theoretical Artificial Intelligence*, vol. 2, 1990.

[8] O. Kramer, "A review of constraint-handling techniques for evolution strategies," *Appl. Comp. Intell. Soft Comput.*, vol. 2010, pp. 1–19, 01 2010. doi: http://dx.doi.org/10.1155/2010/185063

[9] A. Schiendorfer, J.-P. Steghöfer, and W. Reif, "Synthesised Constraint Models for Distributed Energy Management," in *Proceedings of the 2014 Federated Conference on Computer Science and Information Systems*, ser. Annals of Computer Science and Information Systems, M. Ganzha, L. A. Maciaszek, and M. Paprzycki, Eds., vol. 2. IEEE, 2014. doi: 10.15439/2014F49. ISBN 978-83-60810-58-3 pp. 1529–1538. [Online]. Available: http://dx.doi.org/10.15439/2014F49

[10] F. Gieseke and O. Kramer, "Towards non-linear constraint estimation for expensive optimization," in *Applications of Evolutionary Computation*, ser. Lecture Notes in Computer Science, A. Esparcia-Alcázar, Ed. Springer Berlin Heidelberg, 2013, vol. 7835, pp. 459–468. ISBN 978-3-642-37191-2

[11] J. Bremer and M. Sonnenschein, "Model-based integration of constrained search spaces into distributed planning of active power provision." *Comput. Sci. Inf. Syst.*, vol. 10, no. 4, pp. 1823–1854, 2013.

[12] C. A. Coello Coello, "Theoretical and numerical constraint-handling techniques used with evolutionary algorithms: a survey of the state of the art," *Computer Methods in Applied Mechanics and Engineering*, vol. 191, no. 11-12, pp. 1245–1287, Jan. 2002. doi: 10.1016/S0045-7825(01)00323-1

[13] H. Akkermans, F. Ygge, and R. Gustavsson, "Homebots: Intelligent decentralized services for energy management," in *Fourth International Symposium on the Management of Industrial and Corporate Knowledge ISMICK'96*, Rotterdam, NL, Oktober 1996.

[14] R. Gustavsson, "Agents with Power," *Communications of the ACM*, vol. 42, no. 3, pp. 41–47, 1999.

[15] S. Lehnhoff, *Dezentrales vernetztes Energiemanagement - Ein Ansatz auf Basis eines verteilten Realzeit-Multiagentensystems*. Vieweg + Teubner, 2010.

[16] C. Hinrichs, S. Lehnhoff, and M. Sonnenschein, "A Decentralized Heuristic for Multiple-Choice Combinatorial Optimization Problems," in *Operations Research Proceedings 2012*. Springer, 2014. doi: 10.1007/978-3-319-00795-3_43. ISBN 978-3-319-00795-3 pp. 297–302.

[17] A. Nieße and M. Sonnenschein, "A Fully Distributed Continuous Planning Approach for Decentralized Energy Units," in *45. Jahrestagung der Gesellschaft für Informatik e.V. (GI), Informatik, Energie und Umwelt*. Cottbus: Gesellschaft für Informatik, Köllen Druck+Verlag, 2015.

[18] M. Tröschel, *Aktive Einsatzplanung in holonischen Virtuellen Kraftwerken*. Oldenburg: OlWIR, Oldenburger Verl. für Wirtschaft, Informatik und Recht, 2010. ISBN 978-3-939704-55-3

[19] G. Anders, F. Siefert, J.-P. Steghöfer, H. Seebach, F. Nafz, and W. Reif, "Structuring and Controlling Distributed Power Sources by Autonomous Virtual Power Plants," in *Proceedings of the IEEE Power and Energy Student Summit (PESS)*, 2010.

[20] E. Pournaras, "Multi-level Reconfigurable Self-organization in Overlay Services," Ph.D. dissertation, TU Delft. ISBN 9789461860989 2013.

[21] S. Koziel and Z. Michalewicz, "Evolutionary algorithms, homomorphous mappings, and constrained parameter optimization," *Evol. Comput.*, vol. 7, pp. 19–44, 03 1999. doi: http://dx.doi.org/10.1162/evco.1999.7.1.19

[22] J. Bremer and M. Sonnenschein, "Constraint-handling for optimization with support vector surrogate models – a novel decoder approach," in *ICAART 2013 – Proceedings of the 5th International Conference on Agents and Artificial Intelligence*, J. Filipe and A. Fred, Eds., vol. 2. Barcelona, Spain: SciTePress, 2013. doi: 10.5220/0004241100910100. ISBN 978-989-8565-38-9 pp. 91–105.

[23] J. Bremer, B. Rapp, and M. Sonnenschein, "Including environmental performance indicators into kernel based search space representations," in *Information Technologies in Environmental Engineering*, ser. Environmental Science and Engineering, P. Golinska, M. Fertsch, and J. Marx-Gómez, Eds. Springer Berlin Heidelberg, 2011, vol. 3, pp. 275–288. ISBN 978-3-642-19535-8. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-19536-5_22

[24] P. Juszczak, D. Tax, and R. P. W. Duin, "Feature scaling in support vector data description," in *Proc. ASCI 2002, 8th Annual Conf. of the Advanced School for Computing and Imaging*, E. Deprettere, A. Belloum, J. Heijnsdijk, and F. van der Stappen, Eds., 2002, pp. 95–102.

[25] D. M. J. Tax and R. P. W. Duin, "Support vector data description," *Mach. Learn.*, vol. 54, no. 1, pp. 45–66, 2004. doi: http://dx.doi.org/10.1023/B:MACH.0000008084.60811.49

[26] B. Schölkopf, S. Mika, C. Burges, P. Knirsch, K.-R. Müller, G. Rätsch, and A. Smola, "Input space vs. feature space in kernel-based methods," *IEEE Transactions on Neural Networks*, vol. 10(5), pp. 1000–1017, 1999.

[27] A. Ben-Hur, H. T. Siegelmann, D. Horn, and V. Vapnik, "Support vector clustering," *Journal of Machine Learning Research*, vol. 2, pp. 125–137, 2001.

[28] J. Bremer and M. Sonnenschein, "Sampling the search space of energy resources for self-organized, agent-based planning of active power provision," in *27th International Conference on Environmental Informatics for Environmental Protection, Sustainable Development and Risk Management, EnviroInfo 2013, Hamburg, Germany, September 2-4, 2013. Proceedings*, ser. Berichte aus der Umweltinformatik, B. Page, A. G. Fleischer, J. Göbel, and V. Wohlgemuth, Eds. Shaker, 2013. ISBN 978-3-8440-1676-5 pp. 214–222.

[29] J. Bremer, "Constraint-Handling mit Supportvektor-Dekodern in der verteilten Optimierung," Ph.D. dissertation, 2015. [Online]. Available: http://oops.uni-oldenburg.de/2336/

[30] L. Grüne and J. Pannek, *Nonlinear Model Predictive Control: Theory and Algorithms*, 1st ed., ser. Communications and Control Engineering. Springer, 2011. [Online]. Available: http://www.springer.com/978-0-85729-500-2

Software Systems Development & Applications

SSD&A is a FedCSIS conference area aiming at integrating and creating synergy between FedCSIS events that thematically subscribe to the discipline of software engineering. The SSD&A area emphasizes the issues relevant to developing and maintaining software systems that behave reliably, efficiently and effectively. This area investigates both established traditional approaches and modern emerging approaches to large software production and evolution. Events that constitute SSD&A are:

- BTMSPA'16—1st Symposium on Balancing Traditional and Modern Software Process Approaches
- MDASD'16—4th Workshop on Model Driven Approaches in System Development
- MIDI'16 - 4th Conference on Miltimedia, Interaction, Design and Innovation
- SEW-36—The 36th IEEE Software Engineering Workshop

1st Symposium on Balancing Traditional and Modern Software Process Approaches

RESEARCH studies conducted in the area of software engineering have proposed hundreds of techniques and recommendations. On the other hand, software projects are performed in frame of several constraints including required quality, time and budget constraints as well as staff competences. Thus there is a need for selection of the techniques which allow practitioners to achieve a fair balance between time and effort spent on performing these techniques and benefits they provide.

Several traditions exist in software engineering including formal, methodological, user-centered, reuse and agile approaches. Additionally, novel approaches are still being proposed, e.g. collaborative approaches. Each of the approaches focuses on solving a different kind of problems. The proper balance between approaches and, consequently techniques in use, depends on the characteristics of the project at hand.

This symposium aims at exchanging experience and making progress in the knowledge about software process configuration with fair balance between traditional and modern approaches. We invite both experience reports from industry and scientific studies of integrated approaches.

TOPICS

- Balance between agile and disciplined approaches
- Innovation and creativity in software engineering
- Collaborative games in software processes
- Business-IT alignment
- Enterprise integration, business integration and systems integration
- IT-enabled innovations in organizations
- Cooperative, distributed, and collaborative software engineering
- Variability across the software life cycle
- Innovative platforms, architectures and technologies for IS
- Quality assurance and management
- Social media, open data, Internet of Things in business processes
- Methods, tools and human factors in IS/IT management
- Industrial case studies and experience reports related to the above topics

EVENT CHAIRS

- **Bobkowska, Anna,** Gdansk University of Technology, Poland
- **Brudło, Piotr,** Gdansk University of Technology, Poland
- **Przybyłek, Adam,** Gdansk University of Technology, Poland

PROGRAM COMMITTEE

- **Alshayeb, Mohammad,** King Fahd University of Petroleum and Minerals, Saudi Arabia
- **Bauer, Veronika,** Technische Universität München, Germany
- **Belle, Alvine Boaye,** École de Technologie Supérieure, Canada
- **Blech, Jan Olaf,** RMIT University, Australia
- **Borg, Markus,** SICS Swedish ICT AB, Sweden
- **Chatzigeorgiou, Alexandros,** University of Macedonia, Greece
- **Czarnecki, Krzysztof,** Gdańsk University of Technology, Poland
- **Diebold, Philipp,** Fraunhofer IESE, Germany
- **Gregory, Peggy,** University of Central Lancashire, United Kingdom
- **Jasek, Roman,** Tomas Bata University in Zlin, Czech Republic
- **Kaloyanova, Kalinka,** Sofia University, Bulgaria
- **Kapitsaki, Georgia,** University of Cyprus, Cyprus
- **Katić, Marija,** School of Computing, Engineering and Physical Sciences, United Kingdom
- **Knodel, Jens,** Fraunhofer IESE, Germany
- **Kuchta, Jarosław,** Gdansk University of Technology, Poland
- **Madeyski, Lech,** Wrocław University of Technology, Poland
- **Mangalaraj, George,** Western Illinois University, United States
- **Merunka, Vojtech,** Czech Technical University in Prague (Associated Professor), Czech Republic
- **Molhanec, Martin,** Czech Technical University in Prague, Czech Republic
- **Morales Trujillo, Miguel Ehecatl,** National Autonomous University of Mexico, Mexico
- **Nawrocki, Jerzy,** Poznan University of Technology, Poland
- **Norta, Alex,** Tallinn University of Technology, Estonia
- **Noyer, Arne,** University of Osnabrueck and Willert Software Tools GmbH, Germany
- **Özkan, Necmettin,** Turkiye Finans Participation Bank, Turkey
- **Pereira, Rui Humberto R.,** Instituto Politecnico do Porto, Portugal
- **Przechlewski, Tomasz,** Powiślańska Szkoła Wyższa w Kwidzynie, Poland

- **Ramsin, Raman,** Sharif University of Technology, Iran
- **Salnitri, Mattia,** University of Trento, Italy
- **Santos Neto, Pedro de Alcântara dos,** Universidade Federal do Piauí, Brazil
- **Śmiałek, Michał,** Politechnika Warszawska, Poland
- **Soares, Michel,** Federal University of Sergipe, Brazil
- **Soja, Piotr,** Cracow University of Economics, Poland
- **Tarhan, Ayca,** Hacettepe University Computer Engineering Department, Turkey
- **Wiszniewski, Bogdan,** Gdansk University of Technology, Poland
- **Zarour, Nacer Eddine,** University Constantine2, Algeria
- **Zeid, Amir,** American University of Kuwait, Kuwait

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1531–1538
DOI: 10.15439/2016F430
ACSIS, Vol. 8. ISSN 2300-5963

Using ESSENCE ALPHAs in a CMMI level 5 software development organization

Miguel Ehécatl Morales-Trujillo
KUALI-KAANS Research Group and Engineering Faculty of the National Autonomous University of Mexico, Mexico City, Mexico
migmor@ciencias.unam.mx

Hanna Oktaba
KUALI-KAANS Research Group and Science Faculty of the National Autonomous University of Mexico, Mexico City, Mexico
hanna.oktaba@ciencias.unam.mx

María Julia Orozco
Ultrasist, Av. Revolución 1181 8th floor, Merced Gómez, Mexico City, Mexico
mjorozcom@ultrasist.com.mx

Abstract — **Managing a software development project is a challenging task; time and effort is required to monitor the project's health and progress. In this context, organizations look for proposals that would assist them in this task. Recently a new and light alternative was introduced: ALPHAs, which are central elements of ESSENCE – Kernel and Language for Software Engineering Methods OMG standard. This paper presents the experience of a Mexican organization that uses ALPHAs to enhance its processes. The paper summarizes the actual use of ALPHAs in the organization, their advantages and disadvantages, and outlines some advice for organizations wishing to adopt ALPHAs. We conclude that ALPHAs are useful for monitoring and controlling software endeavors. Moreover, their harmonization with the organization's current process was a beneficial factor in renewing the CMMI-DEV and CMMI-SVC level 5 appraisals.**

Keywords: ALPHA, ESSENCE, CMMI, quality, software process.

I. Introduction

SOFTWARE development is a challenging task that involves soft and hard skills, such as technical knowledge to create software products of quality and social abilities to make the participants of a project work together in order to achieve a goal.

The characteristics of a software development project, in particular, asset specificity and uncertainty, affect the choice of governance structure of a project [1]. For that reason, it is important to define a governance structure for monitoring and controlling the project according to its particular context.

According to [2], this governance structure can follow a top-down or a bottom-up approach. The top-down governance approach corresponds to process-centered methodologies and bottom-up governance is similar to agile methodologies.

On the one hand, process-centered methodologies are based on process reference models, standards or body of knowledges. Examples of these are CMMI [3], ISO/IEC 12207 [4] or PMBOK [5]. To follow such a plan driven process is essential since it is the backbone of the endeavor and reduces time and cost deviations; yet, to follow it tightly requires a great effort and does not assure a high quality product.

Altogether, it is important to take into account that executing project management activities alone does not mean *managing* a software development project, and communication is vital in order to determine its health and progress.

On the other hand, agile approaches, like SCRUM [6], KANBAN [7] or ESSENCE [8], are an effective strategy for communicating with work teams in a timely and accurate fashion; their primary focus is on the people involved in the project and on delivering a product that fully satisfies the client's needs. It is important to bear in mind that the use of agile methods does not guarantee the appearance of the mentioned benefits in each project, or their contribution to a higher efficacy of the whole organization [9].

Agile or traditional, all these activities aim at figuring out how the project is progressing and allowing team members to make competent decisions. As a whole, no matter what approach is chosen, project management activities consume a large part of the project's time and effort.

In this context, organizations constantly explore alternatives to improve their processes and product quality and incorporate best practices from different sources, no matter agile or traditional. Some early works provide interesting proposals that show a growing interest in recent years on the part of the software engineering community regarding process improvement environments where multiple models are involved. [10].

In particular, this paper describes two models relevant for this study: ESSENCE and CMMI. ESSENCE is an Object Management Group (OMG) standard of a great value. It provides a domain model for organizing different factors that influence the success of a software engineering endeavor [11]. One of the values added by ESSENCE is to surface unknown issues, support the evaluation of the team, generating reflective team discussions through a thinking framework that is holistic, state-based, goal-driven, and method-agnostic [12].

As for CMMI, it is a well-known set of practices divided into three models: for Development (CMMI-DEV) [13], for

Acquisition (CMMI-ACQ) [14] and for Services (CMMI-SVC) [15]. Its last version 1.3 was published in 2010.

This paper presents the experience of a Mexican organization, evaluated CMMI level 5 that introduced ESSENCE into its processes, and uses it as an agile mechanism to evaluate the progress of its projects and its work products quality.

The structure of this paper is as follows: in Section II a general background of the ESSENCE standard, its ALPHAs and the description of the organization are presented. Section III describes how the ALPHAs were used within a CMMI based organizational process. Section IV concentrates the results and lessons learned. Finally, Section V contains conclusions and future work.

II. Background

This section presents an overview of the ESSENCE standard and its ALPHAs. Later, the context of the organization under discussion is detailed.

A. ESSENCE

ESSENCE – Kernel and Language for Software Engineering Methods was published in 2014. Its origins date back to the SEMAT initiative that supported re-foundation of Software Engineering discipline through the identification of its "common ground" [16] as a set of elements essential for software engineering endeavors. This initiative was launched in 2009 and was endorsed by the OMG in 2010.

ESSENCE consists of two parts: the Kernel and the Language. The Kernel contains a small number of "things we always work with" and "things we always do" when developing software systems [16], while the Language is used to describe methods and practices. The Kernel consists of a set of concepts called ALPHAs that provide an object-oriented state-based model of a software engineering endeavor [11].

According to [11], ESSENCE main benefits are the following:

- It provides a comprehensive model for a large-scale process improvement endeavor;
- It is a context aware model, making visible to practitioners both theory and practice;
- It is an evolvable and participatory model, it can be used in any time and by anybody of the work team.

Another value of ESSENCE comes by providing a structure for analyzing and organizing the context and factors of software engineering endeavors from different dimensions [11].

It is worth mentioning that the initial objective of the standard was to refound Software Engineering, placing it a solid theory base and giving professionals the means to define their own practices and methods. However, this study shows that the main usage addresses project management issues.

B. ALPHAs

ALPHAs are the top-level concepts, which refer to the essential elements of software engineering endeavors, relevant to an assessment of progress and health [17]. In fact, ALPHA originated as an acronym for Abstract Level Progress and Health Attribute, and its main purpose is to determine fast and at any time how the project is doing. ALPHAs provide consistent language and measurable objectives with which to assess the current state, or articulate next steps and goals [18].

An ALPHA's components are the following:

- A representative and unique name.
- A set of states through which an ALPHA passes during its lifecycle.
- A checklist for each state used to determine if the state is reached or not.

This structure allows practitioners to evaluate the project based on the states of each ALPHA. The checklist for an ALPHA state contains a number of checkpoints that can be referenced to determine whether and to what degree the project has reached that state [19]. Therefore, it is possible to establish the state of a software engineering endeavor through the ALPHAs states [11].

There are seven ALPHAs divided into three groups, which are called Customer, Solution and Endeavor areas of concern. Figure 1 displays all the ALPHAs within their areas of concern:

- Customer (green)
  1. Opportunity (6 states)
  2. Stakeholder (6)
- Solution (yellow)
  3. Requirements (6)
  4. Software System (6)
- Endeavor (blue)
  5. Work (6)
  6. Team (5)
  7. Way of Working (6)

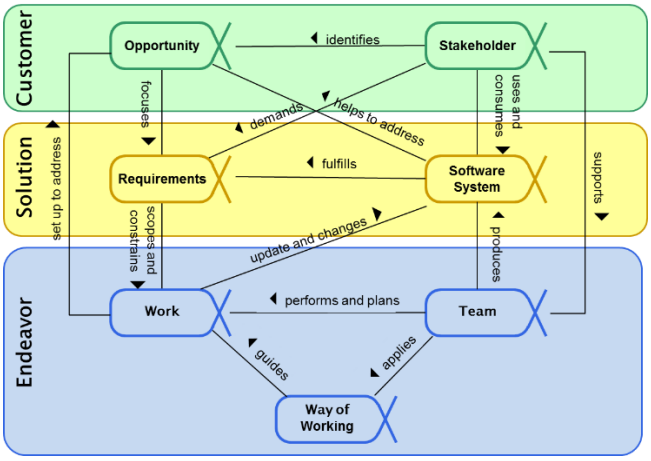


Fig. 1 Areas of concern and their ALPHAs [8]

For an easy and practical use, ALPHAs were represented as cards. Each card corresponds to one state of an ALPHA, and the color code indicates to which area of concern it belongs. As an example, Figure 2 shows the state *Addressed* of the ALPHA Opportunity.

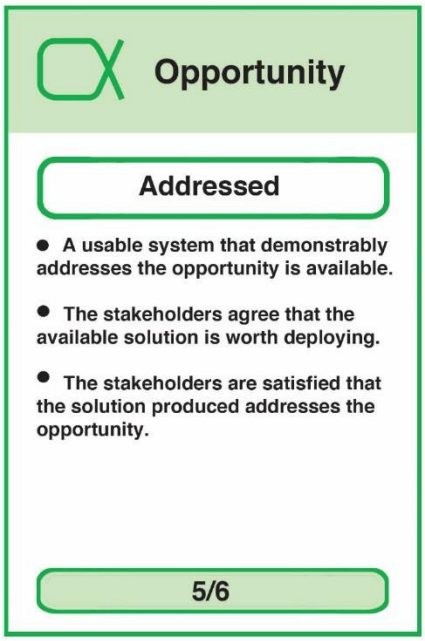


Fig. 2 An ALPHA card, based on [8]

For an opportunity to be *addressed*, the three elements of the checklist should be achieved. Notice that it is the fifth of the six states.

C. Organization

Ultrasist is a Mexican organization founded in 1994. It started as a software development organization, evolving during the last years into a service-oriented enterprise focused on Business Analysis, Enterprise Architecture, Security and Software Quality Assurance. In 2015, they renewed the CMMI-DEV appraisal and obtained CMMI-SVC, both at level 5.

The organization is constantly looking for process optimization and improvement; they carry out a weekly workshop where they discuss the ongoing work and analyze new proposals from the IT community. During such a workshop session, Ultrasist got to know ESSENCE and its ALPHAs through advice from a coauthor of this paper.

In October 2014, they started using ALPHAs for the sake of innovation. At that moment, ESSENCE was close to become an OMG standard, which happened one month later (November 2014).

In January 2015, the ALPHAs were already part of the organization's way of working. The first use of ALPHAs was to verify punctual quality attributes of work products and later on, to develop software quality assurance reviews.

In May 2015, the organization presented their use of ALPHAs as a part of software processes improvement when renewing CMMI level 5 appraisal.

The organization has a hierarchical structure (see Figure 3). The areas colored in green are those directly involved in the use of ALPHAs (the "mid-layer" of the organization):

- Internal SQA
- Sales, Marketing & Clients
- Business Analysis
- Software Construction
- SQA Specialized Services
- Project Management Office

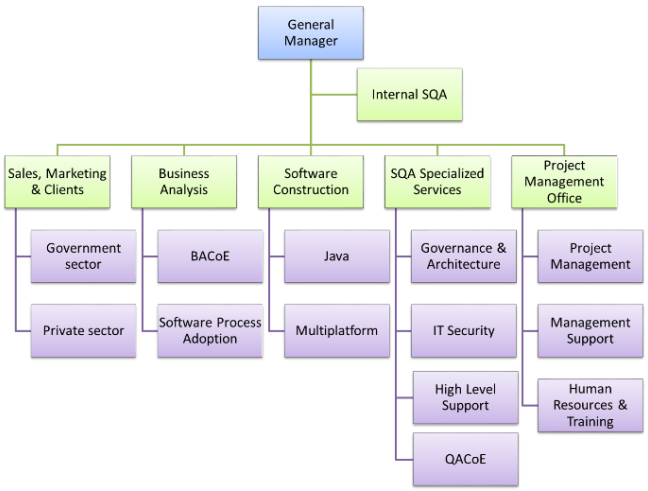


Fig. 3 Ultrasist org chart

One of improvement opportunities recognized by the organization was, for example, the quality of its work products in the Business Analysis area. Particularly, they were interested in a finer verification and validation process of Software Requirements Specification (SRS) work products.

In addition, the Internal SQA wanted to promote a better team integration and communication.

Another and even bigger concern arose after getting the CMMI appraisal, due to which the organizational processes suffered changes. Therefore, the General Manager needed to transmit the new processes to employees, especially new ones, and to generate a pocket guide for technical leaders and project managers.

The first step towards achieving this goal was to carry out a gap analysis versus their process and the actual way of working; next, the organization institutionalized the ALPHAs integrating them into its processes and in the working routine of the people.

Each ALPHA state was analyzed in order to associate organizational processes and the checklists items. When the association was established, the role in charge of the process became responsible for following the ALPHA states. Figure 4 shows the mapping between the organization's processes and the Team ALPHA.

TEAM	State	Responsible	Processes involved
	Seeded	SQA	• APS/Mgm
	Formed	SQA	• APS/Mgm • HR/Training • SQA
	Collaborating	SQA	• APS/Mgm
	Performing	SQA	• APS/Mgm
	Adjourned	SQA	• APS/Mgm

Fig. 4 Mapping between Team ALPHA and organization's process

The next section describes the actual usage of ALPHAs within the organization.

D. Methodology

The data were collected through direct interviews with the people involved in the initiative. Seven persons were interviewed individually in a face-to-face dynamic. Six interviewees lead their respective internal areas, while the seventh is the general manager of the organization.

The interviews were conducted in the following manner: (i) A set of questions on the topic of interest were designed; (ii) The questionnaire was sent to the interviewees; (iii) The interviews took place individually in the organization's facilities; (iv) Each interview was recorded and lasted in average for 20 minutes.

After the data were collected, the recorded interviews were analyzed and the fragments that were relevant for the purposes of the paper were transcribed. Later, the transcriptions were synthesized to obtain the observations listed in the discussion. This work was done by the first and second authors of this paper; this is to enrich the analysis and to moderate the threats to validity in this research.

In addition, the work products arising from the integration of ESSENCE and CMMI were analyzed in order to describe how both proposals, agile and traditional, coexist.

III. ALPHAS USAGE

First, the ALPHAs were translated from English into Spanish; some terms were modified according to the organization's customs and habits. Then, technical leaders got the ALPHA cards printed. They had to define which ALPHA state corresponds to which process or area.

The next step was to start using ALPHAs in a pilot group exclusively for verification and validation activities. Later, the rest of team members started to use the ALPHAs as a means of self-evaluation. Currently, the ALPHAs are being used in other activities and by other roles within the organization; however, this paper is based on the data obtained from the mid-layer roles.

The following paragraphs describe the actual uses of ALPHAs by different areas of the organization.

A. Internal SQA

Internal SQA processes are involved in almost all activities developed in the organization, so the SQA leader makes the most of the ALPHAs. He especially exploits the Requirements ALPHA that helps to trace requirements and verify their level of specification. During his activities, the SQA leader uses ALPHAs for:

- Checking the sufficiency of the work to be done.
- Determining if a work product is finished.
- Integrating agile practices into processes.
- Checkpoints during any time of a project.
- Tracing the requirements.

The Work ALPHA was helpful to determine if the people worked according to the organizational process, and they knew the organizational processes well.

The frequency of ALPHAs use varies depending on the process to monitor. If a SQA process is developed completely within the organization, the ALPHAs are used only during review phases, which happen at least every two weeks. However, if a SQA service is provided to clients, the ALPHAs are used almost daily.

B. Sales, marketing and clients

Sales area uses the Opportunity ALPHA for identifying clients' needs and as a support in creating requests for proposals.

Here, the ALPHAs resulted useful for reporting at what exact point the working team is. The mid-layer has found an adequate way of delivering valuable information of the project to higher levels of the organization who do not need to know all the details, but still need to know the exact progress of the project. Besides, they are especially useful when reporting progress to clients.

Moreover, using ALPHAs with clients helped to get a better understanding of the consequences of not providing a certain feature; the organization found a polite way of showing its clients cumulative effects of product absence as well as presenting missing quality attributes in terms of consequences and negative effects. So, it was possible not only to establish the existence/inexistence of a work product, but also what possible consequences it brought.

The ALPHAs became a favorable alternative for representing a state and a light way to report progress. For example, Figure 5 shows a radial graph that reports in a simple and accurate manner the state of each ALPHA in a particular moment. The ALPHA Opportunity is at state 4

(meaning that the opportunity is *Viable*), Stakeholder at state 3 (*Involved*), Requirements at state 3 (*Coherent*), Software system at state 1 (*Architecture selected*), Team at state 4 (*Performing*), Work at state 3 (*Started*) and, finally, Way of working at state 4 (*In place*).

Fig. 5 ALPHAs radial graph

The graph makes visible a general state of the project and allows work teams to decide which next state of an ALPHA to pursue.

C. Software construction

The JAVA leader guides Daily Scrums and evaluates the team progress using ALPHAs as checklists. The Software System ALPHA, in particular the *Architecture established* state, was took advantage of. He and his team also used other ALPHAs: Stakeholder, Requirements and Team.

D. Business analysis

The SRS Leader used the Stakeholder and Requirements ALPHAs, which were introduced gradually in their daily routine. First, the ALPHAs names were introduced to the work team. Later, when the ALPHAs became more familiar, the states of some ALPHAs were introduced. Thus, the complexity of adopting a new terminology was avoided. Now the SRS leader uses the states names as part of the everyday language when communicating with the team.

On the other hand, the ALPHAs were used to create a SQA Requirements guide (checkpoints for software requirements specification), which aims at improving the work products quality by making a more precise specification of requirements.

The ALPHAs states were grouped into levels of granularity corresponding to different target groups: the ALPHAs are related to the top management level, their states – to the project management, and checklists of work products – to developers. This partition helped the work team to understand ALPHAs and simplified their adoption.

An interesting fact is that the SRS leader actually uses the printed cards and prefers them in the original English version.

E. SQA specialized services

The IT Security technical leader uses the Team ALPHA as a checklist when he forms a team. He also uses the Stakeholder and Requirements ALPHAs.

Similar to the other areas, here the ALPHAs assist in controlling and improving the product quality and the work team's adherence to the process. He reported that he resorts to the ALPHAs, mostly at the moments of hesitation.

Besides that, the Governance Director uses the Opportunity ALPHA to discover client's needs and to understand the project's objective.

Another use of ALPHAs is to identify risks and to classify defects of a process or product. This, in turn, helps to evaluate the problem and to find a solution.

The Team and Software System ALPHAs contributed to know whether the people are working as a team and the communication is healthy; both are crucial factors for a successful project development.

F. How ALPHAs transformed some work products

The organization developed a series of new artifacts influenced by the ALPHAs. The first the SQA checklist that supports the SQA service provided by the organization to their clients. The SQA service usually consists in a reviewing and monitoring a particular business process of the client.

Initially this service used a checklist that consisted of quality attributes identified by the client, the organization, and more importantly, those defined in CMMI-ACQ. However, upon incorporating the ALPHAs internally, the organization observed that the ALPHAs states and checklists made the process lighter and more productive.

Derived from the observed benefits, Ultrasist decided to incorporate the ALPHAs checklist and states into the original document. In the first place, the quality attributes were classified under the scope of each ALPHA and were defined as "quality rules" in the language of the organization. These quality rules are based on the specific goals (SG) and specific practices (SP) of CMMI-ACQ process areas (PA).

For example, the Acquisition Requirements Development PA has SG3 Analyze and Validate Requirements – SP 3.2 Analyze requirements to balance stakeholder needs and constraints. It is mapped to the Requirements ALPHA, specifically to the *Bounded* state, see Figure 6, first row.

Using this rationale, when the team needs to verify the SP 3.2, it runs the checklist of the particular state, instead of reviewing all the subpractices established by CMMI-ACQ or consulting a work product. Importantly, the use of ALPHAs does not replace generation of the CMMI-ACQ related work products, but facilitates assessing the progress and health of a particular concern. Finally, this approach to ALPHAs use is independent from the reference model.

SQA service	CMMI-ACQ			Quality rule	Questions	Category (ALPHA)	Sub-category (ALPHA states)
Functional requirements	ARD	SG3	SP 3.2	The definition of the functionality and required quality attributes is established	• What is the definition of the architecture functionality and quality attributes?	3. Requirements	3.2. Bounded
Techincal Docs	TS	SG3	SP 3.2	The documentation to install, operate and maintain the system is developed	• What kind of supporting documentation is developed? • Are any standards used to generate the supporting docs? • How do you guarantee that the installation, operation and maintenance specifications will work out according to the plan? • What is the exact moment of generating those docs? • How are the generated docs checked?	4. Software system	4.2. Demonstrable
Project Management	RSKM	SG3	SP 3.2	Implement risk mitigation plans	• How often are the risks monitored? • In what way are the mitigation plans implemented?	6. Work	6.4. Under control

Fig. 6 SQA Checklist fragment

IV. Results

This section collects advantages and disadvantages of using ALPHAs. In addition, some advice from the participants involved in this experience, is listed.

A. Advantages and disadvantages

During the interviews, the practitioners expressed the following advantages of the use of ALPHAs:

- ALPHAs are easy to understand and apply, because *"they represent general concepts and put them in black and white"*.
- ALPHAs provide a common language, *"anybody in the work team can understand what you are talking about"*.
- ALPHAs are compatible with any process or lifecycle.
- ALPHAs colors help to associate them quickly to areas of concern, and the state-machine style makes the states flow clear.
- ALPHAs are useful for maintaining discussions focused and collecting clear arguments to make decisions, "*you can ask how the Opportunity is going and get clear answers*".
- ALPHAs help to accelerate convergence during meetings and discussions.
- ALPHAs work as communication facilitators between members of the organization and with the client.

Several disadvantages were also pointed out:

- ALPHAs have too many states; sometimes the cards sets are not handy.
- Some terminology needs to be adapted to the particular context of the organization, for example the *in-place* state or the *Endeavor* area of concern.
- Some states and checklists may have a different interpretation between team members, mainly between juniors and seniors. *"Experience in software development is needed in order to uniform interpretations"*.
- Compared to the mid-layer, the operational layer of the organization required more time to understand ALPHAs.
- ALPHA is not a *universal* term neither a common software engineering word.
- ALPHAs resulted of little help in bigger projects; their simplicity became an obstacle. For example, how to manage many stakeholders or how to evaluate big quantities of requirements with ALPHAs?

B. Threats to validity

The following threats to validity to this particular research were detected.

The internal validity:

- The trustworthiness of the survey responses can be considered a major issue since one of the authors is the general manager of the organization. It is possible that some negative issues were not completely expressed by the interviewees, which may explain the fact the very few disadvantages of the ALPHAs use were detected.
- The coverage of roles in the organization may be considered a threat as well. Only members of the mid-layer of the organization were interviewed, which could affect the perspective of the benefits and drawbacks of the ALPHAs as compared to the point of view of the rest of the organization.
- On the other hand, the number of interviewees that constitutes the sample is representative of the target group who used ALPHAs.

The external validity:

- The organization was not intentionally selected; they took the initiative to start to use ALPHAs and, then, to share their experience.

As for reliability, it should be mentioned that the second and the third authors of the paper has been involved in the organization for a long time (years) and are fully capable to understand the needs of the organization.

Also, the first author, who conducted the interviews, had worked in the organization and developed a trusting relationship with the participants. This fact minimizes the surveys' trustworthiness threat.

After considering the above mentioned threats, we conclude that the results of this research were not affected; however, the data collection methodology can be improved.

Finally, in order to discuss the results and findings from the interviews, peer debriefing took place, which is an extra point to the validity of the research.

C. Some advice

This experience left some useful lessons learned that are summarized in the following paragraphs.

Avoid implementing all the ALPHAs at the same time. Use the ALPHA you need and then add them one by one depending on what you need in a particular moment. Besides, let the team decide which ALPHAs they want or need to use. *"Take the best (that fits for you) of each world and apply it in your context"*. As [20] establishes, an organization has to understand that the organization itself cannot be agile, but its employees can be.

Actually, in the organization's case, the Way-of-Working ALPHA was used occasionally or almost never because they have a well-defined and mature process; in the words of the general manager: *"Everybody knows what to do and how to do it"*.

Do not be afraid to modify ALPHAs, you won't break it down. Some ALPHAs were adapted to become more familiar for the work teams, for example, the Way-of-Working ALPHA was renamed as Working-Methodology. In fact, the ALPHAs should be adapted to the organization's language in order to fit in its process, however minimal these adaptations are.

Do not see ALPHAs as a sequential set of steps, they are not a process. In many cases they were confused with a sequential method. The ALPHAs were created as a tool to be applied at any moment during a project.

It is not necessary to read the whole standard to understand ALPHAs. Actually, only one practitioner read the whole OMG document and the rest consulted section 8.1 Overview to be able to get ALPHAs going. To be more specific, apart from section 8.1 of the standard, in which ALPHAs are presented, it is advisable to read sections 8.2 through 8.4 that provide detailed descriptions of each area of concern, their related ALPHAs, states and checklists.

Try ALPHAs in the presentation you feel comfortable with. Using the printed cards or the electronic version turned out to be a discussion topic. Some participants consider having to carry all the cards a disadvantage; others believe the opposite and prefer the printed version because *"cards are like a cheat sheet to keep quality monitored"*.

V. Conclusions

ALPHAs brought benefits and improvement even for a mature and solid organization that already had its ways of working at a high level. They represented an innovation factor in order to renew the level 5 in CMMI-DEV and CMMI-SVC. Importantly, the quality of the process and the product was systematically improved with the ALPHAs states guide. ALPHAs are not technical-oriented; they are focused on serving as control points and checklists for the teams.

It was shown that ALPHAs can be integrated with other standards, and their independence is a plus when an organization decides to integrate them to the actual way of working and harmonize the whole process. According to [10] harmonizing processes allow organizations to improve, mature, acquire and institutionalize best practices and management systems from multiple approaches.

On the other hand, there are inconsistencies related to the initial objective of ESSENCE: provide practitioners with the means of describing their methods and practices [11]. This issue was addressed by creating activity spaces "descriptions of the challenges a team faces when developing, maintaining, and supporting software systems" [8]. However, in case that the organization does not execute a formal process, the activity spaces make little sense; on the contrary, if an organization possesses a well-defined process, the activity spaces are not necessary, and in the case of Ultrasist, were not used. Instead of supporting the definition of methods and practices, ALPHAs were used to guide the team's way of working. This issue affects small organizations since the simplicity and flexibility of ALPHAs require a well-predefined way-of-working.

Nowadays, the ESSENCE standard is being widely applied in many countries and with diverse objectives. However, the Latin American context features far less impact of this standard in the industry. We believe that the experience described in this paper will motivate Latin American software development industry to work with ALPHAs.

As future work, three lines are identified: (i) to establish patterns, like *usage scenarios*, of when, how and what particular problem ALPHA(s) could solve; and (ii) to determine how the ALPHAs overlap with other standards in order to define a harmonized multi-model process and not implement each one separately. Lastly, (iii) the organization under discussion started to explore how the ALPHAs could be integrated into enterprise architecture services.

Acknowledgment

The authors thank Ultrasist work team leaders: Ernesto Hernández Uribe, Ricardo Rodríguez López, Félix de la O Díaz, Alejandro Ramírez Ramos and Gibrán Granados Paredes.

This work has been developed under the Postdoctoral Fellowships Program of the General Directorate of the

Academic Staff (DGAPA) of the National Autonomous University of Mexico (UNAM).

REFERENCES

[1] B. Erbas and C. Erbas, "On a theory of software engineering: A proposal based on transaction cost economics". In SEMAT Workshop on General Theory of Software Engineering (GTSE'13), San Francisco, CA, USA. pp. 15–18, DOI: 10.1109/GTSE.2013.6613864 (2013)

[2] A. Kocatas and C. Erbas, "Extending Essence Kernel to Enact Practices at the Level of Software Modules". In SEMAT Workshop on General Theory of Software Engineering (GTSE'14), Hyderabad, India. pp. 32–35, DOI: 10.1145/2593752.2593758 (2014)

[3] Capability Maturity Model Integration (CMMI). Software Engineering Institute, Pittsburgh, PA, USA (2010)

[4]ISO/IEC 12207: Systems and software engineering – Software life cycle processes. ISO/IEC, Geneva, Switzerland (2008)

[5] A Guide to the Project Management Body of Knowledge (PMBOK Guides). Project Management Institute, Piscataway, NJ, USA (2013)

[6] K. Schwaber and J. Sutherland, "The scrum guide – the definitive guide to scrum: The rules of the game". http://www.scrumguides.org/ (Accessed 08/05/2016)

[7] S. Shingo, "A study of the Toyota production system: From an Industrial Engineering Viewpoint". Productivity Press (1989)

[8] ESSENCE – Kernel and language for software engineering methods. Object Management Group, Needham, MA, USA (2014)

[9] A. Kaczorowska, "Traditional and Agile Project Management in Public Sector and ICT". In proceedings of the 2015 Federated Conference on Computer Science and Information Systems pp. 1521–1531, DOI: 10.15439/2015F279 (2015)

[10] C. Pardo, F. García, M. Piattini, F. Pino and T. Baldassarre, "A 360-degree process improvement approach based on multiple models". Revista Facultad de Ingeniería, Universidad de Antioquia, No. 77, pp. 95–104, DOI: 10.17533 /udea.redin.n77a12 (2015)

[11] P.-W. Ng, "Theory Based Software Engineering with the SEMAT Kernel: Preliminary Investigation and Experiences". In SEMAT Workshop on General Theory of Software Engineering (GTSE'14), Hyderabad, India. pp. 13–20, DOI: 10.1145/2593752.2593756 (2014)

[12] C. Preraire and T. Sedano, "Essence Reflection Meetings: Field Study". In International Conference on Evaluation and Assessment in Software Engineering (EASE '14), London, England, United Kingdom DOI:10.1145/2601248.2601296 (2014)

[13] CMMI for Development, Version 1.3 (CMMI-DEV). Software Engineering Institute, Pittsburgh, PA, USA (2010)

[14] CMMI for Acquisition, Version 1.3 (CMMI-ACQ). Software Engineering Institute, Pittsburgh, PA, USA (2010)

[15] CMMI for Services, Version 1.3 (CMMI-SVC). Software Engineering Institute, Pittsburgh, PA, USA (2010)

[16] I. Jacobson, P.-W. Ng, P. McMahon, I. Spence and S. Lidman, "The Essence of Software Engineering: The SEMAT Kernel". ACM queue, Vol 10, No. 10, DOI: 10.1145/2380656.2380670 (2012)

[17] I. Jacobson, P.-W. Ng, P. McMahon, I. Spence, and S. Lidman, "The Essence of Software Engineering". Addison Wesley (2013)

[18] J. Park, P. McMahon and B. Myburgh, "Scrum Powered by Essence". ACM SIGSOFT Software Engineering Notes. Vol. 41, No. 1, DOI: 10.1145/2853073.2853088 (2016)

[19] J. Park, "Essence-Based, Goal-Driven Adaptive Software Engineering". In SEMAT Workshop on General Theory of Software Engineering (GTSE'15), Austin, TX, USA. pp. 33–38, DOI: 10.1109/GTSE.2015.12 (2015)

[20] R. Wendler, "Development of the Organizational Agility Maturity Model". In proceedings of the 2014 Federated Conference on Computer Science and Information Systems pp. 1197–1206, DOI: 10.15439/2014F79 (2014).

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1539–1543
DOI: 10.15439/2016F509
ACSIS, Vol. 8. ISSN 2300-5963

Adopting collaborative games into Open Kanban

Adam Przybyłek, Marcin K. Olszewski
Gdansk University of Technology, Faculty of Electronics, Telecommunications and Informatics
Narutowicza 11/12, 80-233 Gdańsk, Poland
Email: adam.przybylek@gmail.com, marolszak@vp.pl

***Abstract*—The crucial element of any agile project is people. Not surprisingly, principles and values such as "Respect for people", "Communication and Collaboration", "Lead using a team approach", and "Learn and improve continuously" are an integral part of Open Kanban. However, Open Kanban has not provided any tools or techniques to aid the human side of software development. Moreover, as a Lean initiative, it is not as comprehensively defined process as Scrum or XP. Accordingly, inexperienced Kanban teams may feel a bit lost. To deal with these challenges, we propose an extension to Open Kanban, which contains a set of 12 collaborative games. The feedback received from three Kanban teams who leveraged our extension in commercial projects, indicates that the adopted games improved participants' communication, commitment, motivation and creativity.**

I. Introduction

In more recent years, the software industry has started to look at Lean as a new approach that could complement Agile Software Development [16]. Lean is a general term for finding ways to eliminate waste and increase efficiency [13]. One of the agile methodologies that adds a vast Lean heritage is Open Kanban [11]. The distinguished characteristics of Open Kanban are: (1) visualization of the workflow with Kanban board, (2) limitation of the work in progress (WIP), and (3) measurement of the lead time [15]. The motivation behind visualization and limiting WIP was to identify the constraints of the process and let each member of a team focus on a single item at a time [1]. This technique promotes the pull approach, which means that the team "pull" work when they are ready, rather than having it "pushed" in from the outside [14].

In contrast to other agile methodologies, Open Kanban leaves almost everything open. It does not prescribe iterations, it does not define roles or meetings, and finally, it does not contain process artifacts [14], [15]. Moreover, although Open Kanban emphasizes the human factor of software development and its founding declaration states that "without teamwork Kanban fails" [11], it does not provide any tools or techniques to aid the human side of software development. Accordingly, inexperienced Kanban teams may feel a bit lost.

Fortunately, Open Kanban can be extended by organizations that wish to create an Agile and Lean method that is customized for their particular audience. We took this opportunity to equip our teams with a set of collaborative games, structured as an extension to Open Kanban. Our research was inspired by the ActiveAction workshop [17], which combines classical and game-based techniques to foster stakeholders' involvement and collaborative identification of objectives and risks.

II. Research Method

Our study was conducted as Action Research (AR). In AR, the researcher works in close collaboration with a group of practitioners, acting as a facilitator, to solve a real-world problem while simultaneously studying the experience of solving the problem [5]. The researcher brings his knowledge of action research while the participants bring their practical knowledge and context [3]. The goal of AR is to improve practical matters as well as to improve scientific knowledge [3]. A precondition for action research is to have a problem owner willing to collaborate to both identify a problem, and engage in an effort to solve it. The problem owner in this research was a software development department of a world wide aviation IT provider (the company wishes to remain anonymous). The department experienced typical challenges of adopting a new methodology (i.e. Open Kanban). Its authorities were open to new ideas and willing to implement collaborative games. Three teams that participated in our research are presented in Table I.

Table I.
Participating teams

Team	Comments
T1, 6-8 people	The multicultural team of junior developers who provided services for external customers. All team members were familiar with Open Kanban.
T2, 8-10 people	The team of developers and testers guided by an agile coach. They developed solutions for internal departments. All team members had over 6 months of experience with agile development, but they got started with Open Kanban.
T3, 6-8 people	The distributed team of instructors with an extensive experience in programming or project management. They worked on a project designed to train engineers within the company. All team members were experienced with Agile practices, but they got started with Open Kanban.

III. Identification of Deficiencies

Action Research assumes that theory and practice can be closely integrated by learning from the results of intervention

that is planned after a thorough diagnosis of the problem context [5]. To identify deficiencies in the adoption of Open Kanban, we prepared a survey that contained a total of 13 questions (Fig. 1). Respondents rated, on a Likert scale of 5 points, their degree of agreement regarding the implementation of Kanban values and practices in their teams. In total, 18 respondents from 3 teams (T1, T2, T3) completed the survey. The respondents were also asked to provide their free comments on the usage of Open Kanban by their organization. The survey was anonymous, so we assumed that the responses were honest.

It is not accidental that "Plan-Do-Check-Act cycle", "Systematic approach to improvement", and "Collecting feedback on the process" received the lowest rates. Open Kanban does not define retrospective meetings, nor does it prescribe timeboxed iterations. Thereby, the investigated teams had practiced only occasional retrospectives.

In the open-ended comment question a few respondents mentioned that their team had problems with complex work items. Indeed, in Open Kanban, no particular item size is prescribed. Since there is no requirement to break down items so they are small enough to fit into a specific time box, the management of the workflow is cumbersome. The results of the survey also suggest that the respondents were familiar with Open Kanban, but their knowledge was incomplete (most of them were neutral on "Understanding of Kanban mechanisms"). Besides, "Work-In-Progress limits" scored slightly above 0, so the limits probably needed adjustment. Furthermore, the detailed results show that only half of the respondents reported communication between team members as satisfactory.

IV. Our extension to Open Kanban

For each deficiency identified in the previous section, we suggest collaborative games that might be a remedy for it. Collaborative games refer to several structured techniques inspired by game play and designed to facilitate collaboration, foster customer involvement, and stimulate creative thinking [12]. Fig. 1 presents the mapping between the problematic issues and Open Kanban principles with collaborative games superimposed. The following subsections explain how we intend to enrich Open Kanban by our extension.

A. Visualize the workflow

Although the teams used a Kanban Board to visualize work, different work item types, and WIP limits, we found room for improvement. Work items were not estimated. As a consequence, it was difficult to manage the workflow and make commitments. Therefore, we set the rule that a work item could not make its way from the backlog onto the board until it had been estimated with Planning poker [9].

B. Learn and improve continuously

Before our research went into work, every time someone saw an issue which seemed worth reviewing, the team started an ad-hoc kaizen meeting. However, Kanban suggests to make incremental improvements to the existing processes at regular intervals called cadences. Accordingly, during our research we chose a four weeks cadence for retrospectives. A key element of a retrospective is that the team must agree, together, to trust each other and to believe that every comment or suggestion is intended for the sole purpose of improving the team's performance [12]. We expected that collaborative games would make the team feel safe to discuss any issue that concerns them.

The Snake Game stimulates memories and helps the team to gather data from many perspectives [6]. Its objective is to create a shared picture of what has happened since the last retrospective. Participants write sticky notes to represent memorable, personally meaningful events and then post them in chronological order on a large poster of a snake. The more recent the event is, the closer to the head it should be posted. The collected notes can constitute an input to other games for retrospectives.

The Perfection Game [www.mccarthyshow.com/online] is a tool for continuous improvement of the process, team and organization. With this game team members are invited to participate in improvements as they give feedback to each other. To get feedback, team members are asked to: rate (on a scale of 1 to 10) the action, process, item or event being considered; state what they like; suggest what to do to make it perfect. Participants can only withhold points if they provide suggestions to improve the considered issue. If they cannot say how to make it better, the default score is a 10. If participants give a high rating then they have to state what went good, what makes it so good, where does the value come from, etc. Since participants have to motivate their ratings (the rating is coupled to what participants like about it and what they think can be done to do it better), the quality of the feedback is improved.

The Coaching Cards Game uses a deck of colorful cards with various images to represent team members' feelings. At the beginning of each round, each participant chooses one card that illustrates his feelings related to the event being considered. Then, the team discusses their feelings. The game creates a non-threatening opportunity to gather data about feelings during the last release cycle by connecting the feelings to events that happened in the cycle [7]. Even though someone does not want to express his opinion directly, the game allows the team to gather the opinion indirectly. With this game, the team can identify events that provide benefits and events that cause problems.

The Sailboat Game [7] lets a team think about their impediments, risks, good practices, and where they go as a team. The game starts by drawing a sailboat, rocks, wind, and an island. The island represents the team's objectives/vision. The rocks represent the risks the team

might encounter along the way. The anchor on the boat is everything that slows them down on their journey. The wind represent everything that helps them to reach their objectives [7]. Next, participants write ideas on sticky notes and then post the ideas into the different areas according to the picture. Then, the team discusses how to continue the practices that are written on the clouds/wind area, how to mitigate the identified risks, and what actions can be taken to fix the problems.

C. Limit Work-in-Progress

Open Kanban advices limiting the WIP so as to optimize the workflow of the system in accordance with its capacity [11]. A limit on WIP constrains how many work items can be in each workflow step at a time [2]. Limiting the WIP has two major benefits: it reduces the time it takes to get any one thing done (lead time); and it improves quality by giving greater focus to fewer tasks [13]. We suggest two games that demonstrate this principle to new teams.

The Ball Flow Game [availagility.co.uk/2010/11/17/the-ball-flow-game]. The aim of this game is to pass as many balls as possible through the team in 2 minutes. However, the activity is constrained by the following rules: (1) balls cannot be passed to a direct neighbor (the team arrange themselves in a circle); (2) each ball must be touched at least once by every player; (3) each ball must have air-time as it is passed between players; (4) each ball must return to the same player who introduced it into the system; and (5) if a ball drops, it cannot be picked up. The game is played a total of 3 times with 1-minute breaks in between to inspect and adapt the process. Before each round, the team estimates of how many balls they can pass through the system. In addition, the team has two minutes of preparation time for the first round to self organize and discuss the strategy. The game demonstrates that every system has a natural velocity and to improve the system significantly, it is often not a case of working harder, but a case of changing the process. Players will find out that when balls are pushed into the system, it results in dropped balls and decreases performance. Therefore, they will arrange a pull system – i.e. a system where the balls are not passed until the downstream player is ready.

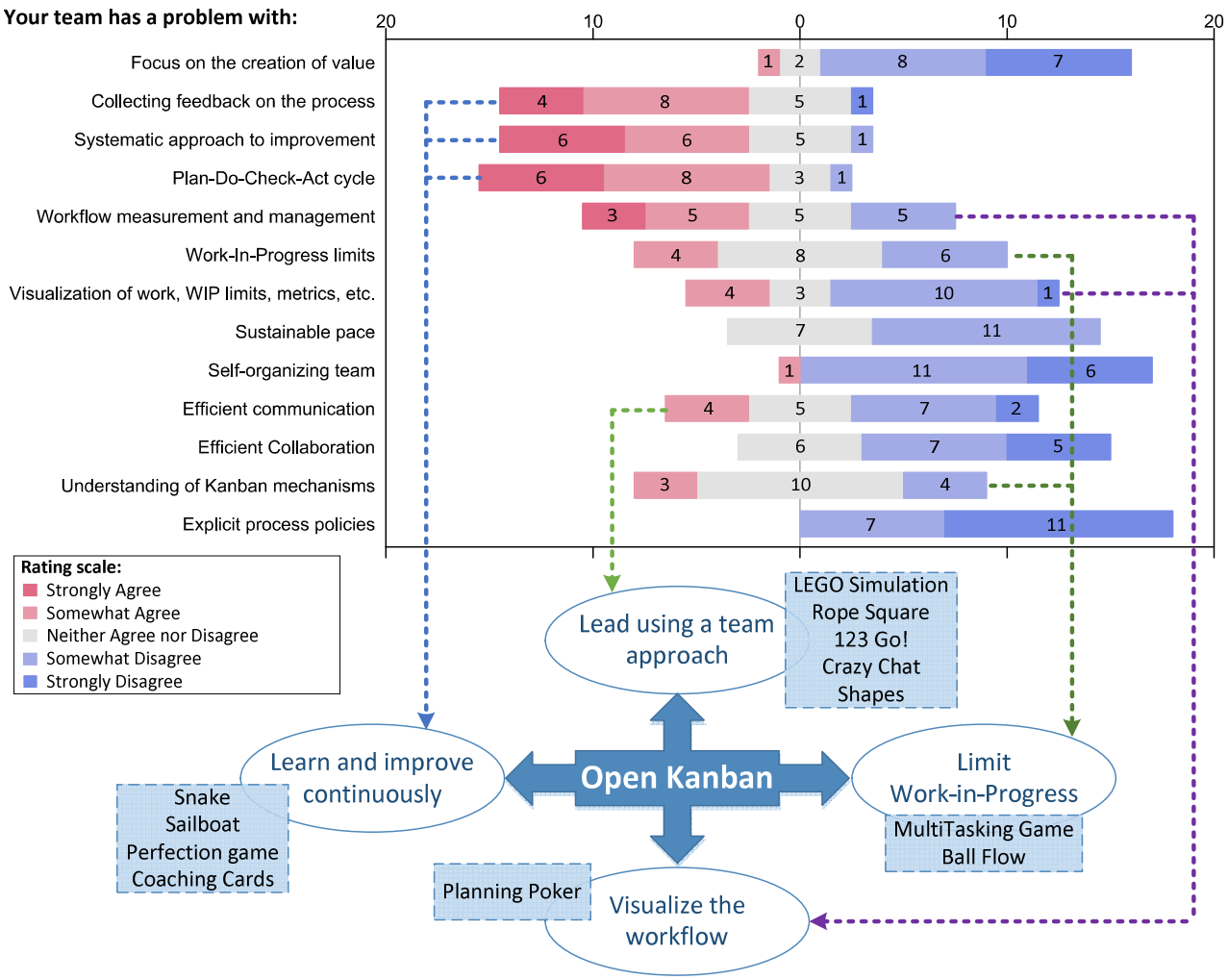


Fig. 1 Summary of questionnaire responses and the proposed extension

The Number Multitasking Game [10] illustrates the effect of multi-tasking and context switching. It consists of two rounds. In each round, each participant is given a page divided into three columns. The task is to fill out the left column with the Roman numerals I through X, the middle column with the letters A through J, and the right column with the Arabic numerals 1 through 10. In the first round, the participants are asked to write row by row, while in the second round, they are asked to work column by column. The game shows why limited WIP improves lead times and lets developers understand Kanban better.

D. Lead using a team approach

Building successful teams and team leadership are crucial to deliver value. At the center of teamwork are effective collaboration and communication. We suggest playing Rope Square and LEGO Simulation as a means of team building and a way to socialize. Besides, these games entail collaboration in decision-making. Other games in this section emphasize the importance of communication.

The Rope Square Game [4] is an icebreaker for team building. The team has to form a perfect square with the rope. Each member must be blindfold and grab the rope with both hands. When the team feels like they are finished and all agree that the rope is in a perfect square, they put the rope to the ground. The game reveals how people fulfill different roles in a group. It also allows the team to build cooperation and trust between members.

LEGO Simulation. The team's task is to build a space base on Mars with LEGO bricks. The facilitator plays the role of the Product Owner. He writes stories, is available to answer questions and to provide feedback. The game encourages the team to communicate with the Product Owner, work in cooperation and reach a consensus.

The 123-go! Game. In this game [8], the facilitator explains that after he counts to three and says "go", everyone should clap their hands. The task for the facilitator is to count to three slowly then clap his hands, pause for a second and say "go". Usually, people will clap when the facilitator claps, rather than when he says "go". This game emphasizes the importance of listening before acting when working in a team.

The Crazy Chat Game teaches players to be more aware of the importance of paying attention. The team split themselves into pairs. Then, for one minute one person talks about something he is most passionate about in life. The other person has to act as disinterested as possible. After a minute they should switch roles.

The Shapes Game. This game teaches that the way we communicate impacts our ability to succeed. The team has to form pairs. Each pair decides who will draw and who will instruct. The instructor is given a picture with shapes. The drawer can not see the shapes or ask any questions. The instructor has to describe the picture giving only verbal instructions, while the drawer has to reproduce the original shapes. The instructor can see what the drawer is doing and provide feedback.

V. EVALUATION AND RESULTS

The evaluation took place in 2015 and 2016. All games presented in Section IV were implemented in both the T1 and T2 teams. In addition, Planning Poker, Coaching Cards, Perfection Game, Rope square, and Tennis Balls were implemented in the T3 team. The implementation of each game was planned with Agile Coach, Service Leader, or Project Manager.

After each game session, we issued a questionnaire. The participants were asked to indicate their level of agreement with statements about game-playing activity. The responses were on a Likert scale of 1 to 5 (Table II). At the end of the survey, the participants were also invited to specify any additional remarks. We used two different sets of questions – one for the retrospective games and Planning Poker (games A), the other for the remaining games (games B). For each game and question, we first took the average per team, and then the average of the averages (the detailed charts are available on http://przybylek.wzr.pl/FedCSIS16). The differences in averages between teams were always less than 1 point, except for Coaching Cards and LEGO Simulation.

Games A aim to directly support teams in their work and were perceived positively. Note, that Planning Poker got a low score in "fostering creativity", but it is not a downside because this game has different objectives. All participants appreciated this game and wanted to use it on a regular basis. Among the retrospective games, Sailboat performed the best but should be used interchangeably with the Perfection game as suggested by a few respondents. Indeed, teams should have a set of possible games to avoid monotony that leads to fatigue and a lack of motivation. In turn, Coaching Cards was not appreciated by the T1 team. They rated the game lower than both other teams and commented that they did not feel comfortable enough to share their problems, opinions and concerns. In addition, someone mentioned that the results strongly depend on the openness and honesty of participants. Surprisingly, only 2 persons in both the T1 and T2 teams wanted to use the Snake game in the future, even though the game obtained high scores in improving commitment and communication.

Games B make teams aware of some of the key Kanban values or mechanisms and generally should be played only once by each team. According to the comments received, these games should be used during training sessions. Note, that the 123-Go! and Crazy Chat game were carried out during one meeting and jointly evaluated due to their common purpose. Although some games received low scores in certain aspects, this is not a downside because they still meet their objectives. For instance, everyone strongly agreed that Shapes, 123-Go! and Crazy Chat revealed the importance of effective communication. In turn, Number Multitasking demonstrated the cost of context switching.

TABLE II.
SUMMARY RESULTS

Rating scale: 1 (Strongly disagree) 2 (Disagree) 3 (Neutral) 4 (Agree) 5 (Strongly Agree) The game:	**Games A**					**Games B**					
	Planning Poker	**Snake Game**	**Perfection Game**	**Coaching Cards**	**Sailboat Game**	**Ball Flow**	**Number Multitasking**	**Rope square**	**LEGO Simulation**	**123 Go!/Crazy Chat**	**Shapes Game**
- produces better results than the standard approach	4,6	3,2	4,7	3,5	4,6						
- should be implemented permanently instead of the standard approach	4,2	3,0	4,7	2,5	4,7						
- may be considered as complementary to the standard approach	2,7	3,2	4,7	4,3	4,4						
- fosters participants' creativity	2,4	4,0	3,5	4,5	4,3						
- fosters participants' commitment and motivation	4,3	4,5	4,0	3,2	4,6						
- improves participants' communication	4,8	4,6	3,0	3,5	4,7						
- is easy to understand and play	4,9	5,0	4,9	5,0	5,0	4,8	5,0	4,3	4,2	4,8	5,0
- makes participants aware of the importance of effective communication						4,9	3,4	4,7	4,6	5,0	5,0
- makes participants aware of the importance of team collaboration						5,0	2,7	5,0	4,9	2,4	2,7
- allows participants to better understand Kanban mechanisms						4,1	4,4	3,1	4,7	3,0	2,9

VI. CONCLUSIONS

This paper reports on an Action Research project designed to explore the ways in which collaborative games could benefit Kanban teams. We started by carrying out a survey among three Kanban teams of a world wide aviation IT provider in order to identify deficiencies in their current practices. Each problematic issue was mapped to the affected Open Kanban principle. Based on the survey results, we proposed an extension to Open Kanban, which specifies 12 collaborative games, divided into four categories in compliance with four Open Kanban principles.

The feasibility of our extension was evaluated by three Kanban teams with encouraging results. We found that the adopted games: (1) improved participants' communication, commitment, motivation and creativity; (2) helped the teams understand the main mechanism, values or practices of Open Kanban; (3) produced better results than the standard approach; and (4) were easy to understand and play. Moreover, the teams intended to continue playing the games after the project finished. We hope that the reported experience will encourage other practitioners to implement collaborative games in their projects. Future studies may consider examining other collaborative games or adopting games into other software development processes.

REFERENCES

[1] Ahmad, M.O., Markkula, J., Oivo, M.: Kanban for software engineering teaching in a software factory learning environment. In: World Transactions on Engineering and Technology Education vol. 12(3), 2014, http://dx.doi.org/10.1109/EDUCON.2014.6826129

[2] Anderson D.J.: Kanban: Successful Evolutionary Change for Your Technology Business. Blue Hole Press, 2010

[3] Baskerville, R., Myers, M.D.: Special issue on action research in information systems: making IS research relevant to practice—foreword. In: MIS Quart 28(3), pp. 329–335, 2004

[4] Davison, P.: Group Warmup and Team Building Activities. 2009

[5] Davison, R.M., Martinsons, M.G., Kock, N.: Principles of Canonical Action Research. In: Inf. Syst. J. 14(1), pp. 65-86, 2004, doi: 10.1111/j.1365-2575.2004.00162.x

[6] Derby, E., Larsen. D.: Agile Retrospectives: Making Good Teams Great. Pragmatic Programmers, 2006

[7] Gonçalves, L., Linders, B.: Getting Value out of Agile Retrospectives: A Toolbox of Retrospective Exercises. Leanpub, 2014

[8] Greaves, K., Laing, S.: Collaboration Games from the Growing Agile Toolbox. Leanpub, 2014

[9] Grenning, J.: Planning Poker or How to avoid analysis paralysis while release planning. In: Renaissance Software Consulting, vol. 3, 2002

[10] Hammarberg, M., Sunden, J.: Kanban in Action. Manning Publications, 2014

[11] Hurtado, J.: Open Kanban. github.com/agilelion/Open-Kanban, 2013

[12] International Institute of Business Analysis (IIBA): Agile Extension to the BABOK®Guide. Toronto, Canada, 2013

[13] Klipp, P.: Getting Started with Kanban. Kanbanery, 2011

[14] Kniberg, H., Skarin, M.: Kanban and Scrum: making the most of both. C4Media Inc, 2010

[15] Nikitina, N., Kajko-Mattsson, M., Stråle, M.: From scrum to scrumban: a case study of a process transition. In: International Conference on Software and System Process, Zurich, Switzerland, 2012, http://dx.doi.org/10.1109/ICSSP.2012.6225959

[16] Rodriguez, P., Markkula, J., Oivo, M., Turula, K.: Survey on agile and lean usage in finnish software industry. In: ACM-IEEE International Symposium on Empirical Software Engineering and Measurement, Lund, Sweden, 2012, doi: 10.1145/2372251.2372275

[17] Trujillo, M.M., Oktaba, H., González, J.C.: Improving Software Projects Inception Phase Using Games: ActiveAction Workshop. In: 9th International Conference on Evaluation of Novel Approaches to Software Engineering (ENASE'14), Lisbon, Portugal, 2014

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1545–1548
DOI: 10.15439/2016F212
ACSIS, Vol. 8. ISSN 2300-5963

Towards the participant observation of emotions in software development teams

Michal R. Wrobel
Faculty of Electronics, Telecommunications and Informatics,
Gdansk University of Technology, Poland
Email: wrobel@eti.pg.gda.pl

***Abstract*—Emotions, moods and temperament influence our behaviour in every aspect of life. Until now plenty of research has been conducted and many theories have been proposed to explain the role of emotions within the working environment. However, in the field of software engineering, interest in the role of human factors in the process of software development is relatively new. In the paper the research design process that has been proposed details a novel approach using the method of participant observation in order to investigate the role of emotions in IT projects. The participant observation protocol presented was developed in an iterative manner, where successive phases were validated in two preliminary studies.**

I. Introduction

COMPUTER programs are used in almost every aspect of our lives. At the same time, a considerable part of IT projects fail. Budgets or schedules are exceeded, applications are low quality. The Standish Group estimated in their "2015 CHAOS Report" that 19% of the IT projects failed, and only 29% were entirely successful [1]. Available techniques and methods of software engineering solve the problems only partially. One of the directions of research that can improve the success factors of an IT project is the analysis of human factors in the process of software development. Emotions, moods and temperament influence our behaviour in every aspect of life. Such impact is sometimes more, sometimes less significant. Modern research of emotions in the workplace were initiated in 1983 by Hochschild's book "The Managed Heart". Since then, numerous research has been conducted and many theories have been proposed so as to explain the role of the emotions in work [2].

However, in the field of software engineering interest in the role of human factors in the process of software development is relatively new. Software development is considered as one of the most complicated tasks performed by humans, not only because of the technical complexity, but also due to human factors [3]. Nowadays, software development require a high level of cooperation between a number of individuals with different personalities, which may raise interpersonal conflicts. While the technical and management issues have been covered in numerous studies, the awareness of importance of human aspects is still minimal.

This low awareness of the role of emotions in the software engineering is partly due to the small number of research in this field. In Section II studies on emotions in the software development process are presented. In most studies quantitative methods were used, however, these methods allow confirmation of the hypothesis, or more accurate data on the phenomenon to be obtained, which, at least in part, is understandable. Whereas knowledge about the impact and the role of emotions in the software development is still negligible. To fill this gap, I propose to conduct experiments using the method of participant observation, taken from ethnography science. In Section III the experiment design is proposed. To confirm the validity of the assumptions and design, preliminary studies were conducted. Information about its execution and results are presented in Section IV.

With the current state of knowledge concerning the role of emotions in the work of IT professionals, it is necessary to carry out further detailed observations. The application of the proposed and validated participant observation study should allow a better insight on the impact of emotions in the software development process.

II. Studies on emotions of software developers

Credible studies on the role of emotions in the software development process started only in the XXI century. However, so far they have been conducted mainly without immersion in the work of software developers. Graziotin et al. compared self-assessed affective states with self-assessed productivity of the programmers. Their study revealed correlation between productivity and both valence and dominance [4]. A similar study was conducted by Khan et al. In the experiment, participants watched short video clips that were meant to affect their mood. Before conducting programming and debugging tasks, they assessed their emotional state. Results showed weak correlation between arousal and productivity [5].

Recently a few studies on emotions of software developers were conducted using sentiment analysis technique. This is a completely passive method of gathering information regarding the attitude of the authors based on the analysis of their texts. Garcia et al. performed such study on the basis of artifacts generated by the Gentoo Community, such as mailing lists and bug tracker. Results of their experiment showed a correlation between developers' emotional states and activity [6]. A similar experiment, performed by Muriga et al. showed that only a small subset of emotions can be detected based on data mining techniques [7].

The common element among all this research is the lack of direct contact with respondents during data collection. Collecting such confidential and ambiguous data, in the form of information concerning emotions, require gaining the trust of the respondent. Another issue is the problem with the naming of emotions. People ambiguously interpret names of emotions, which leads to biased data.

A novel approach to recognizing the emotions of software developers was proposed by Muller et al. In two experiments, they used biofeedback devices to gather EEG, EDA, fMRI and eye-tracker signals of developers during their work [8]. Biometric signals provide objective information about reactions of the human body to emotional stimuli. Nevertheless, current knowledge does not allow unambiguous assignment of emotional states on the basis of this data [9].

Another approach was proposed by Graziotin et al. During their study, a researcher constantly supervised the experiment and respondents were observed while programming. Before and after the development task, a face to face interview was performed [10]. Such approach allows more accurate observation of the behaviour of developers, and thus better understanding of the impact of emotions upon their work.

III. The research design

To gather significant insights in the role of emotions within the software development process, I propose to use a participant observation method, known from ethnography science. Review of the emotion management literature performed by Fisher et al. showed usefulness of participant observation and interviews in examining emotions in the workplace [2].

Observation is a research method that allows detail qualitative information to be gathered. It is defined as systematic description of the events and behaviours in environment [11]. A particular type of this method is participant observation, where the researcher takes an active part in daily events and activities of the group and interacts with its members [12]. Seaman et al. confirmed that participant observation allows to obtain comprehensive data about software development process [13].

For the purpose of the study, in order to obtain information about emotions and moods of the developers and the impact of these factors on their work, participant as an observer method was chosen. It means that the observer is a member of the group and she or he takes full part in its activities. Other members of the group are aware of being the subject of the experiment. The observer during work can collect data about the colleagues, but also observe her or his own behaviour and the factors that influence them. This permits deeper, though subjective, behavioural analysis. In order to eliminate subjective factors it is planned to make a number of observations by independent researchers.

Research is supposed to be fully overt. Employers of the researchers should provide formal agreement to place studies in their facilities. Also, team members will be informed about studies and will have to give their approval. All data will be anonymised to prevent identification of individual employees. The employer will not have access to the collected data, and will only receive a final report containing summary information.

Information will be collected in a continuous manner. The observer will separately take notes about emotional states, as well as events occurred throughout the work day.

A. Preliminary studies

The participant observation study was designed in an iterative manner. Initial experiment assumptions about the observation were verified during two subsequent preliminary studies. Such approach allowed the identification of the design flaws and to propose a better, more mature research design.

There were two participant observations conducted. Both were performed by software developers with bachelors' degrees in Computer Science (polish engineer's degree). They were students of second level studies and they were working part-time (20 hours per week) in two different IT companies as software developers.

The first observation took place in May 2015 and lasted for three weeks. During this time the observer made 121 notes about himself, and his colleagues made 79 notes about their own emotional states.

The second observer worked with 3 other software developers. During four weeks in June 2015, he made 162 observations about his own emotions and 78 notes about emotions of his colleagues. Furthermore, his co-workers made 81 notes about their own emotions.

B. Notes templates

For the purpose of the observation initially two template notes were prepared, named as the Sheet S and the Sheet A.

Information on the emotions of the observer will be collected continuously, i.e. during work, she or he must pay attention to emotional states and well-being, as well as to analyse their impact on the productivity. After every hour the observer makes notes about his emotions and productivity using Sheet S. Additionally, at the end of the working day, each co-worker is asked to describe her or his emotional state and productivity during the day, their insights are then documented on the Sheet A. Each observation is documented on the sheet by filling in four fields: emotional state, fatigue, productivity and additional notes.

For purposes of denoting emotional states the Russell circumplex model of affect was chosen [14]. According to this model, emotions are described with two continuous dimensions of valence and arousal. The valence (pleasure) dimension differentiates positive from negative emotions, and as it is continuous, both directions might be graded. Values close to zero correspond to neutral emotional states. The dimension of arousal (activation) allows the differentiation of active and passive emotional states. The value of zero means relaxed or calm, while four indicates excited or stimulated [15].

In the fatigue field it is possible to indicate whether it was present or not. Information regarding the productivity is denoted using a grade 5 scale, where 0 represents no

Participant observation of software developers

Observer code:	**Date:**	**Experiment day:**	**Work hours:**	**S**

Work hour	**Emotional state**	**Fatigue**	**Productivity**	**Notes**
1	Valence -2 -1 0 1 2 Arousal: 0 1 2 3 4	yes / no	0 1 2 3 4	
2	Valence -2 -1 0 1 2 Arousal: 0 1 2 3 4	yes / no	0 1 2 3 4	
3	Valence -2 -1 0 1 2 Arousal: 0 1 2 3 4	yes / no	0 1 2 3 4	
4	Valence -2 -1 0 1 2 Arousal: 0 1 2 3 4	yes / no	0 1 2 3 4	
5	Valence -2 -1 0 1 2 Arousal: 0 1 2 3 4	yes / no	0 1 2 3 4	
6	Valence -2 -1 0 1 2 Arousal: 0 1 2 3 4	yes / no	0 1 2 3 4	
7	Valence -2 -1 0 1 2 Arousal: 0 1 2 3 4	yes / no	0 1 2 3 4	
8	Valence -2 -1 0 1 2 Arousal: 0 1 2 3 4	yes / no	0 1 2 3 4	
9	Valence -2 -1 0 1 2 Arousal: 0 1 2 3 4	yes / no	0 1 2 3 4	
General notes:				

Fig. 1. Observer notes template – Sheet S

productivity, and 4 the highest productivity. Furthermore, for each observation as to the events influencing the emotional states and productivity should be described within the notes field.

After the first participant observations, lack of regularity and credibility in fulfilling sheets by co-workers was noticed. Therefore, an additional template, marked as Sheet O, was prepared. It is meant for taking notes by the observer about emotions of colleagues and their productivity. During his work, observer should non-invasively watch her or his co-workers. Particular attention should be paid to the events affecting their emotions.

Sheet S (Fig. 1) contains 9 rows of the fields described above, each row for every hour of the working day. The number of rows in the Sheet A, corresponds to the number of co-workers observed. Sheet O contains just one row. Sheets also includes additional information, such as observer code, date of observation and working hours. After each working day n+2 sheets should be collected, where n is the number of co-workers.

IV. Results of preliminary study

In the context of designing the participant observation research, the most important outcome of the preliminary study refers to the issue of observing co-workers. The first observation revealed, and the second confirmed, that co-workers are not willing to systematically fill observation sheets, even if it is required only once a day. Only 25% of the notes were taken on time – at the end of the working day, or at the beginning of the next one, the remaining were delivered with delay.

Therefore, during the second execution, the observer was asked to take notes about her or his colleagues. These notes have been compared with the notes taken by the co-workers observed. The analysis showed that the external evaluation agreed only on not more than 25% of notes with the self-assessment. While the assessment of valence and arousal were sometimes under, and sometimes over-estimated, in the case of productivity, co-workers always have reported greater diligence than suggested by the external evaluation.

TABLE I
Number of reported high and low productivity depending on the Arousal and the Valence

Productivity	**Arousal**				
	0	*1*	*2*	*3*	*4*
Low	31	42	63	42	1
High	9	16	123	51	2
Total	59	98	285	106	3

Productivity	**Valence**				
	-2	*-1*	*0*	*1*	*2*
Low	5	31	61	54	28
High	0	11	37	121	40
Total	6	67	154	250	82

These results indicate the necessity to involve objective methods for identifying emotions and productivity evaluation. Without proper tools support [16], observation of emotional states and productivity of co-workers is useless. Therefore, in further studies only self-assessment of a trained observer will be conducted.

Based on the collected data, analysis of the impact of emotional states on productivity has been conducted. Fig. 2 and Table I show the number of reported high and low

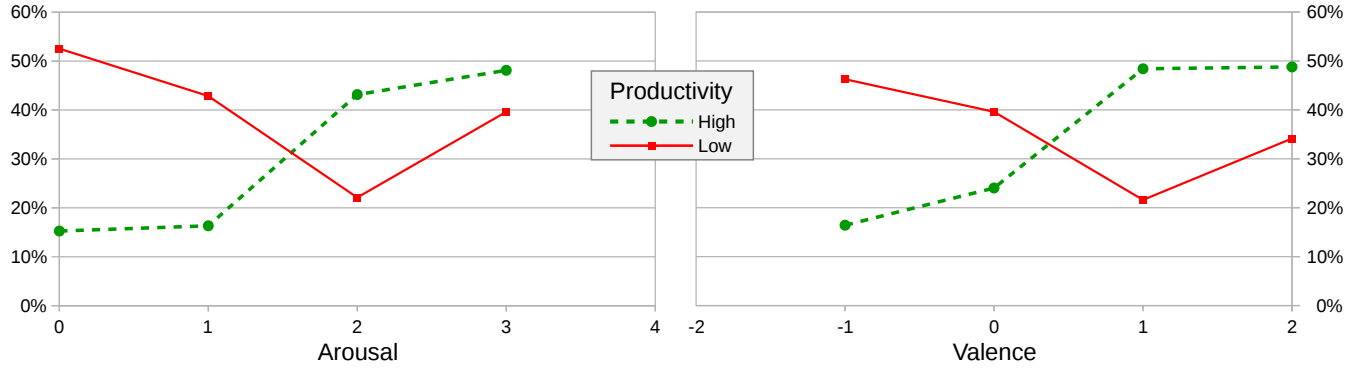


Fig. 2. Percentage of high and low productivity of software developers, depending on their arousal and valence

productivity periods depending on the arousal and valence. Due to the small number of observations, results with the highest arousal and the lowest valence were removed from the chart.

Low productivity is least likely to occur at an average arousal and higher value of valence. Correlation between high productivity and both arousal and valence seems to be more linear. High value of productivity was reported more frequently along with high arousal and high valence. Considering the results above, it appears that the optimum in terms of productivity of software developers, is medium arousal and high valence.

V. Conclusion

Preliminary studies have proved that the proposed research design may bring new and useful information on the role of emotions in the software development process. The use of qualitative research methods may allow for a better understanding the role of emotions in productivity of IT professionals.

Preliminary studies have also discredited the legitimacy of both external and self-assessment of co-workers. Therefore, it is proposed that observers make out notes only on their own emotional states and productivity. Observation of other team members will be possible only after the development of a reliable and non-intrusive methods of emotion recognition and productivity evaluation.

In addition, the results of the two preliminary participant observation studies showed the correlation between high productivity and both arousal and valence. Optimal, in the case of productivity, emotional states include, among others, joy, amusement, surprise and anxiety.

Understanding the role of emotions in the software engineering may lead to the development of new affect-aware extensions to both traditional and modern software development management methodologies.

Acknowledgment

I thank my students R. Lewandowski and J. Falenczyk for conducting the preliminary participant observations described in the paper.

References

[1] The Standish Group, “2015 Chaos Report,” Tech. Rep., 2015.
[2] C. D. Fisher and N. M. Ashkanasy, “The emerging role of emotions in work life: an introduction,” *Journal of Organizational Behavior*, vol. 21, pp. 123–129, 2000. doi: 10.1002/(sici)1099-1379(200003)21:2<123::aid-job33>3.0.co;2-8
[3] F. L. Capretz, “Bringing the human factor to software engineering,” *IEEE Software*, vol. 31, no. 2, pp. 104–104, 2014. doi: 10.1109/MS.2014.30
[4] D. Graziotin, X. Wang, and P. Abrahamsson, “Are Happy Developers More Productive?” in *Proceedings of the 14th International Conference PROFES*, 2013, pp. 50–64. doi: 10.1007/978-3-642-39259-7
[5] I. A. Khan, “Mood Independent Programming,” in *Proceedings of the 14th European conference on Cognitive ergonomics: invent! explore!*, 2007, pp. 269–272. doi: 10.1145/1362550.1362606
[6] D. Garcia, M. S. Zanetti, and F. Schweitzer, “The role of emotions in contributors activity: a case study on the gentoo community,” in *Cloud and Green Computing (CGC), 2013 Third International Conference on*. IEEE, 2013, pp. 410–417. doi: 10.1109/CGC.2013.71
[7] A. Murgia, P. Tourani, B. Adams, and M. Ortu, “Do developers feel emotions? an exploratory analysis of emotions in software artifacts,” *Conference on Mining Software*, pp. 262–271, 2014. doi: 10.1145/2597073.2597086
[8] T. Fritz and S. C. Müller, “Stuck and Frustrated or In Flow and Happy: Sensing Developers Emotions and Progress,” in *37th International Conference on Software Engineering (ICSE 2015)*, 2015. doi: 10.1109/icse.2015.334
[9] A. Landowska, “Emotion monitor-concept, construction and lessons learned,” in *Federated Conference on Computer Science and Information Systems (FedCSIS)*. IEEE, 2015, pp. 75–80. doi: 10.15439/2015F264
[10] D. Graziotin, X. Wang, and P. Abrahamsson, “Do feelings matter? On the correlation of affects and the self-assessed productivity in software engineering,” *Journal of Software: Evolution and Process*, pp. 1–21, 2014. doi: 10.1002/smr.1673
[11] C. Marshall, C. B. Rossman, and G. B. Rossman, *Designing qualitative research*. Sage publications, 2010. ISBN 141297044X
[12] K. M. DeWalt, B. DeWalt R., and C. B. Wayland, *Participant Observation: A Guide for Fieldworkers*. AltaMira Press, 2010. ISBN 0759100446
[13] C. B. Seaman, “Qualitative methods in empirical studies of software engineering,” *IEEE Transactions on Software Engineering*, vol. 25, pp. 557–572, 1999. doi: 10.1109/32.799955
[14] J. A. Russell, “A circumplex model of affect,” *Journal of personality and social psychology*, vol. 39, no. 6, p. 1161, 1980. doi: 10.1037/h0077714
[15] A. Kolakowska, A. Landowska, M. Szwoch, W. Szwoch, and M. R. Wrobel, “Modeling emotions for affect-aware applications,” in *Information Systems Development and Applications*. Faculty of Management, University of Gdask, Poland, 2015.
[16] A. Kołakowska, A. Landowska, M. Szwoch, W. Szwoch, and M. R. Wróbel, “Emotion recognition and its applications,” *Human-Computer Systems Interaction: Backgrounds and Applications 3*, pp. 51–62, 2014. doi: 10.1007/978-3-319-08491-6_5

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1549–1552
DOI: 10.15439/2016F360
ACSIS, Vol. 8. ISSN 2300-5963

AgileSafe – a method of introducing agile practices into safety-critical software development processes

Katarzyna Łukasiewicz
Gdańsk University of Technology ul. Narutowicza 11/12, 80-233 Gdańsk, Poland
Email: katarzyna.lukasiewicz@pg.gda.pl

Janusz Górski
Gdańsk University of Technology ul. Narutowicza 11/12, 80-233 Gdańsk, Poland
Email: jango@pg.gda.pl

Abstract—This article introduces AgileSafe, a new method of incorporating agile practices into critical software development while still maintaining compliance with the software assurance requirements imposed by the application domain. We present the description of the method covering the process of its application and the input and output artefacts.

I. Introduction

AGILE software development methods have been introduced in response to particular concerns emerging from the changing needs of the market. In practice, volatile requirements, demanding clients from diverse backgrounds and a growing need for shortening time-to-market made a growing number of software companies seek alternatives to their own traditional approaches. A similar tendency can be presently observed in many domains, including the public sector [1] and further in what seemed to be a leading plan-driven environment - the safety-critical software domain. In this case, however, strictly agile methods will not be the answer as they are insufficient on the safety assurance and certification side. The question is if and how to enrich the agile practices with safety and risk management practices without sacrificing agility and still providing the necessary assurance level.

While process optimization is vital to the business and economical aspect of a software development project, in the safety-critical software domain its profits will not be sufficient unless a company is able to conform to standards and guidelines, which regulate a particular industry. Clients demand their products to be of high quality, on time and within a reasonable budget but at the same time the software need to be certified by an appropriate authority in order to be licenced for use in its destined environment. For this reason, in safety-critical software domain it is not enough to improve the software development process to provide financial profits to the company. It is also necessary that the safety requirements are adequately identified and assured throughout the process. Changing the process will likely result in changes in the safety evidence collected during the development, which may affect the scope and structure of the certification process. Therefore, each change to the process should be carefully analysed from the viewpoint of future audits and potential consequences to the outcome of these audits. This makes a change in safety-critical software development process a complicated and potentially costly operation depending on how the company is introducing new elements to the process. This may be a problem for SMEs where budget constrains can be a significant barrier in introducing the change.

Attempts to provide a hybrid, disciplined-agile, approaches bringing together best of the two worlds are already in effect for several years. A growing body of evidence, including industrial reports, shows that obtaining the right balance is doable and profitable especially when the companies decide to employ competent experts to develop a custom made approach. Examples of such reports can be found in [2], [3], [4], [5] and were surveyed in [6]. What is more, in 2012 FDA (Food and Drug Administration) recognized the AAMI TIR45:2012 - Guidance on the use of AGILE practices in the development of medical device software [7]. It concludes that agile practices can be successfully used in safety-critical software development and that such practices can be compliant with IEC 62304 [8] standard. It also provides a mapping between agile methods and IEC 62304 activities.

However encouraging these reports are, 'tailoring' a software development method can be a costly and complicated process. If hybrid approaches are to be applied in a larger scale, more available and ready to use solutions are needed. An example can be SafeScrum [9], which concentrates on adapting Scrum into safety-critical software development. The method has been already applied in a number of real life projects and most of them ended in success as well as required standard certification [10]. Another approach is AV-Model [11] combining the traditional V-Model with Scrum and focusing on medical device software development and the IEC 62304 standard.

In this article we propose a new method, called AgileSafe, of incorporating agile practices into critical software development while still maintaining compliance with the software assurance requirements imposed by the application domain, which is complementary to the existing methods. AgileSafe is addressed mainly to SMEs developing safety-critical software, to support them in the process of introducing new practices into their software development environment. AgileSafe provides a user with tools enabling her/him to create, with a help of guidelines and questionnaires, a hybrid agile approach customized for the project. It also provides a tool

978-83-60810-90-3/$25.00

for handling conformance with standards and norms while introducing this new approach.

II. Overview of the AgileSafe Method

To provide and maintain control over the safety requirements and over the scope and level of their assurance, AgileSafe employs evidence-based arguments which are explicitly maintained during the software development process. These arguments follow the ISO/IEC 15026 [12] recommendations on *assurance cases*. The main idea is to provide assurance cases both for the software development process and for the end product itself. While the latter is the essence of demonstrating product conformance with the stated safety objectives, the former is complementary to it and allows to demonstrate adequacy of the chosen software development practices and in particular their conformity with safety requirements imposed by relevant standards. AgileSafe focuses on an explicit development process assurance case demonstrating that the selected range of software development practices is conformant with the requirements of the relevant safety related standards.

Although, for demonstration reasons we concentrate on the medical domain, the method is generic and can be adapted to different safety-critical domains. In order to address a broad range of products resulting from development, the process assurance arguments are based on the standards that are relevant for a particular application domain.

The prerequisite for applying AgileSafe to a particular (planned) safety-related software development project is that we have identified a set of relevant standards we want to be compliant with. Then, the main results of applying AgileSafe to this project are:

- *Project Practices Set* (*PPS*)– a custom prepared hybrid approach composed of plan-based and agile software development practices;
- *Assurance arguments*, for each selected standard.

Figure 1 presents a BPMN model of the AgileSafe.

Based on the *Project characteristics,* prepared during the AS.P.1 *Analyse the project process*, a user is guided through the set *AgileSafe Practices Knowledge Base,* which contains descriptions of software development practices. The method suggestions are based on the good practices for software development as well as the results of experiments conducted in the course of our research.

The customized *Project Practices Set*, prepared in the AS.P.2 *Select practices* process, should be later implemented in the software development process (AS.P.7 *Apply Practices*). For each given *Standard* a *Practices Compliance Argument* need to be *developed/updated* (AS.P.3). These *Practices Compliance Arguments* are then *adapted* (AS.P.4), depending on the *PPS*, into *Project Practices Compliance Arguments*. Based on them, the *Project Compliance Arguments* are *prepared* (AS.P.5) and they are the end products of the method allowing the user to AS.P.6 *Assert conformance*, using the *Evidence* prepared during the AS.P.7 *Apply practices* process.

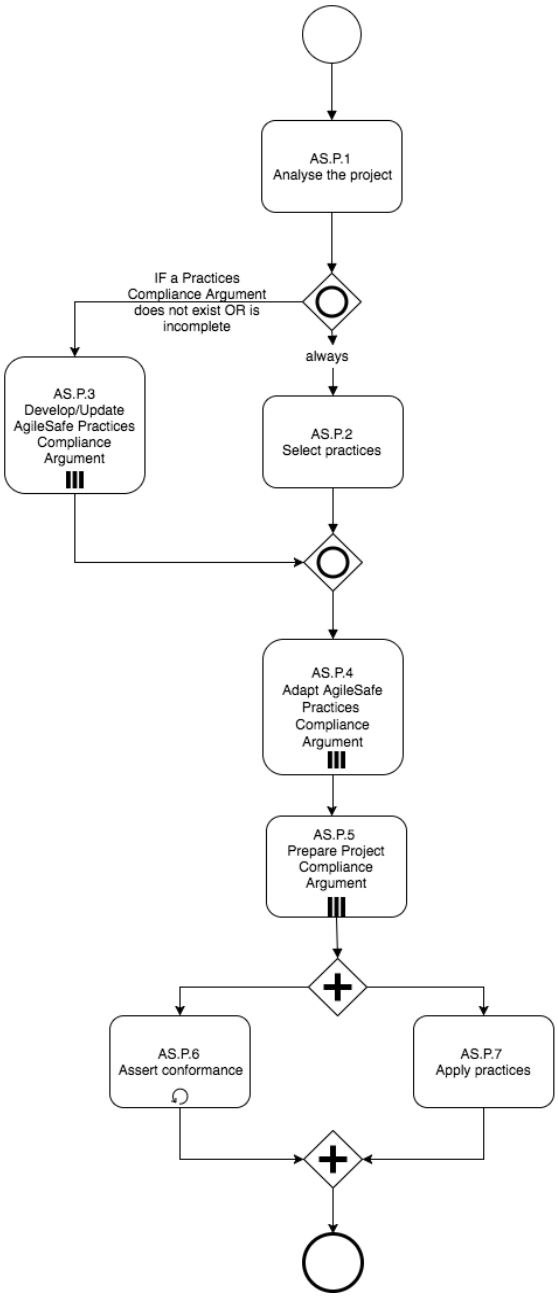


Fig. 1 Diagram of AgileSafe method

In further sections we explain components of the method in more detail.

III. Assurance Arguments in AgileSafe

An assurance argument is a structure of claims, which is based on explicitly provided evidence and demonstrates that a product or a system satisfies a detailed set of requirements. Recommendation on the structure of assurance arguments (called *assurance cases)* can be found in ISO/IEC 15026 standard [12].

Since 2005 the idea of safety assurance arguments has been analysed in depth by both FDA and SEI (Software Engineering Institute) [13]. This partnership resulted in series of documents presenting potential uses of assurance arguments in FDA certification process [13],[14]. With FDA

currently recommending the use of assurance arguments in the process of qualification of medical devices in order to present compliance with safety requirements, explicit use of assurance cases is gaining increasing recognition.

In AgileSafe there are two types of assurance arguments: for process assurance and for product assurance, as shown in Figure 2.

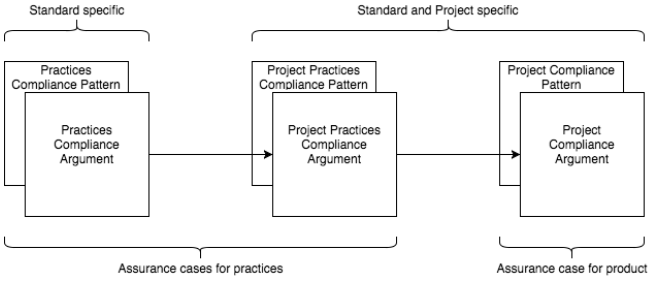


Fig. 2 Assurance cases in AgileSafe

The patterns for assurance arguments hold the information about the argument's structure and are used as a template for creating specific arguments.

In AgileSafe, assurance arguments are developed separately for each applicable standard in order to support certification on the standard by standard basis. A more detailed description of the assurance cases employed in AgileSafe (see Figure 2) is given below.

A. Practices Compliance Argument

Practices Compliance Argument is a template which is developed separately for each relevant standard. Its structure is based on the standard requirements. To make such templates uniform, each template is following the *Practices Compliance Pattern*. This pattern is generic and focuses on the conformance of practices from *AgileSafe Practices Knowledge Base* with particular requirements of the considered standard. For each such requirement it proposes an argumentation strategy and the range of software engineering practices used for collecting evidence demonstrating the compliance. It also contains explicit justification that the argumentation strategy is adequate on the condition that the evidence is collected and integrated with the argument. A list of claims concerning different types of practices, which may contribute to satisfying the standard demand, is presented, each claim postulating the potential of a given practice to generate the evidence needed to demonstrate compliance.

If the *Practices Knowledge Base* is complete, the *Practices Compliance Argument* once prepared for a specific standard remains unchanged and can be used in multiple projects which have to comply with the standard. Nevertheless, in the 'learning period' of the AgileSafe method the *Practices Knowledge Base* is expected to grow acquiring new practices and therefore the *Practices Compliance Arguments* will change and evolve in time.

B. Project Practices Compliance Argument

Project Practices Compliance Argument is a *Practices Compliance Argument* adapted to a specific project and is characterized by the *Project Practices Set* specific to this project. The *Project Practices Compliance Argument* refers only to the practices used in the project along with the description of *Evidence* they are providing. Its structure is defined in the *Project Practices Compliance Pattern*.

C. Project Compliance Argument

Project Compliance Argument is an assurance argument in its traditional form. It is structured around a particular standard and is used to collect the required product related evidence to demonstrate conformance with given standard. The *Evidence* should be collected with accordance to the *Project Compliance Pattern* and be an effect of AS.P.7 *Apply practices* process.

D. Tool support for assurance cases

To handle assurance cases AgileSafe follows the TRUST-IT methodology [15] and uses the Argevide NOR-STA [16] services. NOR-STA provides means for developing, maintaining and assessing assurance cases and integrating them with the supporting evidence. All the assurance cases presented in Section VI were developed using this tool.

In NOR-STA, argument conclusion is represented by a *claim* node. A node of type *argumentation strategy* links the claim with the corresponding premises and uses a *rationale* node to explain and justify the inference leading from the premises to the claim. A premise is a sort of assertion and can be in particular another claim to be further justified by its own premises or a *fact* represented by an assertion to be demonstrated by the supporting evidence. The evidence is integrated by nodes of type *reference* which point to external resources (files of any type, web pages, etc.). In addition, *information* nodes can be used in any place to provide explanatory information.

IV. Practices Selection Process

In order to support users while selecting the software engineering practices from the *AgileSafe Practices Knowledge Base,* AgileSafe offers the guidance, which is based on *Project characteristics.* The practices maintained in the *AgileSafe Practices Knowledge Base* include both plan-driven and agile methods documented in the literature, and may also include custom developed hybrid practices. Upon the selection of the practices an assurance argument is composed along with assurance arguments for selected standards.

Users are able to introduce their own practices into the *Knowledge Base* by following the AS.P.2.1 *Introduce new practice p*rocess.

V. Assessing Conformance

The process of assessing conformance with given standards is based on the set of assurance arguments, mainly the *Project Compliance Arguments*.

The *Project Compliance Arguments* are being developed in parallel to the software development process (AS.P.7 *Apply practices*). The *Evidence* prepared in the course of AS.P.7 *Applying practices* from the *Project Practices Set*

should be placed in the accurate nodes as indicated in the *Project Practices Compliance Arguments*. Upon the certification process for a particular *Standard* User should be able to prove conformance by presenting the applicable *Project Compliance Argument* along with *Project Practices Compliance Argument*, which contains additional reasoning behind the choice of practices (*PPS*) and *Evidence* used in the process.

VI. Conclusion

In this article we presented an overview of AgileSafe, a method of agile software development with simultaneous controlling conformity with selected safety related standards.

The ultimate objective is to help SMEs involved in safety-critical software development to introduce agile practices in the most profitable way while meeting the requirements imposed by safety standards and certification bodies.

Recently we conducted a case study which goal was to use AgileSafe to incorporate selected risk management practices into an agile project with safety requirements. We have followed all of the steps of the AgileSafe method algorithm and prepared the artefacts as specified in the method, collecting all the metrics that we planned to collect. A complete set of AgileSafe assurance cases was prepared for ISO 14971 standard. The basic result of the case study was that it was possible to conduct an agile project while still controlling its conformity to the ISO 14971 requirements.

In the nearest future we plan to interview experts in the field of safety certification as well as practitioners who were involved in adapting agile practices to safety-critical system development in order to obtain their feedback on AgileSafe.

Acknowledgment

This work was partially supported by the Statutory Grant by the Polish Ministry of Higher Education for the Faculty of Electronics, Telecommunications and Informatics of Gdansk University of Technology.

References

[1] A. Kaczorowska, "Traditional And Agile Project Management In Public Sector And ICT," in *Proceedings of the Federated Conference on Computer Science and Information Systems (FedCSIS)*, Łódź, Poland, September 2015. DOI: 10.15439/2015F279

[2] G. B Alleman., M Henderson., C. H. M. Hill, R.Seggelke, "Making Agile Development Work in a Government Contracting Environment Measuring velocity with Earned Value," in *Proceedings of the Agile Development Conference 2003, Salt Lake City, Utah*, June 2003. DOI: 10.1109/ADC.2003.1231460

[3] R. Paige, R. Charalambous, X. Ge, P. Brooke, "Towards Agile Engineering of High- Integrity Systems," in *Proceedings of 27th International Conference on Computer Safety, Reliability and Security (SAFECOMP)*, September 2008. DOI: 10.1007/978-3-540-87698-4_6

[4] K. Petersen, C. Wohlin, "The effect of moving from a plan-driven to an incre- mental software development approach with agile practices," in *Empirical Software Engineering*, 15(6):654–693, 2010. DOI: 10.1007/s10664-010-9136-6

[5] R. Rasmussen, T. Hughes, J. R. Jenks, J. Skach, "Adopting Agile in an FDA Regulated Environment," in *Proceedings of Agile Conference*, Chicago, USA, August 2009. DOI: 10.1109/AGILE.2009.50

[6] J. Górski, K. Łukasiewicz. „Assessment of Risks Introduce to Safety Critical Software by Agile Practices – a Software Engineer's Perspective" in. *Computer Science*, 13(4), AGH University of Science and Technology Press 2012. DOI: http://dx.doi.org/10.7494/csci.2012.13.4.165

[7] AAMI TIR45: 2012 Technical Information Report Guidance on the use of AGILE practices in the development of medical device software

[8] ISO/IEC 62304 Medical device software — Software life cycle processes, standard

[9] SafeScrum, http://www.sintef.no/safescrum

[10] T. Stålhane, T. Myklebust, G. K. Hanssen, "Safety standards and Scrum – A synopsis of three standards", http://www.sintef.no/globalassets/safety-standards-and-scrum_may2013.pdf

[11] M. Mc Hugh, F. Mc Caffery, G. Coady "An Agile Implementation within a Medical Device Software Organisation", in Proceedings of The 14th International SPICE Conference Process Improvement and Capability dEtermination 2014. DOI: 10.1007/978-3-319-13036-1_17

[12] ISO/IEC 15026 Systems and software engineering -- Systems and software assurance

[13] C. B. Weinstock, J. B. Goodenough. "Towards an Assurance Case Practice for Medical Devices". TECHNICAL NOTE Software Engineering Institute October 2009. http://www.sei.cmu.edu/reports/09tn018.pdf

[14] FDA: Guidance – Total Product Life Cycle: Infusion Pump-Premarket Notification Submissions [510(k)], 2010

[15] J. Górski, "Trust Case – a case for trustworthiness of IT infrastructures*"* in Cyberspace Security and Defense: Research Issues, NATO Science Series II: Mathematics, Physics and Chemistry, 196 (). Springer-Verlag, pp. 125-142 (2005). DOI: 10.1007/1-4020-3381-8_7

[16] Argevide NOR-STA, https://www.argevide.com/en

4th Workshop on Model Driven Approaches in System Development

FOR many years, various approaches in system design and implementation differentiate between the specification of the system and its implementation on a particular platform. People in software industry have been using models for a precise description of systems at the appropriate abstraction level without unnecessary details. Model-Driven (MD) approaches to the system development increase the importance and power of models by shifting the focus from programming to modeling activities. Models may be used as primary artifacts in constructing software, which means that software components are generated from models. Software development tools need to automate as many as possible tasks of model construction and transformation requiring the smallest amount of human interaction.

A goal of the proposed workshop is to bring together people working on MD languages, techniques and tools, as well as Domain Specific Languages (DSL) and applying them in the requirements engineering, information system and application development, databases, and related areas, so that they can exchange their experience, create new ideas, evaluate and improve MD approaches and spread its use. The intention is to target an interdisciplinary nature of MD approaches in software engineering, as well as research topics expressed by but not limited to acronyms such as Model Driven Software Engineering (MDSE), Model Driven Software Development (MDSD), Domain Specific Modeling (DSM), and OMG's Model Driven Architecture (MDA).

1st Workshop on MDASD was organized in the scope of ADBIS 2010 Conference, held in Novi Sad, Serbia. From 2012, MDASD becomes a regular bi-annual FedCSIS event.

TOPICS

- MD Approaches in System Design and Implementation – Problems and Issues
- MD Approaches in Software Process Models
- MD Approaches in Databases and Information Systems
- MD Approaches in Software Quality and Standards
- Metamodeling, Modeling and Specification Languages
- Model Transformation Languages
- Model-to-Model, Model-to-Text, and Model-to-Code Transformations in Software Process
- Transformation Techniques and Tools
- Domain Specific Languages (DSL) and Domain Specific Modeling (DSM) in System Specification and Development
- Design of Metamodeling and Modeling Languages and Tools
- MD Approaches in Requirements Engineering and Business Process Modeling
- MD Approaches in System Reengineering and Reverse Engineering
- MD Approaches in HCI development
- MD Approaches in GIS development
- MD Approaches in Document Engineering
- Model Based Software Verification
- Theoretical and Mathematical Foundations of MD Approaches
- Organizational and Human Factors, Skills, and Qualifications for MD Approaches
- Teaching MD Approaches in Academic and Industrial Environments
- MD Applications and Industry Experience

EVENT CHAIRS

- **Luković, Ivan,** University of Novi Sad, Serbia

STEERING COMMITTEE

- **Gray, Jeff,** University of Alabama, United States
- **Mernik, Marjan,** University of Maribor, Slovenia
- **Ristić, Sonja,** University of Novi Sad, Faculty of Technical Sciences, Serbia
- **Tolvanen, Juha-Pekka,** MetaCase, Finland

PROGRAM COMMITTEE

- **Amaral, Vasco,** The New University of Lisbon, Portugal
- **Bryant, Barrett,** University of North Texas, United States
- **Budimac, Zoran,** Faculty of Sciences, Univ. of Novi Sad, Serbia
- **Chen, Haiming,** Chinese Academy of Sciences, China
- **Erradi, Mohammed,** ENSIAS, Mohammed-V University, Morocco
- **Fertalj, Krešimir,** University of Zagreb, Croatia
- **Gray, Jeff,** University of Alabama, United States
- **Ivanović, Mirjana,** University of Novi Sad, Serbia
- **Janousek, Jan,** Czech Technical University, Czech Republic
- **João Varanda Pereira, Maria,** Instituto Politecnico de Braganca, Portugal
- **Karagiannis, Dimitris,** University of Vienna, Austria
- **Kardaş, Geylani,** Ege University International Computer Institute, Turkey
- **Kollár, Ján,** Technical University of Kosice, Slovakia
- **Kosar, Tomaž,** University of Maribor, Slovenia

- **Krdzavac, Nenad,** Michigan State University, United States
- **Liu, Shih-Hsi Alex,** California State University, United States
- **Maćoš, Dragan,** Beuth University of Applied Sciences, Germany
- **Melo de Sousa, Simão,** University of Beira Interior, Portugal
- **Mernik, Marjan,** University of Maribor, Slovenia
- **Milosavljević, Gordana,** University of Novi Sad, Serbia
- **Nešković, Siniša,** University of Belgrade, Serbia
- **Porubän, Jaroslav,** Technical University of Kosice, Slovakia
- **Rangel Henriques, Pedro,** Universidade do Minho, Portugal
- **Ristić, Sonja,** University of Novi Sad, Faculty of Technical Sciences, Serbia
- **Seidl, Martina,** Johannes Kepler University, Austria
- **Selic, Bran,** Malina Software Co., Canada
- **Sierra Rodríguez, José Luis,** Universidad Complutense de Madrid, Spain
- **Slivnik, Boštjan,** University of Ljubljana, Slovenia
- **Suvajdžin-Rakić, Zorica,** University of Novi Sad, Serbia
- **Tolvanen, Juha-Pekka,** MetaCase, Finland
- **Vangheluwe, Hans,** University of Antwerp, Belgium
- **Wimmer, Manuel,** Vienna University of Technology, Austria

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1555–1564
DOI: 10.15439/2016F196
ACSIS, Vol. 8. ISSN 2300-5963

Interoperability of MAS DSMLs via horizontal model transformations

Emine Bircan
International Computer Institute,
Ege University, 35100, Bornova,
Izmir, Turkey
eminebircanbircan@gmail.com

Moharram Challenger
International Computer Institute,
Ege University, 35100, Bornova,
Izmir, Turkey
moharram.challenger@mail.ege.edu.tr

Geylani Kardas
International Computer Institute,
Ege University, 35100, Bornova,
Izmir, Turkey
geylani.kardas@ege.edu.tr

***Abstract*—In this paper, we present our approach which aims at improving the mechanism of constructing language semantics over the interoperability of domain-specific modeling languages (DSMLs) developed for Multi-agent Systems (MAS) and hence providing a more efficient way of extension for the executability of modeled agent systems on various underlying agent platforms. Differentiating from the existing MAS DSML studies, our proposal is based on determining entity mappings and building horizontal model transformations between the metamodels of MAS DSMLs which are in the same abstraction level. The applicability of the approach is demonstrated in the paper by constructing horizontal transformations between two full-fledged agent DSMLs, called SEA_ML and DSML4MAS. Use of these transformations has enabled SEA_ML instance models now to be executable on new agent platforms and that feature has been provided with less effort comparing with the implementation of needed transformations between SEA_ML and those new agent platforms from scratch.**

***Keywords*—Metamodel; Model transformation; Domain-specific Modeling Language; Multi-agent System**

I. Introduction

Multi-agent systems (MASs) are those systems having software agents within an environment where agents interact to solve problems in a competitive or collaborative manner. In MASs, software agents are expected to be autonomous, mostly through a set of reactive/proactive behaviors designed for addressing situations likely to happen in particular domains [1]. Both internal agent behavior model and interactions within a MAS become even more complex and hard to implement when taking into account the varying requirements of different agent environments [2]. Hence, working in a higher abstraction level is of critical importance for the development of MASs since it is almost impossible to observe code level details of MASs due to their internal complexity, distributedness and openness [3].

In order to master the abovementioned problems of developing MASs, agent-oriented software engineering (AOSE) researchers define various agent metamodels (e.g. [4-7]), which include fundamental entities and relations of agent systems. In addition, many model-driven agent development approaches are provided such as [8-10] and by enriching MAS metamodels with some defined syntax and semantics (usually translational semantics [11]), researchers also propose domain-specific languages (DSLs) / domain-specific modeling languages (DSMLs) (e.g. [12-20]) for facilitating the development of MASs. DSLs / DSMLs [21-23] have notations and constructs tailored toward a particular application domain (e.g. MAS) and help to the model-driven development (MDD) of MASs. MDD aims to change the focus of software development from code to models [24], and hence many AOSE researchers believe that this paradigm shift introduced by MDD may also provide the desired abstraction level and simplify the development of complex MAS software [3].

In AOSE, perhaps the most popular way of applying model-driven engineering (MDE) for MASs is based on providing DSMLs specific to agent domain with including appropriate integrated development environments (IDEs) in which both modelling and code generation for system-to-be-developed can be performed properly. Proposed MAS DSMLs such as [13], [17], [19] usually support modelling both the static and the dynamic aspects of agent software from different MAS viewpoints including agent internal behaviour model, interaction with other agents, use of other environment entities, etc. Within this context, abstract syntaxes of the languages are represented with metamodels covering those aspects and required viewpoints to some extent. Following the construction of abstract and concrete syntaxes based on the MAS metamodels, the operational semantics of the languages are provided in the current MAS DSML proposals by defining and implementing entity mappings and model-to-model (M2M) transformations between the related DSML's metamodel and the metamodel(s) of popular agent implementation and execution platform(s) such as JACK[1], JADE[2] and JADEX[3]. Finally, a series of model-to-text (M2T) transformations are implemented and applied on the outputs of the previous M2M transformations which are the MAS models conforming to the related agent execution platforms. Hence, agent software codes, MAS configuration files, etc. pertaining to the implementation and deployment of the modeled agent systems on the target MAS platform are generated automatically.

When we take into account the different abstractions covered by the metamodels of MAS DSMLs and the underlying agent execution platforms, DSML metamodels can be accepted as the platform-independent metamodels (PIMMs) of agent systems while metamodels of the agent execution platforms are platform-specific metamodels

This study is funded by the Scientific Research Projects Directorate of Ege University under grant 16-UBE-001.

[1] http://aosgrp.com/products/jack/ (last access: June, 2016)
[2] http://jade.tilab.com/ (last access: June, 2016)
[3] https://www.activecomponents.org/ (last access: June 2016)

(PSMMs) according to the OMG's well-known Model-driven Architecture (MDA)[4] as also indicated in [5] and [9].

Above described methodology applied in the current MAS DSML development approaches for the derivation of operational semantics unfortunately requires the definition and implementation of new M2M and M2T transformations from scratch in order to make the DSMLs functional for different agent execution platforms. In other words, for each new target agent execution platform, MAS DSML designers should repeat all the time-consuming and mostly troublesome steps of preparing the vertical transformations between the related DSML and this new agent platform.

Motivated by the similarity encountered in the abstract syntaxes of the available MAS DSMLs, we are quite convinced that both the definition and the implementation of M2M transformations between the PIMMs of MAS DSMLs would be more convenient and less laborious comparing with the transformations required between MAS PIMMs and PSMMs in the way of enriching the support of MAS DSMLs for various agent execution platforms. Hence, in this paper, we present our approach which aims at improving the mechanism of constructing language semantics over the interoperability of MAS DSMLs and hence providing a more efficient way of extension for the executability of modeled agent systems on various underlying agent platforms. Differentiating from the existing MAS DSML studies (e.g. [13], [16], [17], [19], [20]), our proposal is based on determining entity mappings and building horizontal M2M transformations between the metamodels of MAS DSMLs which are in the same abstraction level. In this paper, we also investigate the applicability of the proposed DSML interoperability approach by constructing horizontal transformations between two full-fledged agent DSMLs called SEA_ML [19] and DSML4MAS [5] respectively.

The rest of the paper is organized as follows: In Sect. 2, the approach for the MAS DSML interoperability is presented. Applicability of the approach is discussed in Sect. 3 by taking into consideration two MAS DSMLs. In Sect. 4, a case study on the development of an agent-based stock exchange system with using the proposed approach is given. Related work is given in Sect. 5. Finally, Sect. 6 concludes the paper and states the future work.

II. Proposed Approach for the Interoperability of MAS DSMLs

As indicated in the introduction, support of current MAS DSMLs for each agent execution platform is enabled by repetitively defining and implementing a chain of vertical M2M and M2T transformations. Available M2M and M2T transformations are specific for each different agent platform and almost all of them can not be re-used while extending the executability of the MAS models for a new agent platform. Due to the difficulty encountered on repeating those vertical model transformation steps, current MAS DSML proposals (e.g. [13-15], [17], [19]) mostly support the execution of modeled agents on just one agent platform. Rarely, two different platforms are supported (e.g. [5], [9]) and as far as we know, there is no any MAS DSML which provides an operational semantics for more than two different agent execution platforms. In order to increase the platform variety, we propose benefiting from the vertical transformations already existing between the syntax of a MAS DSML (let us call $DSML_1$) and metamodels of various agent platforms for enabling model instances of another MAS DSML (let us call $DSML_2$) executable on the same agent platforms by just constructing horizontal transformations between the PIMMs of the MAS DSMLs in question. Therefore, instead of defining and implementing N different M2M and M2T transformations for N different agent platforms, creation of only one single set of M2M transformations between $DSML_1$ and $DSML_2$ can be enough for the execution of $DSML_2$'s model instances on these N different agent platforms.

Fig. 1 depicts the construction of model transformations between MAS DSMLs and hence re-use of already existing transformations between those DSMLs and agent platforms. Let the abstract syntaxes of $DSML_1$, $DSML_2$ and $DSML_3$ be the metamodels MM_1, MM_2 and MM_3 respectively. Horizontal lines between these MAS DSMLs represent the M2M transformations between these metamodels. According to the figure, agent systems modeled in $DSML_1$ are already executable on the agent platforms A and B (due to the existing vertical transformations for these platforms), while $DSML_2$ model instances are executable on the agent platforms X, Y and Z. Similarly M2M and M2T transformations were already provided for the execution of $DSML_3$ model instances on the agent platforms α, β, θ respectively. If $DSML_1$ is required to support X and Y agent platforms, designers should prepare new model transformations separately for those agent platforms (shown with dotted arrows in Fig. 1) in case of the absence of horizontal transformations between MM_1 and MM_2. Hence, construction of only one set of horizontal M2M transformations between $DSML_1$ and $DSML_2$ enables $DSML_1$'s automatic support on agent platforms X, Y (and also Z). Conversely, same is also valid for extending the $DSML_2$'s support for agent execution platforms. Interoperability between $DSML_1$ and $DSML_2$ over these newly defined horizontal transformations also makes transformation and code generation of $DSML_2$ model instances for the agent platforms A and B. In addition to the important decrease in the number of transformations, construction of horizontal model transformations between the PIMMs of MAS DSMLs is more feasible and easier than the vertical transformations since the DSMLs are in the same abstraction level according to MDA.

III. Interoperability between SEA_ML and DSML4MAS

In this section, we discuss the applicability of the proposed approach by taking into account the construction of the interoperability between two MAS DSMLs called SEA_ML and DSML4MAS. Both DSMLs enable the

[4] http://www.omg.org/mda/ (last access: June 2016)

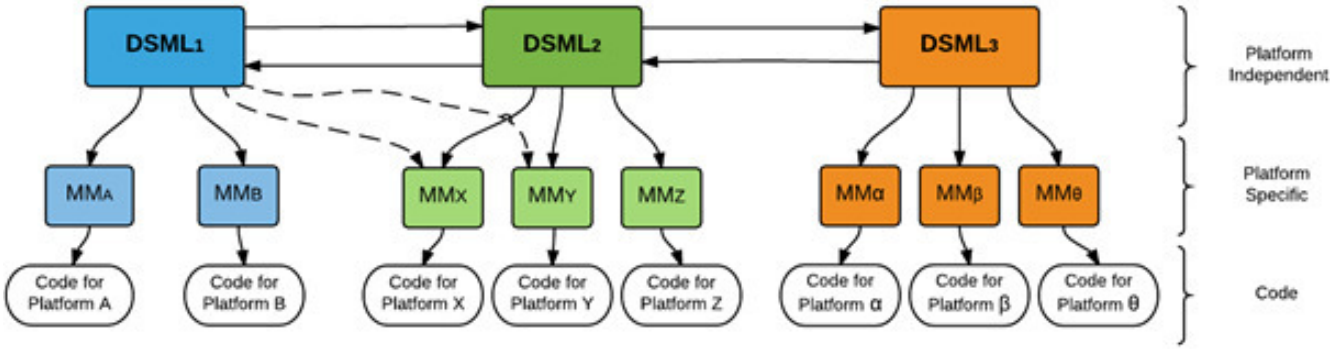


Fig. 1 Interoperability of MAS DSMLs via horizontal model transformations

modeling of agent systems according to various agent internal and MAS organizational viewpoints. They provide a clear visual syntax for MAS modeling and code generation for agent implementation and execution platforms. Moreover, both languages are equipped with Eclipse-based IDEs in which modeling and automatic generation of MAS components are possible. These features of the languages led us to choose them in this study. In the following subsections, brief introduction of these DSMLs and implemented model transformations between these languages are given.

A. SEA_ML

SEA_ML [19] provides a convenient and handy environment for agent developers to construct and implement software agent systems working on various application domains. In order to support MAS experts when programming their own systems, and to be able to fine-tune them visually, SEA_ML covers all aspects of an agent system from the internal view of a single agent to the complex MAS organization. In addition to these capabilities, SEA_ML also supports the model-driven design and implementation of autonomous agents who can evaluate semantic data and collaborate with semantically-defined entities of the Semantic Web, like Semantic Web Services (SWS) [25]. That feature exactly differentiates SEA_ML and makes unique regarding any other MAS DSML currently available. Within this context, it includes new viewpoints which specifically pave the way for the development of software agents working on the Semantic Web environment. Modeling agents, agent knowledgebases, platform ontologies, SWS and interactions between agents and SWS are all possible in SEA_ML.

SEA_ML's metamodel is divided into eight viewpoints, each of which represents a different aspect for developing Semantic Web enabled MASs [19]. *Agent's Internal Viewpoint* is related to the internal structures of semantic web agents (SWAs) and defines entities and their relations required for the construction of agents. It covers both reactive and Belief-Desire-Intention (BDI) agent architectures. *Interaction Viewpoint* expresses the interactions and communications in a MAS by taking messages and message sequences into account. *MAS Viewpoint* solely deals with the construction of a MAS as a whole. It includes the main blocks which compose the complex system as an organization. *Role Viewpoint* delves into the complex controlling structure of the agents and addresses role types. *Environmental Viewpoint* addresses the use of resources and interaction between agents with their surroundings. *Plan Viewpoint* deals with an agent Plan's internal structure, which are composed of Tasks and atomic elements such as Actions. *Ontology Viewpoint* addresses the ontological concepts which constitute agent's knowledgebase (such as belief and fact). *Agent - SWS Interaction Viewpoint* defines the interaction of agents with SWS including the definition of entities and relations for service discovery, agreement and execution. A SWA executes the semantic service finder Plan (SS_FinderPlan) to discover the appropriate services with the help of a special type of agent called SSMatchMakerAgent who executes the service registration plan (SS_RegisterPlan) for registering the new SWS for the agents. After finding the necessary service, one SWA executes an agreement plan (SS_AgreementPlan) to negotiate with the service. After negotiation, a plan for service execution (SS_ExecutorPlan) is applied for invoking the service.

The collection of SEA_ML viewpoints constitutes an extensive and all-embracing model of the MAS domain. SEA_ML's abstract syntax combines the generally accepted aspects of MAS (such as MAS, Agent Internal, Role and Environment) and introduces two new viewpoints (Agent-SWS Interaction and Ontology) for supporting the development of software agents working within the Semantic Web environment [2].

SEA_ML can be used for both modeling MASs and generation of code from the defined models. SEA_ML instances are given as inputs to a series of M2M and M2T transformations to achieve executable artifacts of the system-to-be-built for JADEX agent platform and semantic web service description documents conforming to *Web Ontology Language for Services (OWL-S)* ontology[5]. It is also possible to automatically check the integrity and validity of SEA_ML models [26]. Complete discussion on SEA_ML can be found in [19].

[5] https://www.w3.org/Submission/OWL-S/ (last access: June 2016)

B. DSML4MAS

DSML4MAS [5], [13] is perhaps one of the first complete MAS DSMLs in which a PIMM, called PIM4Agents, provides an abstract syntax for different aspects of agent systems. Similar to SEA_ML's viewpoints, both internal behavior model of agents and agent interactions in a MAS are covered by PIM4Agents views / aspects. *Multiagent view* contains all the main concepts in a MAS such as Agent, Cooperation, Capability, Interaction, Role and Environment. *Agent view* focuses on the single autonomous entity (agent), the roles it plays within the MAS and the capabilities it has to solve tasks and to reach the environment resources. *Behavioural view* describes how plans are composed by complex control structures and simple atomic tasks like sending a message and how information flows between those constructs. In here, a plan is a specialized version of behavior composed of activities and flows. Activities and tasks are minimized parts of the work and flows provide the communication between these parts. *Organization view* describes how single autonomous entities cooperate within the MAS and how complex organizational structures can be defined. Social structure in the system is defined with cooperation entity where agents and organizations take part in. The structure has its own protocol defining how the entities interact in a cooperation. Agents have "domainRoles" for the interaction and these roles are attached to actors by "actorBinding" entities where actors are representative entities within the corresponding interaction protocol. *Role view* examines the behaviour of an agent entity in an organization or cooperation. An agent's role covers the capabilities and information to have access to a set of resources. *Interaction view* describes how the interaction in the form of interaction protocols takes place between autonomous entities or organizations. Agents communicate over the PIM4Agents Protocol which refers to actors and "messageFlows" between these actors. Finally, *Environment view* contains the resources accessed and shared by agents and organizations. Agents can communicate with the environment indirectly via using resources. Resources can store knowledge from BDI agents for changing beliefs by using Messages and Information flows.

As indicated in [5], grouping modelling concepts in DSML4MAS allows the metamodel evolution by adding new modelling concepts in the existing aspects, extending existing modelling concepts in the defined aspects, or defining new modelling concepts for describing additional aspects of agent systems. For instance, SWS integration into the system models conforming to DSML4MAS is provided via introducing the SOAEnvironment entity [27] which extends the Environment entity and contains service descriptions. Agents use service descriptions to specify the Services they are searching for and then service interaction is realized by InvokeWS and ReceiveWS tasks which are inherited from Send and Receive task entities described in PIM4Agents.

Similar to SEA_ML, DSML4MAS also enables the MDD of MAS including a concrete graphical syntax [28] based on the abovementioned PIMM (PIM4Agents) and an operational semantics for the execution of modeled agent systems on JACK or JADE agent platforms. Extensions to the language introduced in [27] provide the description of the services inside an agent environment according to specifications such as *Web Services Modeling Language (WSML)*[6] or *Semantic Annotation of WSDL and XML Schema (SAWSDL)*[7]. Interested readers may refer to [5], [13] and [27] for an extensive discussion on DSML4MAS.

C. Horizontal Model Transformations between SEA_ML and DSML4MAS

We have applied the horizontal transformability approach described in Sect. 2 for establishing the interoperability between SEA_ML and DSML4MAS. As shown in Fig. 2, SEA_ML currently supports the MAS implementation for JADEX BDI architecture and SWS generation according to the OWL-S ontology. In order to extend its platform support capability, new M2M and M2T transformations should be prepared for each new implementation platform. For instance, M2M transformations are needed between the abstract syntax of SEA_ML and PSMM of JADE framework to make SEA_ML instances also executable on the JADE platform. It is worth indicating that definition and application of M2T transformations are also required for the code generation from the outputs of the previous SEA_ML to JADE transformations. Instead, we can follow the approach introduced in Sect. 2 by just writing the horizontal transformation rules between the metamodels of SEA_ML and DSML4MAS and running those transformations on SEA_ML instances for the same purpose: making SEA_ML models executable also on JADE platform. That is possible since DSML4MAS has already support on JADE and JACK agent platforms and SAWSDL and WSDL semantic service ontologies via vertical transformations between its metamodel and metamodels of the corresponding system implementation platforms. Realization of horizontal transformations between SEA_ML and DSML4MAS has extra benefits such as the execution of SEA_ML instances also on JACK platform and/or implementation of the modeled SWS according to SAWSDL or WSDL specifications (Fig. 2).

Before deriving the rules of transformations, we should determine the entity mappings between both languages since the transformations are definitely based on these entity mappings. Comparing with the mappings we previously provided in [15] or [19] for the transformability of SEA_ML instances to MAS execution platforms, we have experienced that the determination of the entity mappings in this study was easier and took less time. We believe that the reason of this efficiency originates from the fact that metamodels of SEA_ML and DSML4MAS are in the same abstraction level and provide close entities and relations in similar viewpoints for MAS modeling.

[6] http://www.wsmo.org/wsml/ (last access: June, 2016)
[7] https://www.w3.org/TR/sawsdl/ (last access: June, 2016)

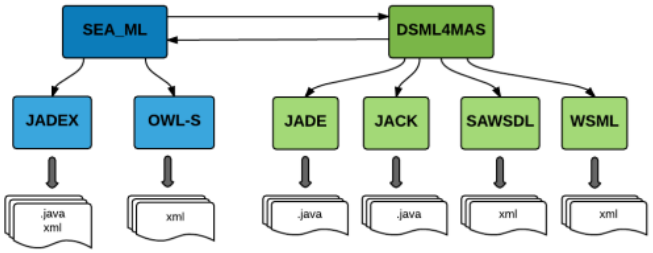


Fig. 2 Interoperability of SEA_ML and DSML4MAS

Table 1 lists some of the important mappings constructed between first-class entities of these two languages. For instance, two agent types (SWA and SSMatchmakerAgent) defined in SEA_ML are mapped onto the autonomous entity Agent defined in DSML4MAS. Likewise, meta-entities pertaining to agent plan types (SS_RegisterPlan, SS_FinderPlan, SS_AgreementPlan and SS_ExecutorPlan) required for the interaction between the semantic services are mapped with the Plan concept of DSML4MAS. Since Actor entity in DSML4MAS has access to resources and owns capabilities needed for agent interactions, SEA_ML's Role entity is mapped onto Actor entity.

One interesting mapping is encountered between SEA_ML's SWS entity and DSML4MAS's SOAEnvironment since it enables the representation of SEA_ML semantic services in DSML4MAS model instances. On the DSML4MAS side, SOAEnvironment entity, which is extended from Environment entity, includes services in general. Hence, SEA_ML SWS entity is mapped onto SOAEnvironment entity and SEA_ML WebService entities are mapped onto Service entities. In SEA_ML WebService definition, every service has Interface, Process and Grounding. Interface entity represents the information about service inputs, outputs and any other necessary information. Process entity has internal information about the service and finally Grounding entity defines the invocation protocol of the web service [19]. DSML4MAS services are described with Blackbox and Glassbox entities [27]. BlackBox is used to define a service's functional and non-functional parameters while Glassbox includes the description of the internal service process. The Functionals are described in terms of service signature that are input and output parameters, and specifications that are preconditions and effects. The NonFunctionals are defined in terms of price, service name and developer. Hence, Interface and Process entities of services defined in SEA_ML are mapped onto DSML4MAS Functionals which have input and output definitions. On DSML4MAS side, agent interactions with services are provided by InvokeWS and ReceiveWS tasks. So, SEA_ML Grounding, which represents the physical structure of the underlying web service executed for the corresponding SWS, is mapped to InvokeWS. Remaining mappings listed in Table 1 (e.g. SEA_ML SWO to DSML4MAS Organization, SEA_ML Environment to DSML4MAS Environment) are very simple to determine since the related entities on both sides have similar or almost same functionality within the syntaxes of the languages.

TABLE I.
ENTITY MAPPINGS BETWEEN THE METAMODELS OF SEA_ML AND DSML4MAS

SEA_ML MM	DSML MM
SWA (Semantic Web Agent)	Agent
SSMatchmakerAgent	Agent
Role	Actor
SWS (Semantic Web Service)	SOAEnvironment
Environment	Environment
WebService	Service
Interface	Functionals
Process	Functionals
Grounding	InvokeWS
Input	Input
Output	Output
Precondition	Precondition
SS_RegisterPlan	Plan
SS_FinderPlan	Plan
SS_AgreementPlan	Plan
SS_ExcecutorPlan	Plan
SWO (Semantic Web Organization)	Organization

After determining the entity mappings between SEA_ML and DSML4MAS, it is necessary to provide model transformation rules which are applied at runtime on SEA_ML instances to generate DSML4MAS counterparts of these instances. For that purpose, transformation rules should be formally defined and written according to a model transformation language. In this study, we preferred to use ATL Transformation Language (ATL)[8] to define the model transformations between SEA_ML and DSML4MAS. ATL is one of the well-known model transformation languages, specified as both metamodel and textual concrete syntax. An ATL transformation program is composed of rules that define how the source model elements are matched and navigated to create and initialize the elements of the target models. In addition, ATL can define an additional model querying facility which enables specifying the requests onto models. ATL also allows code factorization through the definition of ATL libraries. Finally, ATL has a transformation engine and an IDE that can be used as a plug-in on an Eclipse platform. These features of ATL caused us to prefer it as the implementation language for the horizontal transformations from SEA_ML to DSML4MAS.

ATL is composed of four fundamental elements. The first one is the header section defining attributes relative to the transformation module. The next element is the import section which is optional and enables the importing of some existing ATL libraries. The third element is a set of helpers

[8] https://eclipse.org/atl/(last access: June, 2016)

that can be viewed as the ATL equivalents to the Java methods. The last element is a set of rules that defines the way target models are generated from source models.

Following listings include some excerpts from the written ATL rules in order to give some flavor of M2M transformations provided in this study. To this end, rule in Listing 1 enables the transformation of the elements covered by the Agent-SWS Interaction viewpoint of SEA_ML to their counterparts included in DSML4MAS's Multiagent system viewpoint. In line 1, rule is named uniquely. In line 2, the source metamodel is chosen and renamed as swsinteractionvp with "from" keyword. The target metamodel is indicated and renamed as pim4agents with "to" keyword (Line 3). In the following lines (between 4 and 14), instances of SEA_ML SWA and SSMatchmakerAgent entities are selected and transformed to DSML4MAS Agent instances. Transformation of agent roles and plans are also realized by using "Set" and "allInstances" functions. It is worth indicating that types of Plan instances seem to be transformed to DSML4MAS behavior in the given listing although all SEA_ML Plan types are semantically mapped to DSML4MAS Plan as listed in Table 1. That is because some of the DSML4MAS meta-entities are collected with tag definitions in Ecore representations which take the same name with the related viewpoint. For instance, plans are not defined solely with their names; instead they are collected in behavior definitions. Hence, in order to provide the full transformations of the plans with all their attributes, ATL rule is written here as mapping SEA_ML plan instances to the DSML4MAS behaviors. Inside another helper rule, those behaviors are separated into the corresponding plans and so exact transformation of SEA_ML plan instances to DSML4MAS plans are realized.

```
01 rule SWSInteractionVP2MultiagentSystem {
02  from   swsinteractionvp: SWSInteraction!SWSInteractionViewpoint
03  to     pim4agent: PIM4Agents!MultiagentSystem(
04       agent <- Set{SWSInteraction!SemanticWebAgent.allInstances()},
05       agent <- Set{SWSInteraction!SSMatchmakerAgent.allInstances()},
06       role <- Set{SWSInteraction!Role.allInstances()},
07       role <- Set{SWSInteraction!RegistrationRole.allInstances()},
08       behavior <- Set{SWSInteraction!SS_AgreementPlan.allInstances()},
09       behavior <- Set{SWSInteraction!SS_ExecutorPlan.allInstances()},
10       behavior <- Set{SWSInteraction!SS_FinderPlan.allInstances()},
11       behavior <- Set{SWSInteraction!SS_RegisterPlan.allInstances()},
12       environment<-Set{SWSInteraction!SWS.allInstances()},
13       environment<-Set{SWSInteraction!Grounding.allInstances()} )
14 }
```

Listing 1 An excerpt from the SWSInteractionVP2MultiagentSystem rule

Another example can be given for the transformation of SEA_ML SWS instances to DSML4MAS SOAEnvironment instances. In Listing 2, the first rule provides the related transformation. After controlling the name of the SWS by applying the helper rule "nameControl", its attributes are converted to their counterparts in the abstract syntax of DSML4MAS. Again with using helper rules, the web services composed by this semantic service are determined and transformed to DSML4MAS Service instances. Only a fragment of SWS2SOAEnvironment can be given in here due to space limitations. The remaining part of this rule provides the transformation of each SEA_ML semantic service's discovery, engagement and execution components (e.g. Interface, Process and Grounding) to their counterparts in DSML4MAS models according to the mappings given in Table 1.

```
01  rule SWS2SOAEnvironment{
02    from  envIN: SWSInteraction!SWS(envIN.nameControl)
03    to envOUT: PIM4Agents!SOAEnvironment (
04          name <- envIN.setName(),
05          service<-envIN.setWebServices()
06    )
07  }

08  rule WebService2Service{
09    from  webService: SWSInteraction!WebService
10    to service: PIM4SWS!Service (
11          ID<- webService.setName()
12    )
13  }
```

Listing 2 Excerpts from the SWS2SOAEnvironment and WebService2Service rules

In Listing 3, the helper rules used in above transformation rules are given. "nameControl" helper is executed on the SEA_ML SWS instances. It controls whether a SWS has a name attribute. Based on this attribute's existence, the rules returns true (line 4) or false (line 5). That Boolean result is evaluated by the caller rule given in Listing 2. In the second helper rule, the name attribute of a semantic service is controlled. In case of the name is empty, the string given in line 10 is assigned as the value for the transformed service's (SOAEnvironment) name attribute. Otherwise, the name of the source SWS is returned back to the called rule (SWS2SOAEnvironment) to be assigned as the name of the transformed SOAEnvironment instance which will be included in the target DSML4MAS model.

```
01  helper context SWSInteraction!SWS
02    def: nameControl: Boolean =
03      if not (self.name.oclIsUndefined())
04        then  true
05        else  false
06      endif;

07  helper context SWSInteraction!SWS
08    def: setName(): String =
09      if (self.name = '')
10        then 'SERVICE_NAME_IS_EMPTY'
11        else  self.name
12      endif;
```

Listing 3 Helper rules used by the SWS2SOAEnvironment rule

IV. Case Study: Agent-Based Stock Exchange System

In this section, the interoperability of SEA_ML and DSML4MAS is demonstrated by applying the proposed horizontal model transformations for the development of an agent-based stock exchange system. The system is modeled in SEA_ML and transformed to a DSML4MAS instance to use the generation power of DSML4MAS language. In this way, the implementation of this system's agents on JADE (or JACK) platform and services as SAWSDL or WSML ontology instances can be possible by using the operational semantics of DSML4MAS which is already provided for the

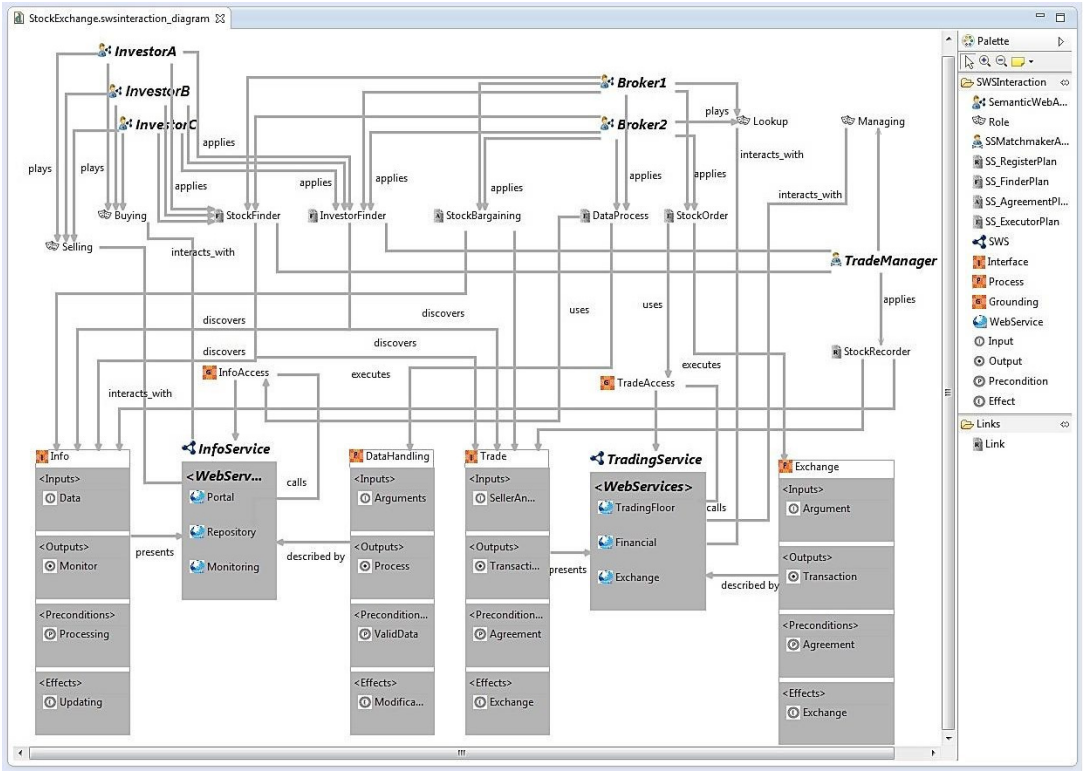


Fig. 3 Instance model of the multi-agent stock exchange system in SEA_ML with including the agents, semantic web services and their relations

execution of agents and the generation of semantic web services (see Fig. 2).

Stock trading is one of the key items in economy and estimating its behavior and taking the best decision in it are among the most challenging issues. Agents in a MAS can share a common goal or they can pursue their own interests. That nature of MASs exactly fits to the requirements of free market economy. Moreover, Stock Exchange Market has lots of services which are offered for Investors (Buyer or Seller), Brokers, and Stock Managers. These services can be represented with semantic web services to achieve more accurate service finding and service matching.

When considering the structure of the system, the semantic web agents work within a semantic web organization for Stock System including sub-organizations for Stock Users where the Investor and Broker agents reside, and the Stock Market where the system's internal agents, e.g. Trade Manager (a SSMatchmaker agent instance) work. The Stock Market organization also has two sub-organizations, the Trading Floor and the Stock Information System. These organizations and sub-organizations have their own organizational roles. These organizations also need to access some resources in other environments. Therefore, they have interactions with the required environments to gain access permissions. For example, agents in the Stock Market sub-organization need to access bank accounts and some security features, so that they can interact with the Banking & Security environment. All of the user agents including Investors and Brokers cooperate with Trade Manager to access the Stock Market. Also, the user agents interact with each other. For instance, Investor Agents can cooperate with Brokers to exchange stock for which Brokers are expert. More information on developing such stock trading agents can be found in [29].

To model the system in SEA_ML, Agent-SWS Interaction viewpoint is considered as the representative for SEA_ML viewpoints. This viewpoint is the most important aspect of MASs working in semantic web environments. Fig. 3 shows a screenshot from the SEA_ML's modeling environment in which instances of both the semantic services and the agent plans required for the stock exchange are modeled, including their relations according to Agent-SWS interaction viewpoint of SEA_ML. Investor and Broker agents can be modeled with appropriate plan instances in order to find, make the agreement with and execute the services. The services can also be modeled for the interaction between the semantic web service's internal components (such as Process, Grounding, and Interface), and the SWA's plans. It is important to indicate that the stock exchange system given in here was already modeled in the SEA_ML environment before this study and instead of re-modeling the whole system (e.g. in DSML4MAS), the existing model is intentionally adopted in here to examine the applicability of the proposed approach. In fact, the model in question is much more complicated and we can only consider the agent-SWS interaction aspect due to page limits of this paper. Discussion on the whole model can be found in [19] and the sources of the model are available at the SEA_ML's distribution website[9].

[9] SEA_ML, its modeling tool and the instance models for the case study are available at: http://serlab.ube.ege.edu.tr/resources.html (last access: June 2016)

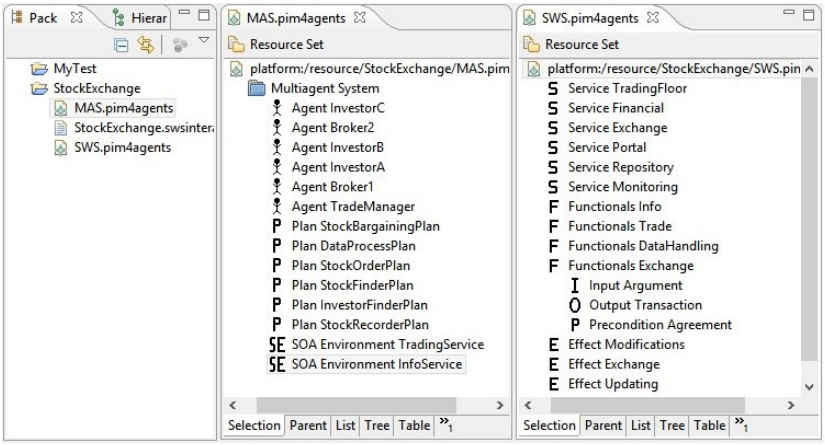


Fig. 4 An excerpt of the output in its Ecore tree view representation achieved after executing the M2M transformations on the SEA_ML instance

We can see from the instance model given in Fig. 3 that an investor agent (e.g. InvestorA) would play the Buying role and apply its StockFinder plan for finding an appropriate Trading service interface of one TradingService SWS in order to buy some stocks. This plan would realize the discovery by interacting with the TradeManager SSMatchmakerAgent which has registered the services by applying the StockRecorder plan. As InvestorA cooperates with Broker1 in order to receive some expert advice for its investment, at the next step, the Broker1 agent applies its StockBargaining plan for negotiating with the already discovered services. This negotiation would be made through the Trade interface of the SWS. Finally, if the result of the negotiation were positive, the agent would apply the StockOrder plan to call the TradingFloor of the SWS by executing its Exchange process and using its TradeAccess grounding with which the service would be realized. In a similar way, Investor agents could cooperate with Brokers and interact with the TradeManager in order to collect some information about the market, e.g. the rate of exchange for a currency or the fluctuation rate for a specific stock.

The designed instance model is controlled based on the provided constraint rules in SEA_ML tool to check its validity. Then, the horizontal model transformations discussed in the previous section are executed on this SEA_ML instance model and as result, we have succeeded to automatically achieve the counterpart models conforming to DSML4MAS. To realize the transformation, the SEA_ML metamodel, the SEA_ML instance models for this case study, and the DSML4MAS metamodel are given to the ATL engine as input and the instance models of the case study in DSML4MAS are generated by the engine with executing our transformation rules. An excerpt of the XMI file containing the output DSML4MAS instance model is given in Fig. 4 in which the output instance can be seen in its Ecore tree view representation.

The generated model conforms to the specification of DSML4MAS's abstract syntax, so it can be handled with DSML4MAS's graphical editor[10]. To visualize the instance model in DSML4MAS, the only thing needed is to add the related graphical notations to the generated instance model. The screenshot given in Fig. 5 shows the appearance of output instance model in the concrete syntax of DSML4MAS. We can examine from the figure that the agents and their relations we modeled in SEA_ML are exactly reflected to a DSML4MAS model after execution of the M2M transformations proposed in this study. From now on, it is straightforward to automatically achieve platform-specific executables and documents of this MAS model for JADE or JACK agent platforms and SAWSDL or WSML semantic service ontologies since DSML4MAS has already own a chain of M2M and M2T transformations for these agent execution platforms and service ontologies as discussed in Sect. 3.2.

V. Related Work

In the last decade, AOSE researchers have noteworthy efforts on the derivation and use of DSLs / DSMLs for MAS. For instance, the Agent-DSL [12] is used to specify the agency properties that an agent needs to accomplish its tasks. However, the proposed DSL is presented only with its metamodel and provides just a visual modeling of the agent systems according to agent features, like knowledge, interaction, adaptation, autonomy and collaboration. Likewise, in [30], the authors introduced two DSMLs. These languages are described by metamodels which can be seen as the representations of the main concepts and relationships identified for each of the particular domains again introduced in [30]. The study included only the abstract syntax of the related DSMLs and does not give the concrete syntax or semantics of the DSMLs.

As previously discussed in this paper, Hahn [13] introduced a DSML for MAS called DSML4MAS. The abstract syntax of the DSML was derived from a platform independent metamodel [5] which was structured into several aspects each focusing on a specific viewpoint of a MAS. In order to provide a concrete syntax, the appropriate graphical notations for the concepts and relations were defined [28]. Furthermore, DSML4MAS supports the

[10]The IDE of DSML4MAS is available at: https://sourceforge.net/projects/dsml4mas/ (last access: June 2016)

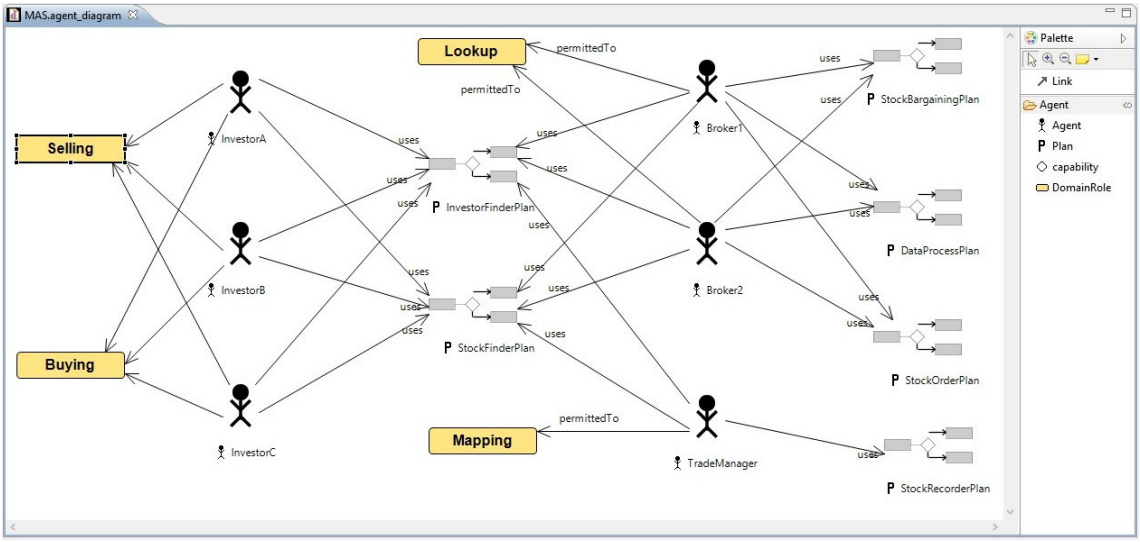


Fig. 5 Partial instance model of the agent-based stock exchange system in DSML4MAS achieved after application of the defined M2M transformations

deployment of modeled MASs both in JACK and JADE agent platforms by providing an operational semantics over model transformations.

Another DSML was provided for MASs in [17]. The abstract syntax was presented using the Meta-object Facility (MOF)[11] , the concrete syntax and its tool was provided with Eclipse Graphical Modeling Framework (GMF)[12], and finally the code generation for the JACK agent platform was realized with model transformations using Eclipse JET[13]. However, the developed modeling language was not generic since it was based on only the metamodel of one of the specific MAS methodologies called Prometheus. A similar study was performed in [14] which proposes a technique for the definition of agent-oriented engineering process models and can be used to define processes for creating both hardware and software agents. This study also offers a related MDD tool based on a specific MAS development methodology called INGENIAS.

Originating from a well-formalized syntax and semantics, Ciobanu and Juravle defined and implemented a language for mobile agents in [16]. They generated a text editor with auto-completion and error signaling features and presented a way of code generation for agent systems starting from their textual description. A recent work conducted in [20] aimed at creating a UML-based agent modeling language, called MAS-ML, which is able to model the well-known types of agent internal architectures, namely simple reflex agent, model-based agent, reflex agent, goal-based agent and utility-based agent. Representation and exemplification of all supported agent architectures in the concrete syntax of the introduced language are given. MAS-ML is also accompanied with a graphical tool which enables agent modeling. However, the current version of MAS-ML does not support any code generation for MAS frameworks which prevents the execution of the modeled agent systems.

Finally, by considering our previous studies, in [15] and [18], we showed the derivation of a DSL for the MDE of agent systems working on the Semantic Web. That initial version of the language was refined and enriched with a graphical concrete syntax in [19]. This new language, called SEA_ML, covered an enhanced version of agent-SWS interaction viewpoint in which modeling those interactions can be elaborated as much as possible for the exact implementation of agent's service discovery, agreement and execution dynamics. We also presented the formal semantics of the language [26] and discussed how the applied methodology can pave the way of evolutionary language development for MAS DSLs [2]. Moreover, qualitative evaluation and quantitative analysis of SEA_ML have been recently performed over a multi-case study protocol [31].

The work presented in this paper contributes to the above mentioned MAS DSL/DSML studies by introducing the interoperability of the languages and hence the proposed MDE technique helps to facilitate the platform support of the MAS DSMLs comparing with the existing agent platform extensibility approaches which deal with the definition and implementation of new M2M and M2T transformations for each execution platform. To the best of our knowledge, the work herein is the first effort on the interoperability of the MAS DSMLs and it is the first study in AOSE which employs horizontal model transformations to enable this interoperability. It is worth indicating that apart from our proposal, only the work conducted in [10] considers the application of horizontal transformations. However, that study just provides the transformation between the metamodels of two specific AOSE methodologies (Prometheus and INGENIAS) to realize MAS implementation on exactly one agent deployment platform and does not consider MAS DSML interoperability or language extensibility on various agent platforms.

[11] http://www.omg.org/mof/ (last access: June 2016)
[12] http://www.eclipse.org/modeling/gmp/ (last access: June, 2016)
[13] https://eclipse.org/modeling/m2t/?project=jet (last access: June 2016)

II. Conclusion

An approach for extending the execution platform support of MAS DSMLs over language interoperability is presented in this paper. The interoperability is provided by defining and implementing horizontal M2M transformations between the agent metamodels which constitute the syntaxes of MAS DSMLs. Due to being at the same abstraction level, both mapping the model entities and implementing the model transformations are more convenient and less laborious comparing with the M2M and M2T transformation chain required in the way of enriching the support of DSMLs for various agent execution platforms. The applicability of the approach is demonstrated by constructing transformations between two full-fledged agent DSMLs.

As the future work, we first plan to extend the applicability of this interoperability approach for some other MAS DSMLs such as MAS-ML [20]. Later, the assessment of the language interoperability will be performed e.g. by taking into consideration the amount and quality of the automatically generated artifacts for MAS software. For this purpose, an improved version of the evaluation framework described in [31] can be employed.

References

[1] C. Badica, Z. Budimac, H. D. Burkhard, M. Ivanovic. 2011. Software agents: Languages, tools, platforms, *Computer Science and Information Systems* 8(2): 255-298, DOI: 10.2298/CSIS110214013B

[2] M. Challenger, M. Mernik, G. Kardas, T. Kosar. 2016. Declarative specifications for the development of multi-agent systems, *Computer Standards & Interfaces* 43: 91-115, DOI: 10.1016/j.csi.2015.08.012

[3] G. Kardas. 2013. Model-driven development of multiagent systems: a survey and evaluation. *The Knowledge Engineering Review* 28(4): 479-503, DOI: 10.1017/S0269888913000088

[4] A. Omicini, A. Ricci, M. Viroli. 2008. Artifacts in the A&A meta-model for multi-agent systems. *Autonomous Agents and Multi-Agent Systems* 17(3): 432-456, DOI: 10.1007/s10458-008-9053-x

[5] C. Hahn, C. Madrigal-Mora, K.Fischer. 2009. A Platform-Independent Metamodel for Multiagent Systems. *Autonomous Agents and Multi-Agent Systems* 18(2): 239-266, DOI: 10.1007/s10458-008-9042-0

[6] G. Beydoun, G. Low, B. Henderson-Sellers, H. Mouratidis, J. J. Gomez-Sanz, J. Pavon, C. Gonzalez-Perez. 2009. FAML: A Generic Metamodel for MAS Development. *IEEE Transactions on Software Engineering* 35(6): 841-863, DOI: 10.1109/TSE.2009.34

[7] I. Garcia-Magarino. 2014. Towards the integration of the agent-oriented modeling diversity with a powertype-based language. *Computer Standards & Interfaces* 36: 941–952, DOI: 10.1016/j.csi.2014.02.002

[8] J. Pavon, J. Gomez-Sanz, R. Fuentes. 2006. Model driven development of multi-agent systems, *Lecture Notes in Computer. Science* 4066: 284–298, DOI: 10.1007/11787044_22

[9] G. Kardas, A. Goknil, O. Dikenelli, N.Y. Topaloglu. 2009. Model driven development of semantic web enabled multi-agent systems, *International Journal of Cooperative Information Systems* 18(2): 261-308, DOI: 10.1142/S0218843009002014

[10] J. M. Gascuena, E. Navarro, A. Fernandez-Caballero, R. Martínez-Tomas. 2014. Model-to-model and model-to-text: looking for the automation of VigilAgent. *Expert Systems* 31(3): 199-212, DOI: 10.1111/exsy.12023

[11] B.R. Bryant, J. Gray, M. Mernik, P. J. Clarke, R. B. France, G. Karsai. 2011. Challenges and Directions in Formalizing the Semantics of Modeling Languages. *Computer Science and Information Systems* 8(2): 225-253, DOI: 10.2298/CSIS110114012B

[12] U. Kulesza, A. Garcia, C. Lucena, P. Alencar. 2005. A generative approach for multi-agent system development. *Lecture Notes in Computer Science* 3390: 52–69, DOI: 10.1007/978-3-540-31846-0_4

[13] C. Hahn. 2008. A Domain Specific Modeling Language for Multiagent Systems. 7th Int'l Conf. on Autonomous agents and Multi-agent systems (AAMAS 2008), pp. 233-240

[14] R. Fuentes-Fernandez, L. Garcia-Magarino, A. Maria Gomez-Rodriguez, J. Carlos Gonzalez-Moreno. 2010. A technique for defining agent-oriented engineering processes with tool support. *Engineering Applications of Artificial Intelligence* 23(3): 432-444, DOI: 10.1016/j.engappai.2009.08.004

[15] S. Demirkol, M. Challenger, S. Getir, T. Kosar, G. Kardas, M. Mernik. 2012. SEA_L: A Domain-specific Language for Semantic Web enabled Multi-agent Systems. 2nd Workshop on Model Driven Approaches in System Development at FedCSIS 2012, pp. 1373-1380

[16] G. Ciobanu, C. Juravle. 2012. Flexible Software Architecture and Language for Mobile Agents. *Concurrency and Computation-Practice & Experience* 24(6): 559-571, DOI: 10.1002/cpe.1854

[17] J. M. Gascuena, E. Navarro, A. Fernandez-Caballero. 2012. Model-Driven Engineering Techniques for the Development of Multi-agent Systems. *Engineering Applications of Artificial Intelligence* 25(1): 159-173, DOI: 10.1016/j.engappai.2011.08.008

[18] S. Demirkol, M. Challenger, S. Getir, T. Kosar, G. Kardas, M. Mernik. 2013. A DSL for the development of software agents working within a semantic web environment. *Computer Science and Information Systems* 10(4): 1525-1556, DOI: 10.2298/CSIS121105044D

[19] M. Challenger, S. Demirkol, S. Getir, M. Mernik, G. Kardas, T. Kosar. 2014. On the use of a domain-specific modeling language in the development of multiagent systems. *Engineering Applications of Artificial Intelligence* 28: 111-141, DOI: 10.1016/j.engappai.2013.11.012

[20] E. J. T. Goncalves, M. I. Cortes, G. A. L. Campos, Y. S. Lopes, E. S. S. Freire, V. T. da Silva, K. S. F. de Oliveira, M. A. de Oliveira. 2015. MAS-ML2.0: Supporting the modelling of multi-agent systems with different agent architectures. *Journal of Systems and Software* 108: 77-109. DOI: 10.1016/j.jss.2015.06.008

[21] M. Mernik, J. Heering, A. Sloane. 2015. When and how to develop domain-specific languages. *ACM Computing Surveys* 37(4): 316-344, DOI: 10.1145/1118890.1118892

[22] M. Joao Varanda Pereira, M. Mernik, D. Da Cruz, P. R. Henriques. 2008. Program Comprehension for Domain-specific Languages. *Computer Science and Information Systems* 5(2): 1-17, DOI: 10.2298/CSIS0802001P

[23] I. Lukovic, M. Joao Varanda Pereira, N. Oliveira, D. Carneiro da Cruz, P. R. Henriques. 2011. A DSL for PIM specifications: Design and attribute grammar based implementation, *Computer Science and Information Systems* 8(2): 379-403, DOI: 10.2298/CSIS101229018L

[24] B. Selic. 2003. The pragmatics of model-driven development. *IEEE Software* 20: 19-25, DOI: 10.1109/MS.2003.1231146

[25] K. Sycara, M. Paolucci, A. Ankolekar, N. Srinivasan. 2003. Automated discovery, interaction and composition of Semantic Web Services. *Journal of Web Semantics*, 1(1): 27-46, DOI: 10.1016/j.websem.2003.07.002

[26] S. Getir, M. Challenger, G. Kardas. 2014. The formal semantics of a domain-specific modeling language for semantic web enabled multi-agent systems. *International Journal of Cooperative Information Systems* 23(3): 1-53, DOI: 10.1142/S0218843014500051

[27] C. Hahn, S. Nesbigall, S. Warwas, I. Zinnikus, K. Fischer, M. Klusch. 2008. Integration of Multiagent Systems and Semantic Web Services on a Platform Independent Level. IEEE/WIC/ACM Int'l Conf. on Web Intelligence and Intelligent Agent Technology, pp. 200-206

[28] S. Warwas, C. Hahn. 2008. The concrete syntax of the platform independent modeling language for multiagent systems. Agent-based Technologies and applications for enterprise interoperability

[29] G. Kardas, M. Challenger, S. Yildirim, A. Yamuc. 2012. Design and implementation of a multiagent stock trading system. *Software: Practice and Experience* 42(10): 1247-1273, DOI: 10.1002/spe.1137

[30] S. Rougemaille, F. Migeon, C. Maurel, M-P. Gleizes. 2007. Model Driven Engineering for Designing Adaptive Multi-agent Systems, *Lecture Notes in Artificial Intelligence* 4995: 318-333, DOI: 10.1007/978-3-540-87654-0_18

[31] M. Challenger, G. Kardas, B. Tekinerdogan. 2016. A systematic approach to evaluating domain-specific modeling language environments for multi-agent systems. *Software Quality Journal*, DOI: 10.1007/s11219-015-9291-5

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1565–1571
DOI: 10.15439/2016F530
ACSIS, Vol. 8. ISSN 2300-5963

Development of Human-friendly Notation for XML-based Languages

Sergej Chodarev
Technical University of Košice, Department of Computers and Informatics, Letná 9, Košice, Slovakia
Email: sergej.chodarev@tuke.sk

***Abstract*—XML is a popular choice for development of domain-specific languages. In spite of its popularity, XML is a poor user interface and a lot of languages can be improved by introducing custom notation. This paper presents an approach for development of custom human-friendly notation for existing XML-based language together with a translator between the new notation and XML. This approach is based on explicit representation of language abstract syntax that can be decorated with mappings to both XML and the custom notation. The approach supports iterative design and development of the language concrete syntax, allowing its modification based on users feedback. Development process is demonstrated on a case study of language for definition of graphical user interface layout.**

I. Introduction

XML is very common and easy to parse generic language, it is well supported by existing tools and technologies and therefore it is a popular basis for domain-specific languages (DSLs). While XML is appropriate choice in many cases, especially for program-to-program communication, it is not well suited for cases, where humans need to manipulate documents. Although they are able to create, modify and read XML documents, it is not a pleasurable experience, because of uniformity and syntactic noise that makes it difficult to find useful information visually [1].

While a more appropriate syntax can be chosen for development of new languages, a lot of languages was already implemented based on XML and their reimplementation would be complicated and time-consuming. One of the possible ways to solve this problem is to develop a translator that would read documents written in a specialized human-friendly notation and output them in the XML for further processing using existing tools. Ideally, the new notation would be specifically tailored to the domain of the language as is usual for DSLs [2].

Development of the translator requires implementation of parser and generator. Proper separation of these two components also involves some internal representation of the language that would be created by the parser and then traversed to generate the XML. Development of all these components may be very tedious, even using parser generators.

This paper presents an approach to development of the translator that simplifies the process and allows to evolve the new syntax iteratively. The main idea of the approach is in extracting definition of language structure into a format that can be easily augmented with the definition of a new notation. For example, Java classes representing the structure of an XML-based language can be generated automatically from the XML Schema using JAXB[1]. The generated classes are already annotated in a way that allows automatic marshalling and unmarshalling their instances in the XML form. Additional annotations can be added to the classes that define their mapping to a different textual notation. In the next step an annotation based parser generator, like YAJCo [3], can be used to generate a parser for the new notation. Connecting the parser with the XML unmarshaller one would get a complete translator from a custom human-friendly notation to the original XML-based.

Main topics discussed in the paper and its contributions are the following:

- It explains the approach to language translator development that is based on explicit representation of language *abstract syntax* in a format that allows attaching definitions of different concrete notations (Section II).
- The approach allows to develop a round-trip translator based on a single specification of abstract syntax. This enables *iterative development* of the notation in contrast to classical approach where complete syntax should be defined upfront. The process of iterative notation development is described in Section III.
- The whole approach is demonstrated on a case study of a language for specifying layout and properties of graphical user interface components (Section IV). The case study shows possible challenges of the approach and can be used as a guide to develop similar translators.

Presented case study also demonstrates that object-oriented programming language like Java can be successfully used as a format for abstract syntax description, provided that it allows attaching structured meta-data [4] (known as annotations or attributes) to program elements. This allows to use numerous existing tools and also avoids the need for special purpose representations and related technologies.

II. Model-driven Development of Language Translator

Similarly to model-driven software development [5] it is possible to drive development of the language translator by the model of the language – metamodel[2]. The metamodel defines language concepts with their properties and relations to other

[1]Java Architecture for XML Binding, available at https://jaxb.java.net/.

[2]If we consider documents written in a language to be *models*, then a model of the language itself is *metamodel*.

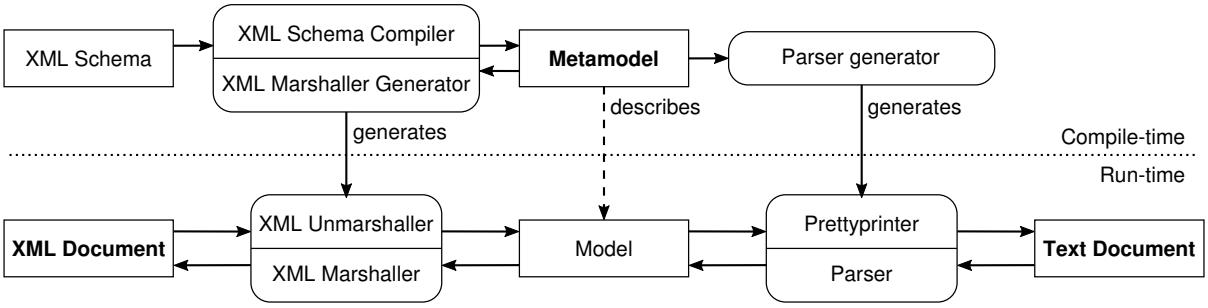


Fig. 1. Model-driven language translator development (arrows represent data-flow)

concepts. It can be annotated with additional information about concrete syntaxes of the language that need to be translated.

Figure 1 shows the whole architecture of model-driven language translator development in the case of translating XML to textual notation and vice versa. The metamodel augmented with definition of concrete notations is the central element. It is used as an input to generate parser and prettyprinter for both the textual notation (using parser generator) and XML (using XML marshaller generator). The generated tools can be connected into a pipeline that handles translation of one notation to the other with the internal representation of the model (defined by the metamodel) as an intermediate format.

What is important, the first version of the metamodel itself can be retrieved from the existing description of XML-based language – XML Schema. This allows to significantly shorten the development process, because large part of the language definition – its abstract syntax – is derived automatically. This style of development also follows the "Single Point of Truth" principle, because the structure of the language is defined only once and its mappings to concrete notations are attached to it.

The described approach does not depend on concrete tools. It, however, requires an XML marshaller and a parser/prettyprinter generator that both use the same format for metamodel specification. In Section IV is presented the case study that uses Java classes to represent the metamodel. They are augmented using annotations, JAXB is used as an XML marshaller and YAJCo as a parser generator. Alternative solution can use Ecore from Eclipse Modeling Framework (EMF) [6] to represent the metamodel and Xtext [7] as a parser generator.

III. Iterative Development of the Translator

Design of notation for an existing language is, actually, design of a user interface. As such, it requires evaluation of various alternatives, and testing new alternatives in conditions similar to real-life. This process is iterative by nature [8]. On the other hand, classical approach to language development assumes that the language syntax is designed upfront (for example [9]). A complete specification of grammar is then augmented with semantic actions and processed to generate a parser. Changes in the syntax often require modification of semantic rules, making the process laborious.

The fact, that the model-driven approach described above allows to easily receive bidirectional translator, makes it possible to use a different process:

1) Extract language metamodel from the XML Schema.
2) Augment the metamodel with initial definition of the new concrete syntax.
3) Generate a prettyprinter based on the definition and convert examples of existing XML documents to the new notation.
4) Evaluate the new notation on examples of converted documents.
5) If the notation is not satisfactory, modify concrete syntax definition and go back to the step 3.
6) If the notation is satisfactory, complete the syntax definition and generate a parser.

This process allows to easily use existing documents for testing new notation instead of some artificial examples. Complete real-life documents in the new notation can be generated automatically immediately after the definition of the syntax has changed. This allows very fast evaluation and modification cycles, so problems in the notation can be spotted and resolved, even if they occur only in complex documents.

This approach also provides a simple method for testing correctness of the developed translator, i.e. that no information is lost or corrupted during translation. A set of example XML documents can be automatically converted to the new notation and then back to the XML. Result of the conversion can be compared with the original XML documents to reveal missing support for some language features or other errors. If the translator is correct, no data is lost and documents are identical (except of differences in formatting that can be removed using normalization before the comparison).

IV. Case Study

The approach can be demonstrated on the development of a new textual notation for the GtkBuilder language. GtkBuilder is a part of the GTK+ GUI toolkit that allows to declaratively specify layout of a user interface using an XML-based language[3]. There is a Glade tool[4] that allows to edit GtkBuilder specifications visually, however it tends to lack support for

[3]Specified at https://developer.gnome.org/gtk3/stable/GtkBuilder.html
[4]Available at https://glade.gnome.org/

```
1  <interface>
2    <object class="GtkDialog" id="dialog1">
3      <child internal-child="vbox">
4        <object class="GtkVBox" id="vbox1">
5          <property name="border-width">10</property>
6          <child internal-child="action_area">
7            <object class="GtkHButtonBox" id="hbuttonbox1">
8              <property name="border-width">20</property>
9              <child>
10               <object class="GtkButton" id="save_button">
11                 <property name="label" translatable="yes">Save</property>
12                 <signal name="clicked" handler="save_button_clicked"/>
13               </object>
14             </child>
15           </object>
16         </child>
17       </object>
18     </child>
19   </object>
20 </interface>
```

Fig. 2. Example of user interface definition using XML notation

newest GTK+ widgets, requiring manual modification of XML files.

The translator was implemented using two tools: JAXB and YAJCo. JAXB is a standard solution for marshalling and unmarshalling Java objects to XML. YAJCo[5] (Yet Another Java Compiler Compiler) is a parser generator for Java that allows to specify language syntax using a metamodel in a form of annotated Java classes [3]. This allows declarative specification of a language and its mapping to Java objects [10]. In addition to parser, YAJCo is able to generate pretty-printer and other tools from the same specification [11].

This section describes a process of development of the translator using the chosen tools. It also explains challenges that arise during the implementation and their solutions. Readers can use it as a guide to develop their own translators. The complete source code of the translator is available for download at http://hron.fei.tuke.sk/~chodarev/gtkbuilder/.

A. GtkBuilder Language

The GtkBuilder UI definition language allows to specify a layout of widgets forming a user interface and their properties using an XML notation. Each instance of a widget is defined using an *object* element, which contains its type, identifier, properties, signal bindings, and child objects. Fig. 2 presents an example UI definition in the XML notation.

The XML notation for the language, while familiar, is very hard to read. Document contains a lot of syntactic noise that makes fast scanning of the definition very hard. The same definition can be expressed using a custom notation as shown in Fig. 3. The notation uses special symbols to provide concise representation for language elements. For example, *object* is expressed using "`[ Class id ... ]`" notation (e.g. line 1), properties are written simply as pairs in a form

[5]Available at https://github.com/kpi-tuke/yajco

```
1  [ GtkDialog dialog1
2  %child vbox :
3     [ GtkVBox vbox1
4     border-width : 10
5     %child action_area :
6        [ GtkHButtonBox hbuttonbox1
7        border-width : 20
8        %child :
9           [ GtkButton save_button
10          label : _ Save
11          clicked -> save_button_clicked ]]]]
```

Fig. 3. Example of user interface definition using custom textual notation

"`name : value`" (e.g. line 4), signal binding is expressed as "`signal_name -> handler`" (line 11), and strings that should be translated in localized versions of UI are marked with underscore (line 10). The notation is short and quite intuitive at the same time.

In the rest of the section the development of the custom notation and conversion tools is described in more detail.

B. Metamodel Extraction

As was mentioned earlier, the metamodel represented by Java classes can be generated based on the existing XML schema using the XML binding compiler (`xjc`) that is a part of JAXB. It generates Java classes corresponding to elements of XML-based language. Generated classes contain annotations that define mapping to XML elements and attributes. JAXB uses these annotations to create instances of the classes and set their properties based on XML document contents. The same annotations are used to serialize objects to the XML form.

This means that after the metamodel was extracted it is possible to use JAXB to read existing UI definition from the XML notation to an internal representation defined by the

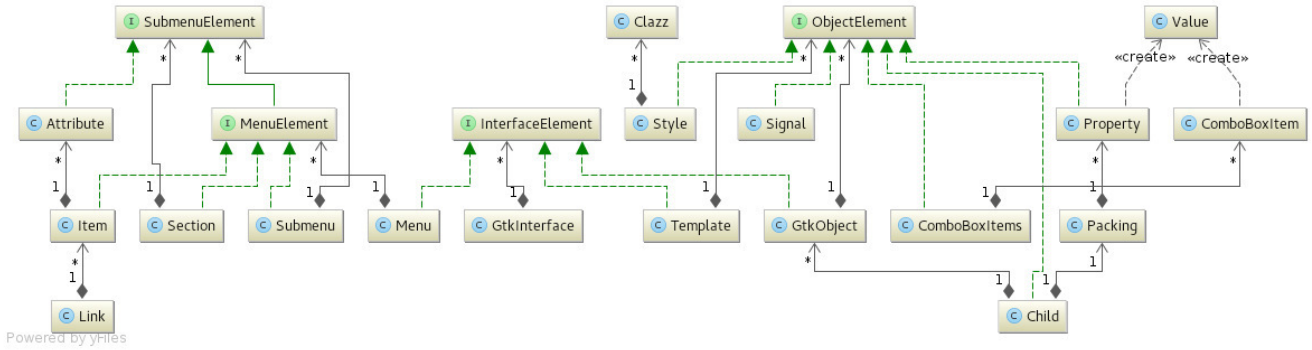


Fig. 4. Class diagram of the GtkBuilder language metamodel

metamodel and also to marshall the internal model back to the XML form.

In the case study, extracted classes directly corresponded to elements of the XML-based language. Therefore, they included classes like *Interface*, *Object*, *Child*, *Property*, *Signal*, classes for definition of menus, etc. In total, 13 classes was generated by JAXB. Full metamodel, including modifications and additions described in next sections is depicted in Fig. 4.[6]

Encountered problems: Unfortunately, the schema of the GtkBuilder language is available only in the RelaxNG format. Because the support for RelaxNG schemas in JAXB is only experimental, it was converted to the XML Schema format using the Trang tool[7].

In addition, the schema does not define the language completely. Each widget type can support additional elements for widget-specific functionality. These elements, however, are not specified in the schema. Instead, arbitrary elements are allowed inside the *object* element.

The support for the most common widget-specific extensions, that was not specified in the schema, was added later by defining new classes in the metamodel. Generated classes obviously need to be modified to include declaration of added child elements of new types.

Shortcomings of the GtkBuilder language definition make it impossible to create the metamodel fully automatically. But on the other hand, it shows that the approach is applicable even in such cases.

C. Syntax Definition

Definition of new concrete syntax is provided in form of annotations added to the metamodel classes. This means that the metamodel generated using JAXB needs to be augmented to include YAJCo-specific annotations.

YAJCo infers abstract syntax of the language from the inheritance relations between the metamodel classes and from their constructors. Each constructor is transformed into a grammar rule and parameters of a constructor determine the

[6]Classes *Object* and *Interface* was renamed to *GtkObject* and *GtkInterface* to avoid clashes with Java keywords in the generated parser.

[7]Available at http://www.thaiopensource.com/relaxng/trang.html

```
@Before("%child")
public Child(@Token("ID") @After(":")
             String internalChild,
             @NewLine @Indent
             List<GtkObject> object) {
    this.object = object;
    this.internalChild = internalChild;
}
```

Fig. 5. Example class constructor with YAJCo annotations

right hand side of the rule. YAJCo annotations are attached to constructors to specify details of the grammar that cannot be inferred automatically. For example, Fig. 5 presents one of the constructors of the *Child* class. It defines that a child can be constructed from a string representing an internal child name and a list of objects (e.g. lines 2 and 5 in Fig. 3). The child definition would start with the "`%child`" token followed by the `ID` token representing an identifier, followed by colon and a sequence of objects. Annotations also contain hints on indentation and new-line placement for prettyprinter (they are ignored by the parser).

Such constructors need to be added to the metamodel classes. Each variation of the element concrete syntax requires its own constructor. For example, the *Child* can be defined with *internalChild* property specified, or without it (e.g. line 8 in Fig. 3) and therefore it needs at least two constructors. In addition to constructors, factory methods can be used as an annotation target. This makes it possible to define different syntaxes even if they have the same types of parameters in Java.

Each class also needs a non-parametrized constructor required by JAXB. This constructor must be marked using the YAJCo `@Exclude` annotation so it would be ignored by the YAJCo tool.

In the following subsections are described some details of the implementation, typical problems and their solutions.

1) Completing the abstract syntax specification: In some cases several alternative values of different types are expected in the same place. For example, *object* definition contains a sequence of properties, child definitions or signal bindings. In

```
public class GtkObject {
    @XmlElements({
        @XmlElement(name = "property",
                    type = Property.class),
        @XmlElement(name = "signal",
                    type = Signal.class),
        @XmlElement(name = "child",
                    type = Child.class)
    })
    protected List<java.lang.Object>
        propertyOrSignalOrChild;
    ...
```

Fig. 6. Alternative types of values by as defined by JAXB

```
public class GtkObject {
    @XmlElements( ... )
    protected List<ObjectElement>
        propertyOrSignalOrChild;
    ...

public interface ObjectElement {}

public class Property implements ObjectElement{
    ...
public class Signal implements ObjectElement {
    ...
public class Child implements ObjectElement {
    ...
```

Fig. 7. Alternative types of values defined using inheritance

object-oriented model this situation can be expressed by inheritance. JAXB, however, does not use this technique in generated metamodel classes. Instead, it uses type *java.lang.Object* in the container and adds `@XMLElements` annotation to specify all possible concrete types that can be used as is shown in Fig. 6.

On the other hand, YAJCo requires the use of inheritance or implementation relations in these situations. So a new marker interface needs to be created and classes of all elements that can appear in specific context are marked to implement it. The container class is then modified to reference the marker interface. An example of all these modifications is presented in Fig. 7.

2) Conflict between reserved keywords and identifiers: The problem arises from the different treatment of keywords in different notations. XML uses special syntax for language elements (tags delimited by angle brackets) and therefore it can allow to use language keywords as identifiers inside XML attributes and text fragments. For example, menu can be named simply "menu": `<menu id="menu">...</menu>`

On the other hand, if element names, like "`menu`" or "`child`", become reserved keywords in the custom notation, they could not be used as identifiers anymore, because standard lexical analyzer would not be able to distinguish them. Such conflicts can be resolved by decorating either language keywords or identifiers with some special symbols that would distinguish them. In our case percent sign was used as a starting symbol of all language keywords. For example, the menu would be defined like this: `%menu menu { ... }`

3) Model transformations: Representation of the metamodel using Java classes allows to implement simple model transformations using constructors. Constructors of the metamodel classes can transform their parameters before storing to class fields. It makes possible to define helper classes with own syntax rules, that are not stored in the model.

For example, at different places in the language it is possible to specify value, that can be a number, a symbol, or a string, all with different notations. Each class where such value can be used would need at least three constructors (for each notation of the value). Instead of this, it is possible to define a new helper class that would handle this aspect of concrete syntax using its own constructors and factory methods. It would also implement appropriate pre-processing of the values (e.g. removing quotation marks from strings). Instances of the helper class would become constructor parameters of the original classes, but only the actual value would be stored in the model, not the instance of the helper class. This allows to avoid modification of the metamodel and XML bindings.

D. Other Implementation Notes

1) Project setup: It is useful to split the project into two submodules: one for the metamodel definition and the code generated based on it, and other for code that uses the generated parser and prettyprinter to translate language sentences. This setup explicitly divides generated code and the code that depends on it.

The second submodule contains implementation of a command-line tool for converting between different notations of the language. This tool instantiates JAXB marshaller and unmarshaller and also YAJCo generated parser and prettyprinter and uses them to produce the internal model from one notation and convert it to the other notation.

In addition, the project contains script that tests the implementation by running round-trip transformation and comparing results with original versions of example documents. In the case study this script was implemented as a *Makefile* that would produce report on differences between documents if modifications of the code would cause errors in translation. This approach helped to find several problems described in this section and led to successful translation of tested examples.

2) Prettyprinter customization: Some syntax constructs can not be handled by the YAJCo generated prettyprinter automatically. For example, strings that should be translated in localized versions of the application are marked using an underscore "_" symbol. In the model, however, it is stored as a value of *True* in the field *translatable*. This correspondence is not inferred by the prettyprinter generator. As a solution, a prettyprinter can be simply extended by a new class, that would override the corresponding method to provide the needed functionality. This is greatly simplified by the fact that the generated prettyprinter is based on the visitor pattern.

V. Related Work

The presented approach and technologies are not limited to development of textual notation for the language. As was shown in the work of Bačíková et al. [12], it is possible to use the same metamodel definition to generate a graphical user interface. This interface would consist of forms allowing to edit language sentences. Input for the metamodel extraction is not limited to XML Schema: it is possible to extract metamodel from some non-XML notation [13], existing application [14] or it user interface [15]. It is also possible to avoid modification of generated metamodel code to augment definition of the metamodel and add different methods of its processing by using aspect-oriented programming [16].

The most similar work to the one presented in this paper is XMLText by Neubauer et al. [17]. They use EMF for representing metamodels and Eclipse Xtext [7] for generating parser, prettyprinter (*serializer* in the Xtext terminology), and editing support for the Eclipse integrated development environment. They integrate these tools and develop round-trip transformation between XML based languages defined by XML Schema and textual notation. Their tool, however, does not directly support custom syntax definition for each language element. On the other hand, customization of the textual notation should be possible using manual modification of the generated Xtext grammar.

Therefore, it should be possible to use the iterative approach described in this paper with EMF and Xtext as well. The main difference compared to technologies presented in this paper is the fact that EMF and Xtext use specialized language for defining metamodel — Ecore, while JAXB and YAJCo rely on Java for this purpose. This allows to lower the entry barrier by minimizing the amount of new technologies needed to be learned. It also allows to implement model transformations in Java using the techniques well-known by industrial programmers. On the other hand EMF promises independence on the concrete programming language. Together with Xtext they also provide a more mature platform for development of languages with their tooling, first of all – editing environment.

A real-life example of migrating UML and XML based modeling language to these technologies was presented by Eysholdt and Rupprecht [18]. They, however, did not use a single metamodel for different notations. Instead, they used model-to-model transformations to migrate models.

Other alternative would be the use of different generic language instead of the XML. YAML is a popular choice, for example, Shearer [19] used it to provide textual representation for ontologies. YAML (Yet Another Markup Language) was specially designed as a human-friendly notation for expressing data structures [20]. Its syntax is readable and quite simple, but the use of generic language does not allow to use specialized short-hand notations tailored for a developed language. While the basic structure of our example language may be expressed similar to the custom notation, problems start in the details. For example, the custom notation allows to mark any string as translatable by simply writing underscore before it, YAML would require a different and more noisy solution.

Similar solution is the use of OMG HUTN (Human-Usable Textual Notation) which specifies generic textual notation for MOF (Meta-Object Facility) based metamodels [21], again without possibility to customize concrete syntax.

The approach presented in this paper is also similar to tools supporting development of DSLs based on existing ontologies [22], [23]. In our case, however, existing XML-based language is used as a basis for a DSL instead of ontology.

VI. Conclusion

Presented case study showed the applicability of the model-driven translator development approach. It also allowed to formulate several advises for practical usage of the approach (described in Section IV). While most of them are specific to the tools used in the study, some may be applicable to other tools as well. The approach itself is tool-independent and can be used with any language metamodel representation that can be mapped to both XML and custom textual syntax.

Future work may include identification, validation and comparison of tools and metamodel representations that support the described translator development approach. The YAJCo tool itself requires further development, especially in the area of generating tool support for the language beside parser and prettyprinter.

Acknowledgment

This work was supported by project KEGA No. 047TUKE-4/2016 “Integrating software processes into the teaching of programming”.

References

[1] T. Parr, “Humans should not have to grok XML,” 8 2001. [Online]. Available: http://www.ibm.com/developerworks/library/x-sbxml/index.html

[2] M. Mernik, J. Heering, and A. M. Sloane, “When and how to develop domain-specific languages,” *ACM Computing Surveys*, vol. 37, no. 4, pp. 316–344, dec 2005. doi: 10.1145/1118890.1118892

[3] J. Porubän, M. Forgáč, M. Sabo, and M. Běhálek, “Annotation based parser generator,” *Computer Science and Information Systems (ComSIS)*, vol. 7, no. 2, pp. 291–307, 2010. doi: 10.2298/csis1002291p

[4] M. Nosáľ, M. Sulír, and J. Juhár, “Source code annotations as formal languages,” in *2015 Federated Conference on Computer Science and Information Systems (FedCSIS)*, Sept 2015. doi: 10.15439/2015F173 pp. 953–964.

[5] T. Stahl, M. Voelter, and K. Czarnecki, *Model-Driven Software Development: Technology, Engineering, Management*. John Wiley & Sons, 2006. ISBN 0470025700

[6] D. Steinberg, F. Budinsky, E. Merks, and M. Paternostro, *EMF: Eclipse Modeling Framework*. Pearson Education, 2008.

[7] S. Efftinge and M. Völter, “oAW xText: A framework for textual DSLs,” in *Workshop on Modeling Symposium at Eclipse Summit*, vol. 32, 2006, p. 118.

[8] J. Nielsen, “Iterative user-interface design,” *IEEE Computer*, vol. 26, no. 11, pp. 32–41, Nov 1993. doi: 10.1109/2.241424

[9] A. V. Aho, M. S. Lam, R. Sethi, and J. D. Ullman, *Compilers: Principles, Techniques, and Tools (2Nd Edition)*. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 2006. ISBN 0321486811

[10] D. Lakatoš, J. Porubän, and M. Bačíková, “Declarative specification of references in DSLs,” in *2013 Federated Conference on Computer Science and Information Systems (FedCSIS)*. IEEE, 2013. ISBN 9781467344715 pp. 1527–1534.

[11] D. Lakatoš and J. Porubän, "Generating tools from a computer language definition," in *Proceedings of International Scientific conference on Computer Science and Engineering (CSE 2010)*, September 2010, pp. 76–83.

[12] M. Bačíková, D. Lakatoš, and M. Nosáľ, "Automatized generating of GUIs for domain-specific languages," in *CEUR Workshop Proceedings*, vol. 935, 2012, pp. 27–35.

[13] J. Porubän, J. Kollár, and M. Sabo, "Abstraction of computer language patterns: The inference of textual notation for a dsl," in *Formal and Practical Aspects of Domain-Specific Languages: Recent Developments*. IGI Global, 2012, pp. 365–385, doi: 10.4018/978-1-4666-2092-6.ch013.

[14] J. Kollár and M. Vagač, "Aspect-oriented approach to metamodel abstraction," *Computing and Informatics*, vol. 31, no. 5, pp. 983–1002, 2012.

[15] M. Bačíková, J. Porubän, S. Chodarev, and M. Nosáľ, "Bootstrapping DSLs from user interfaces," in *Proceedings of the 30th Annual ACM Symposium on Applied Computing - SAC '15*. ACM Press, apr 2015. doi: 10.1145/2695664.2695994. ISBN 9781450331968 pp. 2115–2118.

[16] J. Porubän, M. Sabo, J. Kollár, and M. Mernik, "Abstract syntax driven language development: Defining language semantics through aspects," in *Proceedings of the International Workshop on Formalization of Modeling Languages (FML '10)*. New York, NY, USA: ACM, 2010. doi: 10.1145/1943397.1943399. ISBN 978-1-4503-0532-7

[17] P. Neubauer, A. Bergmayr, T. Mayerhofer, J. Troya, and M. Wimmer, "XMLText: from XML schema to Xtext," in *2015 ACM SIGPLAN International Conference on Software Language Engineering*. ACM, oct 2015. doi: 10.1145/2814251.2814267. ISBN 978-1-4503-3686-4 pp. 71–76.

[18] M. Eysholdt and J. Rupprecht, "Migrating a large modeling environment from xml/uml to xtext/gmf," in *Proceedings of the ACM International Conference Companion on Object Oriented Programming Systems Languages and Applications Companion*, ser. OOPSLA '10. New York, NY, USA: ACM, 2010. doi: 10.1145/1869542.1869559. ISBN 978-1-4503-0240-1 pp. 97–104.

[19] R. Shearer, "Structured ontology format," in *Proceedings of the OWLED 2007 Workshop on OWL: Experiences and Directions*, 2007.

[20] O. Ben-Kiki, C. Evans, and B. Ingerson, "YAML Ain't Markup Language. Version 1.2," Tech. Rep., 2009. [Online]. Available: http://yaml.org/

[21] P.-A. Muller and M. Hassenforder, "HUTN as a Bridge between ModelWare and GrammarWare - An Experience Report," *WISME Workshop, MODELS/UML*, pp. 1–10, 2005.

[22] I. Čeh, M. Črepinšek, T. Kosar, and M. Mernik, "Ontology driven development of domain-specific languages," *Computer Science and Information Systems (ComSIS)*, vol. 8, no. 2, pp. 317–342, 2011. doi: 10.2298/CSIS101231019C

[23] J. M. S. Fonseca, M. J. V. Pereira, and P. R. Henriques, "Converting Ontologies into DSLs," in *3rd Symposium on Languages, Applications and Technologies*, ser. OpenAccess Series in Informatics (OASIcs), vol. 38. Dagstuhl, Germany: Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2014. doi: http://dx.doi.org/10.4230/OASIcs.SLATE.2014.85. ISBN 978-3-939897-68-2. ISSN 2190-6807 pp. 85–92.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1573–1576
DOI: 10.15439/2016F524
ACSIS, Vol. 8. ISSN 2300-5963

Preliminary Report on Empirical Study of Repeated Fragments in Internal Documentation

Milan Nosáľ, Jaroslav Porubän
Department of Computers and Informatics,
Faculty of Electrical Engineering and Informatics,
Technical University of Košice
Letná 9, 042 00, Košice, Slovakia
Email: milan.nosal@gmail.com, jaroslav.poruban@tuke.sk

***Abstract*—In this paper we present preliminary results of an empirical study, in which we used copy/paste detection (PMD CPD implementation) to search for repeating documentation fragments. The study was performed on 5 open source projects, including Java 8 SDK sources. The study shows that there are many occurrences of copy-pasting documentation fragments in the internal documentation, e.g., copy-pasted method parameter description. Besides these, many of the copy-pasted fragments express some domain or design concern, e.g., that the method is obsolete and deprecated. Therefore the study indicates that the cross-cutting concerns are present in the internal documentation in form of documentation phrases.**

I. Introduction

PRESERVING and comprehending developer's concerns (intents) in the source code is still a current challenge in software development [1], [2], [3], [4]. In this paper we analyze the internal documentation (source code comments, JavaDoc, etc.) to recognize repeating documentation fragments that document those concerns (or features [5]). Our research question for this work is: *Does internal documentation contain significant duplication?* To answer this question we performed a copy/paste detection study, in which we analyzed JavaDoc comments in 5 open source projects. In this report we present preliminary results that indicate that there is a significant duplication of text in internal documentation. These repeating documentation fragments constitute documentation phrases discussed in several works – e.g., our previous work [6], or the one by Horie et al. [7]. This way the study has a potential to highlight the importance of those works and it may stimulate further attention to this topic.

II. Documentation Phrases

A documentation phrase is a set of documentation fragments with the same or similar formulation (part of sentence, sentence, a set of sentences) that can be found across the software system or even across multiple systems. Documentation fragments that represent the same documentation phrase usually document the same domain or design property that is shared by the documented program elements [6]. Horie et al. [7] likened the documentation phrases to crosscutting concerns from aspect oriented programming (AOP [8]).

As an example we can use the Swing library for component graphical user interfaces in Java. The library is not thread safe and therefore the programmer has to pay extra caution when using it in multithreaded systems (she has to use Event Dispatch Thread to safely work with the Swing components). Swing JavaDoc documents it and for each affected class includes a warning (see JPanel documentation in Figure 1).

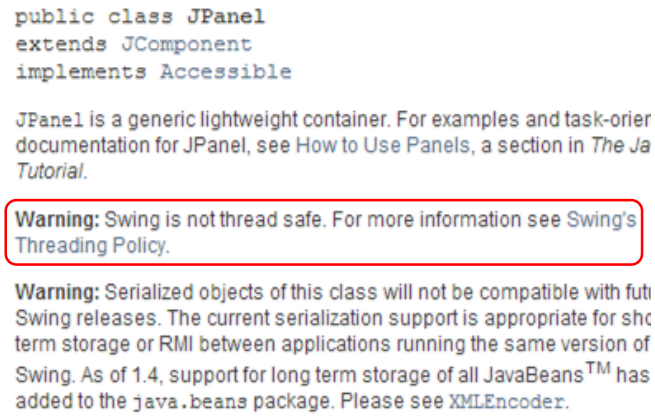


Fig. 1. A NotThreadSafe documentation phrase instance in the Swing JPanel documentation

For this warning to be included in the JavaDoc, its HTML snippet has to be copy-pasted in the JavaDoc comment of each affected class.

III. Documentation Copy/paste Study

In this study we analyzed the internal documentation of several Java frameworks and libraries to detect currently existing documentation phrases. In order to find documentation phrases, we performed a copy/paste detection[1] on the documentation.

We have modified the PMD Copy/Paste Detection (CPD) tool[2] to support copy/paste detection on JavaDoc documentation. The tool was fed preprocessed sources of several libraries and open source projects in Java and it analysed them to detect duplications in documentation that would indicate the potential

[1]Copy/paste detection is usually used to search for code that needs to be refactored, or to detect plagiarism [9].

[2]http://pmd.sourceforge.net/

Algorithm 1 JavaDoc preprocessing example – before

```
package org.tuke;

/**
 * Dummy class.
 * Created by Milan on 5.3.2016.
 */
public class DummyClass {
}
```

Algorithm 2 JavaDoc preprocessing example – after

```
_
_
_
Dummy class.
Created by Milan on 5.3.2016.
 DummyClass.jdoc
     .6.1394135902525.0.-2042003928.
_
_
```

documentation phrases. We will discuss the process phases in more details in following sections.

A. Java Sources Preprocessing

In our experiment we used the tool to detect simple non-parametrized documentation phrases in the JavaDoc documentation. However, the PMD CPD was designed to be a code analysis tool and as such its purpose was to detect duplication in programming languages. Documentation phrases are fragments of documentation in a natural language.

PMD CPD tool works with language lexical tokens that it compares to detect duplications in their usage. To reduce tokenization complexity we pre-processed the sources to remove all the characters and tokens that are not the documentation. In other words the preprocessed sources are source files with only JavaDoc comments in their comments. Let us consider a simple class with JavaDoc from Listing 1.

This source file would be transformed to the content presented in Listing 2[3]

Java lexical tokens were discarded along with asterisks indicating that following lines are part of JavaDoc (lines are preserved for backtracking to original sources). At the end of each JavaDoc comment we added a randomly generated unique "anchor" (DummyClass.jdoc.6.1394135902525 .0.-2042003928. in the example) that prevented detection of duplicates spanning multiple comments.

B. PMD CPD Modification

PMD CPD tool uses a tokenizer to read files and obtain lexical tokens of the language. In our experiment we used our custom tokenizer that divided the preprocessed files into sentences. Each sentence in the file was a single token. If we consider the `DummyClass` example from section III-A, the tokenizer would return following three tokens:

- "Dummy class."
- "Created by Milan on 5.3.2016."
- "DummyClass.jdoc.6.1394135902525.0.-2042003928."

Separators for the tokenization were characters '.', '?', and '!' followed by a whitespace character (therefore the date in the second token from the example was not divided in multiple tokens), or a new line character followed by an empty line.

C. Document Phrases Detection

For the duplication detection process we used the standard PMD CPD implementation (according to http://pmd.sourceforge.net/pmd-4.3.0/cpd.html they use Karp-Rabin string matching algorithm). We registered our modification in `LanguageFactory` and `GUI` classes and used the graphical user interface provided by the `GUI` class to run the tool.

In the setup of the copy/paste detection we set the 'Report duplicate chunks larger than:' option to a single token. This way PMD CPD reported even duplication of a single token – a single line in the documentation. The results were serialized as XML so that we could use XPath with XSLT to process them. First post-processing removed all the results that did not have at least 4 duplications – we considered 4 instances of a documentation phrase a reasonable threshold for considering it a significant documentation phrase.

IV. Results

We performed the experiment on the following open source Java projects:

- sources of Java 8 standard edition[4] with 7703 source files,
- PicoContainer[5] with 1067 source files,
- JasperReports library[6] with 2834 source files,
- JoSQL[7] (SQL for Java Objects) with 85 source files, and
- jEdit[8] with 573 source files.

A. Copy/paste Detection Results

The PMD CPD tool discovered 6102 duplicated fragments of various lengths in Java 8 source code. 2221 of them were duplicated fragments that had 4 instances. The highest number of instances of a single duplicated fragment was 344. Second highest were two duplicated fragments both with 299 instances. Figure 2 shows an overview of obtained results. We will provide a more detailed analysis of these data below.

PicoContainer contained 70 duplicated fragments ranging from 36 duplicated fragments with 4 instances to a single fragment duplicated 634 times. Figure 2 presents results for PicoContainer in a simple chart. Closer inspection of the results showed that duplicated fragments with the highest

[3]We used underscores here to highlight empty lines.

[4]http://www.oracle.com/technetwork/java/javase/downloads/java-archive-javase8-2177648.html, JDK version 8u20

[5]https://github.com/picocontainer/picocontainer, commit `0f8172b`

[6]http://sourceforge.net/projects/jasperreports/files/jasperreports/, version 6.0.0

[7]http://sourceforge.net/projects/josql/files/josql/, version 2.2

[8]http://sourceforge.net/projects/jedit/files/jedit/, version 5.2.0

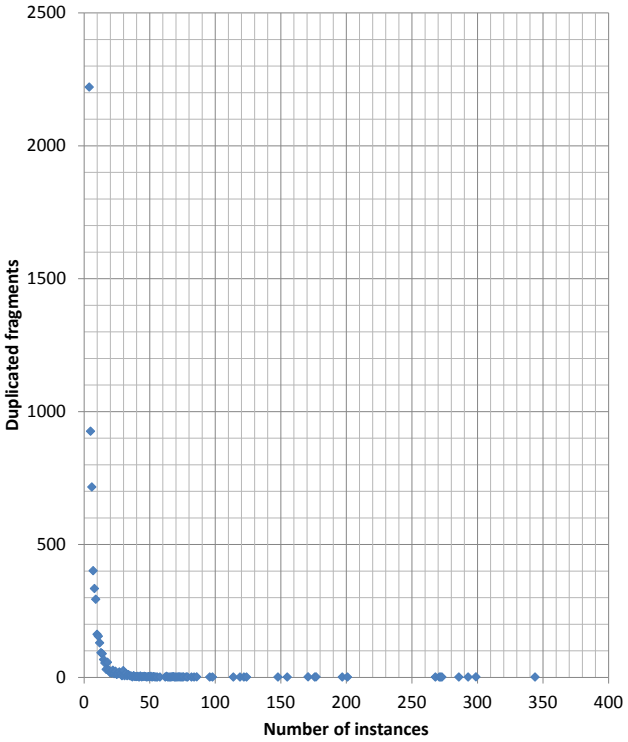


Fig. 2. Duplicated fragments detected by PMD CPD in standard Java

numbers of instances were just lines consisting of asterisks (*) probably used as a visual separator in the documentation.

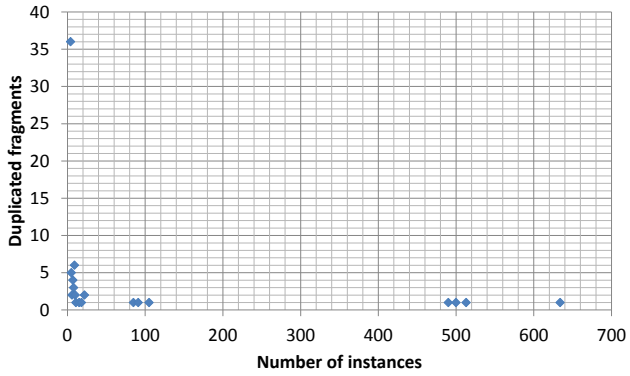


Fig. 3. Duplicated fragments detected by PMD CPD in PicoContainer sources

JasperReports contained 131 duplicated fragments ranging from 44 duplications with 4 instances to a single duplicated fragment with 106 instances. Figure 4 presents results for JasperReports in a simple chart. In this case the duplicated fragment with 106 instances was a documentation phrase reporting that the documented program elements were deprecated and to be removed: '@deprecated To be removed.'. Deprecation naturally expresses design concern – the given program element became obsolete and should not be used anymore.

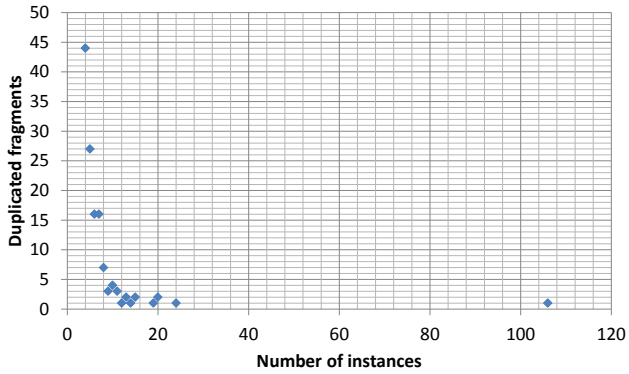


Fig. 4. Duplicated fragments detected by PMD CPD in JasperReports sources

In JoSQL the PMD CPD discovered 31 duplicated fragments. The most 'potent' duplicated fragment with 54 instances was a paramater description: '@param q The Query object.'. Parameter descriptions, especially this simple, could hardly be considered reasonable documentation phrases. The rest of results can be seen in Figure 5.

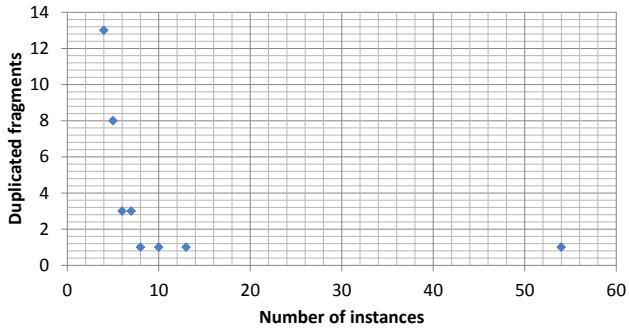


Fig. 5. Duplicated fragments detected by PMD CPD in JoSQL sources

jEdit project sources manifested 76 duplicated fragments with 4 or more instances. Again, the most 'potent' duplicated fragment (with 78 instances) was a line of asterisks. However, the second most 'potent' duplicated fragment was a sentence reporting that the documented method is thread-safe: 'This method is thread-safe.', thus showing that the thread-safe documentation phrase would be useful even beyond the scope of standard Java sources. The results overview can be seen in Figure 6.

We can conclude that duplicated fragments (documentation phrases) are common in practice.

V. Threats to Validity

We should mention several threats to validity that should to be considered for this study. First, the copy/paste detection found all the duplicated sentences in the documentation, even those that could hardly be assigned a concern. In those

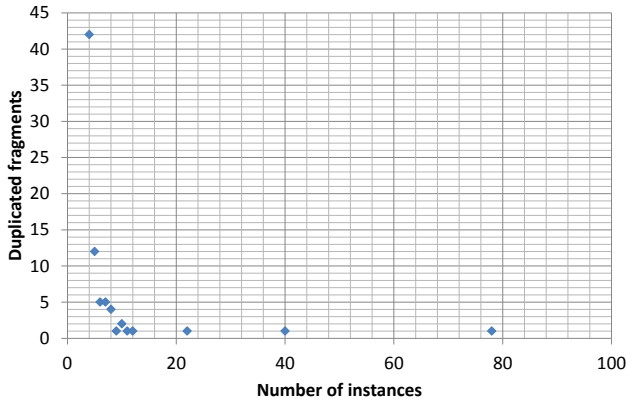


Fig. 6. Duplicated fragments detected by PMD CPD in jEdit sources

cases using documentation weaving [7] would be impractical. Further examination would be useful.

Second, the modified PMD CPD detected solely static phrases – fragments of the documentation that were copy/-pasted in documentation. Inclusion of parametrized documentation phrases [6] would be welcome.

VI. Related Work

Maalej et al. in [10] present a study of knowledge patterns in API reference documentation. They define patterns as knowledge types that categorize types of information expressed by a particular documentation unit (a fragment of API documentation documenting one API element). As example we can mention types like the *Functionality and Behavior* knowledge type that describes what the API does or the *Code Examples* that provides a code sample showing how to use the API.

The work of Horie et al. [7] discusses documentation phrases from the aspect-oriented viewpoint. They view documentation phrases as cross-cutting concerns. Their tool CommentWeaver is able to weave documentation phrases the same way as advices are woven in aspect-oriented programming. Our work presented in [6] continues in their work. There we propose using source code annotations to indicate program elements that should be documented by a given documentation phrase.

Shi et al. in [11] present an empirical quantitative study of API documentation evolution. They analyze the documentation to detect which parts of documentation are frequently revised, how often these revisions indicate behavioral changes in API and how often do these revisions occur. The contribution of their work is in emphasizing the importance of API documentation evolution in order to prevent defects in software using the given API.

VII. Conclusion

In conclusion, the presented results underline the significance of approaches like the one we presented in [6], or the one by Horie et al. [7], which centralize the management of such a documentation phrase into one place and thus ease their maintenance and evolution.

In the future work we need to further examine the results and confirm the significance of fragments that can be considered a concern (intent). An interesting modification of the experiment would be inclusion of the parametrized fragments as well.

Acknowledgment

This work was supported by project KEGA No. 019TUKE-4/2014 Integration of the Basic Theories of Software Engineering into Courses for Informatics Master Study Programmes at Technical Universities – Proposal and Implementation.

References

[1] V. Vranić, J. Porubän, M. Bystrický, T. Frťala, I. Polášek, M. Nosál', and J. Lang, "Challenges in preserving intent comprehensibility in software," *Acta Polytechnica Hungarica*, vol. 12, no. 7, pp. 57–75, 2015. doi: 10.12700/aph.12.7.2015.7.4. [Online]. Available: http://dx.doi.org/10.12700/aph.12.7.2015.7.4

[2] J. Kollár, M. Sičák, and M. Spišiak, "Towards Machine Mind Evolution," in *2015 Federated Conference on Computer Science and Information Systems*, ser. FedCSIS 2015, Sept 2015. doi: 10.15439/2015F210 pp. 985–990. [Online]. Available: http://dx.doi.org/10.15439/2015F210

[3] J. Juhár and L. Vokorokos, "A review of source code projections in integrated development environments," in *2015 Federated Conference on Computer Science and Information Systems*, ser. FedCSIS 2015, Sept 2015. doi: 10.15439/2015F289 pp. 923–927. [Online]. Available: http://dx.doi.org/10.15439/2015F289

[4] E. Pietriková and S. Chodarev, "Profile-driven source code exploration," in *2015 Federated Conference on Computer Science and Information Systems*, ser. FedCSIS 2015, Sept 2015. doi: 10.15439/2015F238 pp. 929–934. [Online]. Available: http://dx.doi.org/10.15439/2015F238

[5] R. Táborský and V. Vranić, "Feature Model Driven Generation of Software Artifacts," in *2015 Federated Conference on Computer Science and Information Systems*, ser. FedCSIS 2015, Sept 2015. doi: 10.15439/2015F364 pp. 1007–1018. [Online]. Available: http://dx.doi.org/10.15439/2015F364

[6] M. Nosáľ and J. Porubän, "Reusable software documentation with phrase annotations," *Central European Journal of Computer Science*, vol. 4, no. 4, pp. 242–258, 2014. doi: 10.2478/s13537-014-0208-3. [Online]. Available: http://dx.doi.org/10.2478/s13537-014-0208-3

[7] M. Horie and S. Chiba, "Tool Support for Crosscutting Concerns of API Documentation," in *Proceedings of the 9th International Conference on Aspect-Oriented Software Development*, ser. AOSD '10. New York, NY, USA: ACM, 2010. doi: 10.1145/1739230.1739242. ISBN 978-1-60558-958-9 pp. 97–108. [Online]. Available: http://dx.doi.org/10.1145/1739230.1739242

[8] V. Vranić and B. Kuliha, "Realizing changes by aspects at the design level," in *Proceedings of the 2015 IEEE 19th International Conference on Intelligent Engineering Systems*, ser. INES 2015, Sept 2015. doi: 10.1109/INES.2015.7329736 pp. 369–374. [Online]. Available: http://dx.doi.org/10.1109/INES.2015.7329736

[9] J. Genči, *About One Way to Discover Formative Assessment Cheating*. Cham: Springer International Publishing, 2015, pp. 83–90. ISBN 978-3-319-06764-3. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-06764-3_11

[10] W. Maalej and M. P. Robillard, "Patterns of Knowledge in API Reference Documentation," *IEEE Transactions on Software Engineering*, vol. 39, no. 9, pp. 1264–1282, Sept 2013. doi: 10.1109/TSE.2013.12. [Online]. Available: http://dx.doi.org/10.1109/TSE.2013.12

[11] L. Shi, H. Zhong, T. Xie, and M. Li, "An Empirical Study on Evolution of API Documentation," in *Proceedings of the 14th International Conference on Fundamental Approaches to Software Engineering: Part of the Joint European Conferences on Theory and Practice of Software*, ser. FASE'11/ETAPS'11. Berlin, Heidelberg: Springer-Verlag, 2011. doi: 10.1007/978-3-642-19811-3_29. ISBN 978-3-642-19810-6 pp. 416–431. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-19811-3_29

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1577–1580
DOI: 10.15439/2016F408
ACSIS, Vol. 8. ISSN 2300-5963

A Model-to-Model Transformation of a Generic Relational Database Schema into a Form Type Data Model

Sonja Ristić, Slavica Kordić, Milan Čeliković, Vladimir Dimitrieski, Ivan Luković
University of Novi Sad,
Faculty of Technical Sciences,
Trg D. Obradovića 6, 21000 Novi Sad, Serbia
Email: {sdristic, slavica, milancel, dimitrieski, ivan}@uns.ac.rs

Abstract—**An important phase of a data-oriented software system reengineering is a database reengineering process and, in particular, its subprocess – a database reverse engineering process. In this paper we present one of the model-to-model transformations from a chain of transformations aimed at transformation of a generic relational database schema into a form type data model. The transformation is a step of the data structure conceptualization phase of a model-driven database reverse engineering process that is implemented in IIS*Studio development environment.**

I. Introduction

A MODEL-DRIVEN (MD) approach to information system (IS) and software (re)engineering addresses complexity through abstraction. A complex system consists of several interrelated models organized through different levels of abstraction and platform specificity. Through a forward engineering process models need to be refined and integrated and used to produce code and therefore they would undergo a series of transformations. Each transformation adds levels of specificity and detail. A chain of model-to-model (M2M) transformations is completed starting from an initial model at the highest level of abstraction (Platform Independent Model, PIM), through the less abstract models, with different levels of platform specificity (Platform Specific Models, PSMs), and resulting in an executable program code that represents a model at the lowest level of abstraction (fully PSM). Conversely, in a reverse engineering process, the abstraction level of models and degree of platform independency are increasing throughout the chain of transformations.

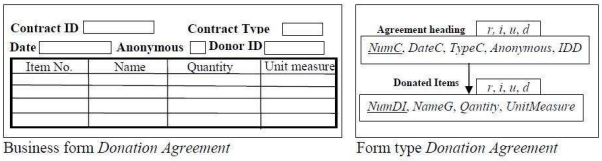
Through a number of research projects on MD intelligent systems for IS development, maintenance and evolution, we have developed the IIS*Studio development environment. It is aimed to provide the IS design, generating executable application prototypes and IS reverse engineering. Our approach is mainly based on the MD information system and software engineering [1] and domain specific language (DSL) paradigms ([2], [3]). In [4] we discuss the importance of meta-modeling in the context of database reverse engineering and review different database meta-models (MM) that are used in the database reengineering process applied in IIS*Studio. In [5] we propose an MD approach to data structure conceptualization phase of database reverse engineering process that is conducted through a chain of M2M transformations. In this paper we present the final step of the conceptualization phase—the M2M transformation of a generic relational database schema into a form type model.

The form type concept and the IIS*Studio architecture are given in Section 2. Classifications of form types and relation schemes are described in Section 3. The transformation of a generic relational database schema into a form type data model is presented in Section 4 and related work is discussed in Section 5.

II. Form Type Concept

A form type is central IIS*Studio PIM concept, used to model the structure and constraints of various business forms that are broadly used in organizations to conduct daily operations and to communicate with their affiliated entities. They are a source for eliciting user information requirements and for designing and developing user-oriented information systems. Initially, each form type (FT) is an abstraction of a business form. However, it may be enriched by additional specifications like specifications of: key and unique constraints; check constraints; allowed database CRUD (Create, Retrieve, Update and Delete) operations applied by means of screen computerized forms to manipulate data of an IS; functionalities concerning relationships between generated screen forms, i.e. transaction programs. The business form *Donation Agreement* (*DA-bf*) is presented on the left-hand side of Fig. 1. It is business form used in the Safe House Center (SHC) that provides support for those children impacted by domestic violence. The SHC is in a great extent based on donations and SHC IS has to support donation process. One of the activities is keeping track about the donation agreements. The business form *Donation Agreement* may be modeled by the form type *Donation Agreement* (*DA-ft*). The simplified representation of the structure of the *DA-ft*, which generalizes the *DA-bf*, is presented on the right-hand side of Fig. 1. A form type is a hierarchical structure of form type components. The form type *Donation Agreement* (Fig. 1) has two component types: *Agreement Heading* and *Donated Items*.

Research presented in this paper was supported by Ministry of Education, Science and Technological Development of Republic of Serbia, Grant III-44010, Title: Intelligent Systems for Software Product Development and Business Support based on Models.



Business form *Donation Agreement* Form type *Donation Agreement*

Fig. 1 The business form *Donation Agreement* and its form type

A form type in IS design by means of IIS*Studio has a dual role. On the one hand it provides an important input data for database design, and on the other hand it is a source for the generation of a sole transaction program and its screen or report form. IIS*Studio introduces *FT data model* based on FT concept [6] aimed at conceptual database design.

IIS*Studio comprises: IIS*Case, IIS*UIModeler, and IIS*Ree tools that communicate by means of shared repository aimed at storing project specifications. The IIS*Case tool supports IS forward engineering process. The IIS*UIModeler is aimed at modeling of graphic user interface (GUI) static aspects via UI templates. The IIS*Ree tool enable reverse engineering (RE) of relational databases to conceptual data models. The RE process is implemented by means of a series of M2M transformations between database models (*database model transformations*) based on meta-models that are conformed by the source and target database models. A blueprint of IIS*Studio database RE process is presented in [5]. Here we present one step of that process aimed at transformation of a generic relational database schema into a form type data model.

III. Classifications of Form Types and Relation Schemes

A *form type* $\mathcal{F}$ is a named tree structure, whose nodes are called *component types* (CTs). Let $C(\mathcal{F})$ denotes a set of CTs making up the form type $\mathcal{F}$. Each CT is identified by its name within the scope of a FT, and has nonempty sets of attributes and keys, and a possibly empty set of unique constraints. Formally, a CT is a named pair $N(Q, O)$, where N denotes name of the CT, Q is the set of CT attributes $Q=\{A_1, .., A_n\}$ and O is a set of CT constraints. O is a union of: a set of key constraints, a set of unique constraints and a singleton containing a tuple constraint. The tuple constraint of a CT refers to a set of attribute-based constraints (attribute data type specification and not-null constraint) paired with a tuple-based constraint (constraint on tuple value).

Let $C(\mathcal{F}) = \{N_i(Q_i, O_i) \mid i = 1, ... m\}$. $W(\mathcal{F})$ denotes a set of the form type attributes that satisfies (1) and (2).

$$\bigcup_{i=1}^{m} Q_i = W(\mathcal{F}) \text{ and} \quad (1)$$

$$(\forall N_i, N_j \in C(\mathcal{F}))(i \neq j \Leftrightarrow Q_i \cap Q_j = \varnothing). \quad (2)$$

Three categories of FTs can be identified: $\mathcal{F}\_Basic$—an elementary form type containing only one root component type (Fig. 2); $\mathcal{F}\_Tree^2$—a form type containing a root component type with only one child component type (Fig. 3); and $\mathcal{F}\_Tree^n$—a form type that apart from a root component type contains an arbitrary number of child component types (Fig. 4).

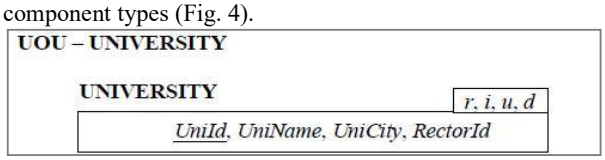


Fig. 2 An example of $\mathcal{F}\_Basic$ form type

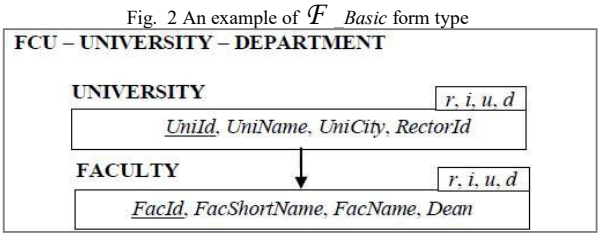


Fig. 3 An example of $\mathcal{F}\_Tree^2$ form type

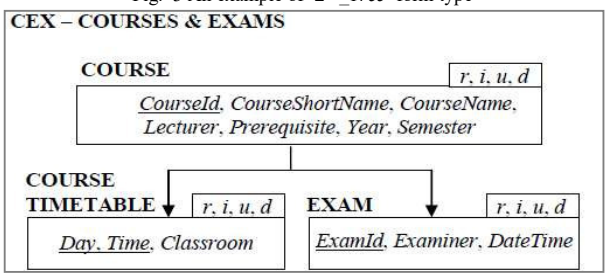


Fig. 4 An example of $\mathcal{F}\_Tree^n$ form type

Hammer et al. [7] have proposed a classification of relation schemes in the context of the transformation of a relational database schema into Entity-Relationship (ER) database schema. Here we present a classification that is adapted according to the target FT data model. There are three kinds of relation schemes: basic (BR); weak (WR); and all keys (AK) relation scheme. A BR relation scheme is a relation scheme whose PK does not properly contain a key attribute of any other relation. A WR relation scheme N satisfies the following three conditions: i) a proper subset of its PK contains key attributes of other basic or weak relation schemas; ii) the remaining attributes of its PK do not contain key attributes of any other relation scheme; and iii) it has an identifying parent relation scheme and properly contains the PK of its parent relation scheme. An AK relation scheme contains only key attributes of other relation schemes, and does not contain any other self-inherent attributes.

A graphic representation of a relational database schema is presented in Fig. 5. Underlined attributes belong to a key of a relation scheme. If a relation scheme has two candidate keys their attributes are underlined with different lines. The relation schemes *University* and *Project* are BR relation schemes. *Faculty*, *Department* (first version), *Employee*, the second version of *Department* relation schema (below the first version) and relation scheme *Lecturer* are WR relation schemes. The relation scheme *WorkOn* is an example of AK relation scheme.

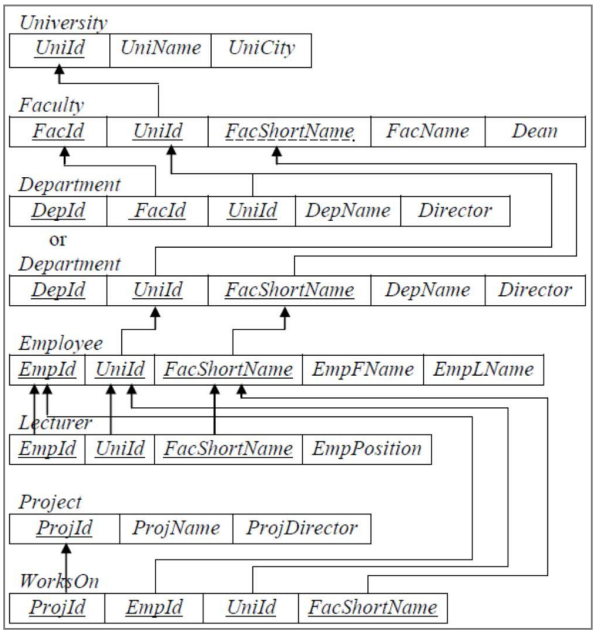


Fig. 5 An example of a relational database schema

IV. Transformation of a Generic Database Schema into a Form Type Data Model

The transformation of a model conformant with generic relational database meta-model into a model conformant with FT meta-model is PSM2PIM transformation. It generates all relevant combinations of form types. It is a user who chooses form types to be introduced in the form type data model. The remaining form types are deleted. The proposed transformation is carried out in three steps. In the first step, a $\mathcal{F}$ _Basic form type is created for each relation scheme from a relational database schema. The input parameters for this transformation are:

$$S = \{N_i(R_i, O^{RS}_i) \mid i = 1,.., n\}, \text{ and} \quad (3)$$

$$O^{RS} = K^{RS} \cup UQ^{RS} \cup CH^{RS}, \quad (4)$$

where S is a set of the relation schemes $N_i(R_i, O^{RS}_i)$. N_i is relation scheme name, R_i is nonempty set of its attributes and O^{RS}_i (4) is a union of three sets containing: key constraints, unique constraints, and tuple constraints.

In Fig. 6 parts of generic relational schema MM and FT meta-model are presented in the first row. In the next raw ATL rules are presented that specify a mapping between the concepts of these MMs alongside with five lazy rules aimed at mapping: optional and mandatory relation scheme attributes to optional and mandatory CT attributes; and relation scheme key, unique and tuple constraints to CT key, unique and tuple constraints, respectively.

In the next step of the transformation a $\mathcal{F}$ _$Tree^2$ form type is generated for each referential integrity constraint, having WR relation scheme on the left-hand side. The set of input parameters in addition to (3) and (4) contains a set of referential integrity constraints of a relational database schema (5):

$$RIC = \{ric_i: N_l[LHS] \subseteq N_r[RHS] \mid i = 1,.., m\}, \quad (5)$$

where N_l and N_r are relation schemes, *LHS* and *RHS* are subsets of attribute sets R_l and R_r of relation schemes N_l and N_r, respectively. For each ric_i from *RIC*, with N_l that is WR relation scheme, a $\mathcal{F}$ _$Tree^2$ FT is created.

In the last step of the transformation an $\mathcal{F}$ _$Tree^n$ form type is created for each relation scheme that is referenced by at least two WR relation schemes. Besides, an $\mathcal{F}$ _$Tree^n$ form type is created for each relation scheme that is referenced by at list one WR relation scheme that is referenced by some WR relation scheme, too.

V. Related Work

Hainaut et al. in [8] describe main steps of database RE. Vendor-specific physical or standard relational meta-model mainly are found on the source side of RE transformations. On the other side, ER [9], object-oriented [9]–[11], standard/vendor-specific relational [12] or object-relational [13] MMs occur on the target side. There are various research works about the use of forms in different contexts: Tsichritzis [14] introduces the concepts of form type, Shu [15] proposed using forms to specify system requirements, in [16] is presented a usage of business forms as input data for the process of database schema design. A form-based approach for reverse engineering of relational databases is proposed in [17].

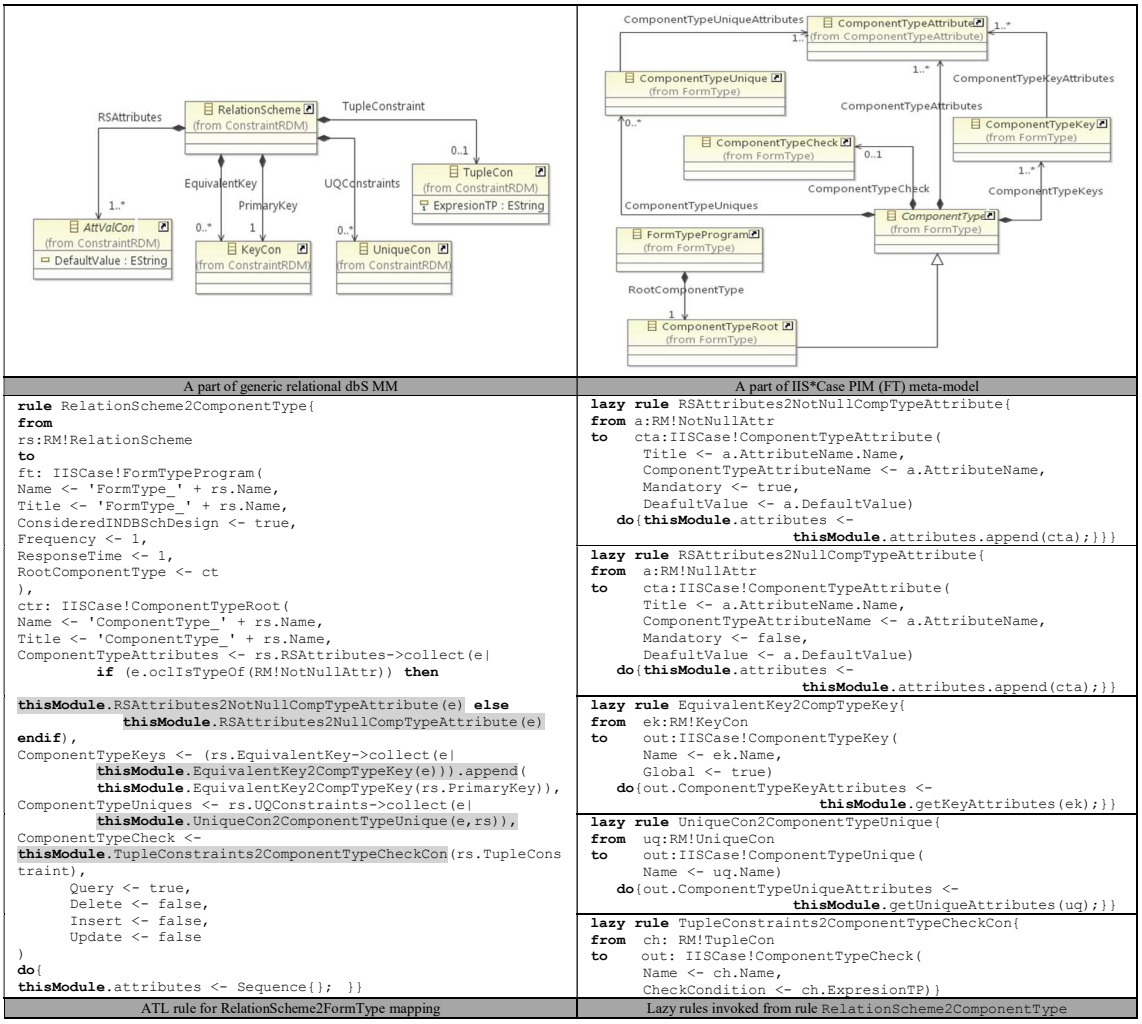
VI. Conclusion

The main reason to develop our IIS*Ree reverse engineering tool was to take advantage of our approach to database schema generation during: the integration of independently designed ISs, legacy database schema restructuring and improvement of empirically designed database schemas. FT specification models system as-is in a platform independent way. At the same time, the specification is platform independent prescription model of future screen and report forms and input for series of M2M transformations that ends up with M2T transformation generating application prototype. The MMs and models that we use in our approach are intensional data models. System evolution can be supported by automatic MD data migration and extensional database MMs could play important role in its implementation. Our future research has to consider extensional database MMs and possible usage of category theory [18] for PIM specification of model transformations in order to automate the process of database model transformations generation.

References

[1] J.M Favre, " Foundations of Model (Driven) (Reverse) Engineering: Models.", *Dagstahl Seminar Proceedings*, 2005.

[2] T. Kosar, N. Oliveira, M. Mernik, V. J. M. Pereira, M. Črepinšek, C. D. Da, and R. P. Henriques, "Comparing general-purpose and domain-specific languages: An empirical study," *Computer Science and Information Systems,* vol. 7 (2), pp. 247–264, 2010. DOI: 10.2298/CSIS1002247K



A part of generic relational dbS MM

A part of IIS*Case PIM (FT) meta-model

```
rule RelationScheme2ComponentType{
from
rs:RM!RelationScheme
to
ft: IISCase!FormTypeProgram(
Name <- 'FormType_' + rs.Name,
Title <- 'FormType_' + rs.Name,
ConsideredINDBSchDesign <- true,
Frequency <- 1,
ResponseTime <- 1,
RootComponentType <- ct
),
ctr: IISCase!ComponentTypeRoot(
Name <- 'ComponentType_' + rs.Name,
Title <- 'ComponentType_' + rs.Name,
ComponentTypeAttributes <- rs.RSAttributes->collect(e|
        if (e.oclIsTypeOf(RM!NotNullAttr)) then

thisModule.RSAttributes2NotNullCompTypeAttribute(e) else
          thisModule.RSAttributes2NullCompTypeAttribute(e)
endif),
ComponentTypeKeys <- (rs.EquivalentKey->collect(e|
        thisModule.EquivalentKey2CompTypeKey(e))).append(
        thisModule.EquivalentKey2CompTypeKey(rs.PrimaryKey)),
ComponentTypeUniques <- rs.UQConstraints->collect(e|
        thisModule.UniqueCon2ComponentTypeUnique(e,rs)),
ComponentTypeCheck <-
thisModule.TupleConstraints2ComponentTypeCheckCon(rs.TupleCons
traint),
      Query <- true,
      Delete <- false,
      Insert <- false,
      Update <- false
)
do{
thisModule.attributes <- Sequence{}; }}
```

ATL rule for RelationScheme2FormType mapping

```
lazy rule RSAttributes2NotNullCompTypeAttribute{
from a:RM!NotNullAttr
to   cta:IISCase!ComponentTypeAttribute(
       Title <- a.AttributeName.Name,
       ComponentTypeAttributeName <- a.AttributeName,
       Mandatory <- true,
       DeafultValue <- a.DefaultValue)
   do{thisModule.attributes <-
                    thisModule.attributes.append(cta);}}
lazy rule RSAttributes2NullCompTypeAttribute{
from  a:RM!NullAttr
to    cta:IISCase!ComponentTypeAttribute(
       Title <- a.AttributeName.Name,
       ComponentTypeAttributeName <- a.AttributeName,
       Mandatory <- false,
       DeafultValue <- a.DefaultValue)
   do{thisModule.attributes <-
                    thisModule.attributes.append(cta);}}
lazy rule EquivalentKey2CompTypeKey{
from  ek:RM!KeyCon
to    out:IISCase!ComponentTypeKey(
       Name <- ek.Name,
       Global <- true)
   do{out.ComponentTypeKeyAttributes <-
                    thisModule.getKeyAttributes(ek);}}
lazy rule UniqueCon2ComponentTypeUnique{
from  uq:RM!UniqueCon
to    out:IISCase!ComponentTypeUnique(
       Name <- uq.Name)
   do{out.ComponentTypeUniqueAttributes <-
                    thisModule.getUniqueAttributes(uq);}}
lazy rule TupleConstraints2ComponentTypeCheckCon{
from  ch: RM!TupleCon
to    out: IISCase!ComponentTypeCheck(
       Name <- ch.Name,
       CheckCondition <- ch.ExpresionTP)}
```

Lazy rules invoked from rule `RelationScheme2ComponentType`

Fig. 6 Relation scheme – to – $\mathcal{F}$ _Basic Form Type transformation

[3] I. Dejanović, G. Milosavljević, B. Perišić, M. Tumbas, "A Domain-Specific Language for Defining Static Structure of Database Applications", *Computer Science and Information Systems*, (ComSIS), ISSN:1820-0214, Vol. 7, No. 3, pp 409-440. 2010.

[4] S. Ristić, S. Aleksić, M. Čeliković, V. Dimitrieski, and I. Luković, "Database reverse engineering based on meta-models," *Central European Journal on Computer Science*, vol. 4(3), pp: 150–159, 2014. DOI: 10.2478/s13537-014-0218-1

[5] S. Ristić, S. Kordić, M. Čeliković, V. Dimitrieski, I. Luković, "A Model-driven Approach to Data Structure Conceptualization" in *Proceedings of the 2015 FEDCSIS„* Vol. 5, DOI: http://dx.doi.org/10.15439/ 978-83-60810-66-8, pp. 977–984. 2015.

[6] I. Luković, P. Mogin, J. Pavićević, and S. Ristić, "An approach to developing complex database schemas using form types," *Software: Practice and Experience*, vol. 37 (15), pp. 1621–1656, 2007. doi: 10.1002/spe.820

[7] M. Hammer, M. Schmalz, W. O'Brien, S. Shekar, N. Haldevnekar, *Knowledge Extraction in the SEEK Project Part I*, Tecnical Report TR-02-008, July 2002.

[8] J-L. Hainaut, J. Henrard, V. Englebert, D. Roland, J-M. Hick J-M, "Database Reverse Engineering", In: *Encyclopedia of Database Systems*, L. Liu and Özsu, T. (ed), Springer-Verlag, 2009.

[9] M. Gogolla, A. Lindow, M. Richters, and P. Ziemann, "Meta-model transformation of data models", WISME at the UML, 2002.

[10] J. Perez, I. Ramos, and V. Anaya, "Data reverse engineering of legacy databases to object oriented conceptual schemas," *Electronic Notes in Theoretical Computer Science*, vol. 74(4), pp. 1–13, 2002.

[11] A. Boronat, J. Perez, J. A. Cars, and I. Ramos., "Two Experiences in Software Dynamics," *Journal of Universal Computer Science*, vol. 10(4), pp. 428–453, 2004.

[12] Beggar O. E., Bousetta B., Gadi T., Getting Relational Database from Legacy Data-MDRE Approach, Computer Engineering and Intelligent Systems ISSN 2222-1719 Vol 4, No.4, 2013.

[13] J. Vara, B. Vela, V.A. Bollati E. Marcos, "Supporting model-driven development of object-relational database schemas: a case study", in: *Theory and Practice of Model Transformations*, R. Paige (Ed.), Heidelberg, Springer Berlin, 2009. pp. 181–196.

[14] D. Tsichritzis, "Form management",. *Communications of the ACM* 25 (5), pp. 453–478. 1982.

[15] N.C. Shu "FORMAL: a form-oriented, visual-directed application development system", *Computer*, pp 38– 49. 1985.

[16] J. Choobineh, S.S. Venkatraman, "A methodology and tools for derivation of functional dependencies from business form", *Information Systems* 17 (3), pp 269–282. 1992.

[17] M. Malki, A. Flory, and M. K. Rahmouni, "Extraction of Object-oriented Schemas from Existing Relational Databases: a Form-driven Approach," *INFORMATICA*, vol. 13(1), pp. 47–72, 2002.

[18] W. Steingartner, and D. Radaković, "Categorical structures as expressing tool for differential calculus", In: Proceedings of: The 12th Conference Informatics'2013 Technical University of Kosice, Slovakia, pp. 77–82, 2013.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1581–1591
DOI: 10.15439/2016F250
ACSIS, Vol. 8. ISSN 2300-5963

Towards OntoUML for Software Engineering: Transformation of Rigid Sortal Types into Relational Databases

Zdeněk Rybola, Robert Pergl
Faculty of Information Technology
Czech Technical University in Prague
Thákurova 9, 16000 Praha 6
Email: {zdenek.rybola, robert.pergl}@fit.cvut.cz

***Abstract*—OntoUML is an ontologically well-founded conceptual modelling language that distinguishes various types of classifiers and relations providing precise meaning to the modelled entities. Efforts arise to incorporate OntoUML into the Model-Driven Development approach as a conceptual modelling language for the PIM of application data. In a prequel paper, we have introduced and outlined our approach for a transformation of OntoUML PIM into a PSM of a relational database. In this paper, we discuss the details of various variants of the transformation of Rigid Sortal types of OntoUML.**

I. Introduction

SOFTWARE engineering is a demanding discipline that deals with complex systems [1]. The goal of software engineering is to ensure high quality software implementation of these complex systems. To achieve this, various software development approaches have been developed.

Model-Driven Development (MDD) is a very popular approach in the recent years. It is a software development approach based on elaborating models and performing their transformations [2]. The product to be developed is described using various types of models specifying the requirements, functions, structure and deployment of the product. These models are used to construct the product using transformations between models and code.

The most usual part of the MDD approach used in the practice is the process of *forward engineering*: transformations of more abstract models into more specific ones. The most common use-case of such process is the development of conceptual data models and their transformation into source codes or database scripts.

To achieve a high-quality software system, high-quality expressive models are necessary to define the requirements for the system [1]. To use such models in the Model-Driven Development approach, the model should define all requirements and all constraints of the system. Moreover, it should hold that more specific models persist the constraints defined in the more abstract models [3].

This research was partially supported by grant by Student Grant Competition No. SGS16/120/OHK3/1T/18.

OntoUML was formulated in 2005 as a graphical modelling language for developing ontologically well-founded conceptual models [3]. As it is based on cognitive science and modal logic, it helps to create expressive models that are able to describe the domain very precisely. As OntoUML is domain-agnostic, it may be used for any domain. In our research, we focus on the domain of software application data and therefore we use OntoUML to create the PIM of the system. Such model can be then transformed into a PSM of the data persistence. However, as OntoUML uses various types of entities and relations to provide additional ontological meaning to the model elements, the transformation needs to deal with these aspects.

As relational databases represent a very common type of data storage, we focus on the transformation of an OntoUML PIM of application data into an ISM of a relational database. To achieve that, we divide the transformation into the following steps:

1) Transform an OntoUML PIM into a UML PIM including all the aspects defined by the OntoUML constructs.
2) Transform the UML PIM with the additional constraints into a PSM of a relational database including the required additional constraints.
3) Transform the PSM with the additional constraints into the ISM to define the constructs in the database to hold the data and maintain the constraints.

In the prequel paper [4], we outlined the various possibilities of the transformation of Sortal universal types used in OntoUML. In this paper, we discuss the details of the transformation of Rigid Sortal types (Kinds and Subkinds) and illustrate various possibilities on examples. The parallel research focused on the transformation of OntoUML Anti-rigid universal types is discussed in the parallel paper [5].

The structure of the paper is as follows: in section II, the work related to our approach including the OntoUML notation is discussed; in section III, the running example of the OntoUML PIM is explained; in section IV, our approach is discussed and illustrated on the running example; in section V, discussion to our approach is provided; finally, in section VI, the conclusion of the paper results is provided.

978-83-60810-90-3/$25.00

II. Background and Related work

A. Model-Driven Development

Model-Driven Development (MDD) is a very popular approach in the recent years. It is a software development approach based on elaborating models and performing their transformations [2]. The product to be developed is described using various types of models specifying the requirements, functions, structure and deployment of the product. These models are used to construct the product using transformations between models and code generation.

MDD was originally based on Model-Driven Architecture (MDA) [6] designed by OMG in 2001. MDA defines these types of models:

- Computation Independent Model (CIM),
- Platform Independent Model (PIM),
- Platform Specific Model (PSM),
- Implementation Specific Model (ISM) [7].

Although established already in 2001, there is still deep interest in this approach, as can be seen in recent publications. The book *Model-Driven Software Development: Technology, Engineering, Management* by Stahl et al. [8] provides a great overview of the MDD approach including the terminology, specifications, transformations and case studies. Another book *Model-Driven Software Engineering in Practice* by Brambilla et al. [9] presents the foundations of MDSE approach and also deals with the technical aspects of MDSE including the basics of domain-specific languages, transformations and tools. Also, the survey by da Silva [10] provides a good overview of the MDD approach and terminology related to MDE, MDD and MDA. Another survey was published by Whittle et al. [11] that focused on the support of the MDE approach in tools and provides a taxonomy of tool-related considerations.

The most usual part of the MDD approach used in the practice seems to be the process of *forward engineering*: transformations of more abstract models into more specific ones. The most common use-case of such process is the development of conceptual data models and their transformation into source codes or database scripts. In our research, we focus on the modelling of application data creating a PIM in OntoUML and performing transformations to generate creation scripts of a relational database schema.

B. UML

Unified Modeling Language (UML) [12], [13] is a popular modelling language for creating and maintaining variety of models using diagrams and additional components [10]. UML defines a set of building blocks – various types of elements (e.g. classes, use cases, components, etc.), relations (e.g. association, generalization, dependency, etc.) and diagrams (e.g. class diagram, use case diagram, sequence diagram, etc.). It defines also the syntax and semantics of models and a general architecture of the model [7]. In context of the data modelling, UML Class Diagram is the notation mostly used to define conceptual models of application data. Also, to describe the structure of a relational database schema, UML Data Model profile as an extension to the UML Class Diagrams may be used [14].

The main elements of a UML Class Diagram are classes, which serve to classify various types of objects in the domain of interest and specify their features and behaviour [13]. Between the classes, associations and generalization/specialization relations can be defined. The associations are used to define the fact that various instances of one class can be related to some instances – according to association multiplicities – of the other class.

Generalization is a taxonomic relationship between a more general class – *superclass* – and a more specific class – *subclass* [13]. It is used in situations when there are multiple special cases of the more general class with additional features and/or specialised meaning. In such a situation, the subclasses inherit all features of their superclass and add their own features, so their instances have all the features of both the superclass and the subclass[1]. As UML is designed following the object-oriented programming approach, an object can be an instance of only one class [7]. As UML is based on object-oriented paradigm, an object is either an instance of the superclass or an instance of the subclass, inheriting the features from the superclass but not being its direct instance.

The subclasses of the same superclass may form a *generalization set* to define a partition of subclasses with common meaning [13]. For each generalization set, two meta-properties should be set to restrict the relation of an instance to the individual subclasses: *isCovering* – expressing whether each instance of the superclass must be also an instance of some subclass in the generalization set – and *isDisjoint* – expressing whether an object can be an instance of multiple subclasses in the set at the same time. The default setting of these properties differ in the various versions of UML: UML 2.4.1 [12] and older define the `{incomplete, disjoint}` as default, while UML 2.5 [13] defines the `{incomplete, overlapping}` as default. As each object is an instance of exactly one class in the most current programming languages, the concept of generalization sets can be used only in conceptual models and it must be transformed before its realization.

C. OCL

Object Constraint Language (OCL) [15] is a specification language that is part of the UML standard. It can be used for the following purposes:

- to access model elements and their values,
- to define constraints and restrictions for model elements and their values,
- and to define query operations [7].

Several types of OCL constructs may be used to define the constraints for the model elements. *Invariants* are defined in context of certain class of the attached UML model and they are used to define constraints which must be satisfied

[1] In fact, the features of the superclass may be overriden by features of the subclass, but this situation is not considered here.

by all contextual instances at any moment. *Preconditions* and *postconditions* are defined in context of certain method of a class in the attached UML model. Preconditions define the constraints that must be satisfied before executing the method (e.g. the values of the method parameters), while postconditions define the constraints which must be satisfied after the method execution (e.g. the value of the result).

In [16], the authors define basic syntax and semantics of OCL constructs and introduce several tools that support modelling and evaluation of OCL constraints. In [17], the authors define a technique for transformations of OCL constructs into other equivalent forms to support their definition, validation and transformation.

In our approach, we use OCL invariants to define the constraints on the UML PIMs and PSMs derived from the semantics of OntoUML universal types that cannot be expressed directly in the diagram.

D. OntoUML

OntoUML is a conceptual modelling language focused on building ontologically well-founded models. It was formulated in Guizzardi's PhD Thesis [3] as a light-weight extension of UML based on UML profiles.

The language is based on *Unified Foundational Ontology* (UFO), which is based on the cognitive science and modal logic and related mathematical foundations such as sets and relations. Thanks to this fact, it provides expressive and precise constructs for modellers to capture the domain of interest. Unlike other extensions of UML, OntoUML does not build on the UML's ontologically vague "class" notion, but builds on the notion of *universals* and *individuals*. It uses the basic notation of UML Class Diagram like classes, associations and generalization/specialization together with stereotypes and meta-attributes to define the nature of individual elements more specifically. On the other hand, it omits a set of other problematic concepts (for instance aggregation and composition) and replaces them with its own ontologically correct concepts.

UFO and OntoUML address many problems in conceptual modelling, such as part-whole relations [18] or roles and the counting problem [19]. The language has been successfully applied in different domains such as interoperability for medical protocols in electrophysiology [20] and the evaluation of an ITU-T standard for transport networks [21].

However, being domain-agnostic, we believe that it may be suitable even for conceptual modelling of application data in the context of MDD. Using OntoUML, we can create very precise and expressive models of application data. These models can be later transformed into relational database schema containing various domain-specific constraints to maintain consistency according to the OntoUML model.

The following description of the OntoUML and UFO aspects is based on [3].

1) Universals and individuals: UFO distinguishes two types of things. *Universals* are general classifiers of various objects and they are represented as classes in OntoUML (e.g. `Person`). There are various types of universals according to their properties and constraints as discussed later. *Individuals*, on the other hand, are the individual objects instantiating the universals (e.g. `Mark`, `Dan`, `Kate`).

The fact that an individual is an instance of a universal means that – in the given context – we perceive the object *to be* the Universal (e.g. `Mark` is a `Person`). Important feature of UFO is the fact that an individual may instantiate multiple universals at the same time but all the universals must have a common ancestor providing the identity principle (e.g. `Mark` is a `Person` and he is a `Student` as well).

2) Identity principle: Identity principle is a key feature of UFO, which enables individuals to be distinguished from each other. Various universals define different identity principles and thus different ways how to distinguish their individuals (e.g. a `Person` is something else than a `University`); different individuals of the same universal have different identities (e.g. `Mark` is not `Kate` even when both are `Persons`).

Each individual always needs to have a single specific identity, otherwise there is a clash of identities (e.g. `Mark` is a `Person` and therefore it can never be confused with another concept such as a `University`). The identity of an individual is determined at the time the individual comes to existence and it is immutable – it can never be changed (e.g. `Mark` will always be `Mark` and he will always be a `Person`).

The types of universals that provide the identity principle for their instances are called *Sortal universals* (e.g. `Person`, `Student`). The types of universals not providing the identity principle are called *Non-Sortal universals* (e.g. a `Customer` may be a `Person` or a `Company`). In this paper, we discuss only the transformations of the Sortal types of universals, as they form the basis of models.

3) Rigidity: UFO and OntoUML are built on the notion of worlds coming from Modal Logic – various configurations of the individuals in various circumstances and contexts of time and space. *Rigidity* is, then, the meta-property of universals that defines the fact if the extension of the universal (i.e. the set of all instances of the universal) is world invariant [22]. UFO distinguishes rigid, anti-rigid and semi-rigid universals:

- `Rigid universals` are such types of universals whose extension is rigid – instances of the Rigid universals cannot cease to be their instances without ceasing to exist (e.g. `Mark` will always be a `Person`). Certain types of both Sortal and Non-Sortal universals are rigid.
- `Anti-rigid universals` are such types of universals which, in one world, contain an instance in their extension, which is not included in the extension in another world. It means that an individual that is an instance of the Anti-rigid universal in one world may not be an instance of that universal in another world without ceasing to exist (e.g. `Mark` is a `Student` now, but he will not be a `Student` 50 years later). Certain types of both Sortal and Non-Sortal universals are anti-rigid.
- `Semi-rigid universals` are such types of universals that can include both rigid and anti-rigid instances in

their extension. Only Non-Sortal types of universals are semi-rigid.

In this paper, we focus only on the Rigid Sortal types of universals and we discuss the details of the transformation of such universals into the relational databases.

4) Generalization and Specialization: In contrast to UML, in UFO and OntoUML, the generalization relation defines the inheritance of the *identity principle*. According to that, an individual which is an instance of the subclass is also an instance of the superclass automatically through inheriting the identity principle from the superclass. Also, the relation is rigid in UML – when an instance of the superclass is also an instance of the subclass, it cannot cease to be so without losing its identity – while in OntoUML, the relation may be non-rigid: a single individual may be an instance of both the superclass and subclass in one world and it may be an instance of only the superclass in another world.

The generalization sets in OntoUML are much more common as they define the required identity for various universal types. Unless altered, *{incomplete, non-disjoint}* is considered the default value of the meta-properties.

5) Kinds and Subkinds.: The backbone of an OntoUML model is created by Kinds. *Kind* is a Rigid Sortal type of universals that defines the identity principle for its instances, thus defining the way how we are able to distinguish individual instances of that universal. In OntoUML, the Kind universals are depicted as classes with the $\ll Kind \gg$ stereotype. Examples of Kind universals are a `Person` and a `University`.

Subkind is a Rigid Sortal universal type that does not define its own identity principle, but it inherits it from its ancestor and provides it to its instances. Therefore, Subkind universals form generalization sets of other Kind or Subkind universals; they form inheritance hierarchies with the root in a Kind universal. In other words, each instance of a Subkind universal is automatically – through the transitive generalization relation – also an instance of all the ancestral Kind and Subkind universals, receiving the identity principle from the root Kind universal. The inheritance may have any combination of values of the *isDisjoint* and *isCovering* meta-properties. Examples of Subkind universals may be a `Man` and a `Woman` as subkinds of a `Person`.

6) Other universal types.: UFO and OntoUML define several other universal types such as Role, Phase, Relator, Mixin, Quantity et al. However, they are out of scope of this paper.

E. Tools

There are tools supporting certain parts of the transformation process described in section IV. Although none of them supports the full transformation, they can be used for the individual steps or serve as an inspiration for a complex tool to be developed.

Enterprise Architect[2] is a complex CASE tool supporting the whole software development process. Beside the modelling in UML and other notations, it offers transformation between models and source code generation. In context of our work, the transformation of a class model into a database model and the generation of SQL DDL scripts are useful. Beside Enterprise Architect, there are many other tools providing similar functions for UML and relational databases (e.g. Visual Paradigm[3]).

There are also several tools supporting definition of OCL constraints and their evaluation on a given model instance, such as DresdenOCL[4], OCLE[5] or USE[6]. DresdenOCL even provides functions to generate Java source code with AspectJ for the OCL constraints or SQL DDL scripts with views for the OCL constraints.

For OntoUML, there are a few tools available, as well. OntoUML lightweight editor (OLED)[7] is an environment for modelling with OntoUML which also offers functions for model visualisation, validation and transformation into OWL. However, it does not offer transformation into UML nor into relational databases. Menthor Editor[8] is a successor of OLED, providing more convenient environment for modelling and providing transformations of an OntoUML model along with OCL constraints into OWL, RDF and UML. As for other tools, there is an Enterprise Architect plugin[9] and a palette for UMLet editor[10] available for OntoUML modelling.

F. Previous work

In our previous work, we focused on the transformation of special multiplicity values in a UML PIM into PSM for relational databases [23] and the possible realizations of such constraints [24]. The approaches described in these papers may be used for the realization of the constraints derived from the OntoUML constructs used in the PIM as discussed in this paper.

In [25] we focused on the transformation of an ontological conceptual model in OntoUML into a pure object implementation model in UML and also the instantiation of such model to validate it. In the paper [4], we outlined our approach to the transformation of OntoUML PIM into an ISM of a relational database. In the parallel paper [5], we discuss the details of the transformation of OntoUML Anti-rigid Sortal types. This paper presents the parallel research focused on the transformation of OntoUML Rigid Sortal types (Kinds and Subkinds).

III. Running Example

Our approach to the transformation of the Sortal Rigid universal types in an OntoUML PIM into ISM of a relational database is illustrated on the running example shown in Figure 1. The model shows an excerpt of the domain of an automotive company. The company takes care about

[2]http://www.sparxsystems.com.au/products/ea/

[3]https://www.visual-paradigm.com/

[4]https://github.com/dresden-ocl

[5]http://lci.cs.ubbcluj.ro/ocle/

[6]http://sourceforge.net/projects/useocl/

[7]https://github.com/nemo-ufes/ontouml-lightweight-editor

[8]http://www.menthor.net/menthor-editor.html

[9]http://www.menthor.net/ea-plugin.html

[10]https://zenodo.org/record/51859

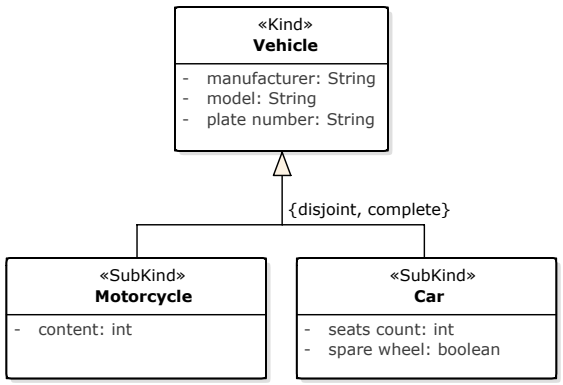


Fig. 1. OntoUML PIM of vehicle types

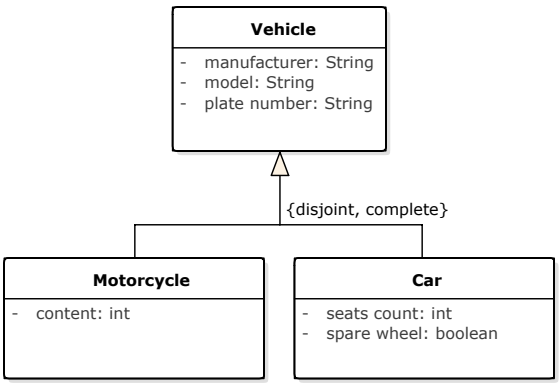


Fig. 2. UML PIM of vehicle types

motorcycles and personal cars in their fleet. These two types of vehicles are represented by the `Motorcycle` and `Car` Subkinds of the common ancestral Kind `Vehicle` defining the identity of a vehicle.

As the company uses only motorcycles and personal cars, the generalization set is `complete`. Also, it is not possible for a single vehicle to be both the motorcycle and the car, therefore the generalization set is `disjoint`.

IV. Our Approach

Our approach to the transformation of a PIM in OntoUML into its realization in a relational database consists of three steps which are discussed in the following sections:

1) subsection IV-A discusses the transformation of an OntoUML PIM into a UML PIM,
2) subsection IV-B discusses the transformation of the UML PIM into a PSM for relational database,
3) subsection IV-C discusses the transformation of the PSM into an ISM of the relational database.

As mentioned in the introduction, it should hold that no information should be lost when transforming from a more abstract model into a more specific one. As OntoUML applies certain constraints based on the OntoUML type used for an entity, these constraints should be carried over to the other models. In our approach, we use OCL to define such constraints in the UML models that cannot be expressed directly in the diagrams.

Although we may formulate a direct transformation from OntoUML into the relational database, the transformation via an auxiliary UML model enables to leverage all the available knowledge (e.g. [26], [24] and tools for transformation of a UML PIM into database models such as Enterprise Architect[11]. Also, various optimizations and refactoring may be applied whenever possible (e.g. when the entity does not hold any attributes, they can be expressed by a mere attribute of the superclass).

In the approach presented here, we assume the (most common) situation where all attributes of the model classes have multiplicities `[1..1]`. In the conclusions, we discuss how the situation changes for different multiplicities.

[11]http://www.sparxsystems.com.au/products/ea/

A. Transformation of OntoUML PIM into UML PIM

This phase of the transformation deals with the transformation of various types of universals in an OntoUML model into a pure UML model while preserving all the semantics defined by the universal types.

In this phase of the transformation, the semantics of the OntoUML model is mostly realized by the multiplicities of the relations between the classes in the UML model.

As various OntoUML universal types define different semantics, they are also transformed in a different manner. However, we discuss only the transformation of Rigid Sortal types (Kinds and Subkinds) and their variants in this paper.

1) Kinds and Subkinds: As both Kind and Subkind universals in OntoUML are rigid, their instances cannot cease to be their instances without ceasing to exists. The same applies in UML for the relation between the instances and their classes. Therefore, the representation of Kinds and Subkinds in UML may stay the same: each $\ll Kind \gg$ and $\ll Subkind \gg$ class is transformed into a standard UML class keeping all its features – attributes and relations.

The resulting transformed PIM into UML for the vehicle domain shown in Figure 1 is shown in Figure 2. Each of the $\ll Kind \gg$ and $\ll Subkind \gg$ classes has been transformed into standard UML class.

2) Generalization sets: A Subkind in OntoUML represents a special case of a Kind or other Subkind, forming a generalization set together with other Subkinds. As both Kinds and Subkinds are rigid, also the generalization set is rigid: when an object is an instance of the Subkind, it is also an instance of its rigid ancestor – a Kind or another Subkind – and it cannot cease to be the instance of any of them without loosing its identity.

Thanks to the rigidity, the generalization sets of $\ll Subkind \gg$ classes in the OntoUML model can be transformed into the standard UML generalization set. Also, the meta-properties `isDisjoint` and `isCovering` of the

generalization set remain the same. The example of this transformation can be seen in Figure 2.

B. Transformation of PIM into PSM

The second step is the transformation of the UML PIM into a PSM of a relational database. The UML Data Model profile – an extension to the UML class diagrams – is used in the examples to define the structure of relational databases in UML [14]. Additional constraints required to preserve the semantics derived from the OntoUML model are defined as OCL invariants, as OCL is part of the UML standard and there are tools supporting the transformation of OCL constraints into database constructs such as DresdenOCL[12]. The basics of this transformation was already discussed in [24]. In this paper, we focus on the transformation of the constraints derived from the OntoUML Rigid Sortal universal types.

In general, when performing transformation from a UML PIM into a PSM of a relational database, classes are transformed into database tables, class's attributes are transformed into table columns and associations are transformed into FOREIGN KEY constraints. Also, PRIMARY KEY constraints are defined for unique identification of individual rows in the tables.

The transformation of classes representing various Kind universals is straightforward – the class with its attributes is transformed into a table with its columns as discussed in [24]. However, more complicated situation arises for the subclasses representing the Subkind universals, as they always form generalization sets. There are multiple standard variants of the transformation of generalization [27], however, they have certain limitations regarding the OntoUML Subkind universal constraints, as discussed in the following sections.

1) Single table: In this variant of generalization realization, the superclass and all its subclasses are realized by a single table. Such table contains the columns for all the attributes of the superclass and all its subclasses. Instances of the superclass are represented by rows with the subclasses' columns containing NULL values, instances of a subclass contain values only in the superclass columns and their respective subclass columns – the other columns remain NULL. Usually, a special column to discriminate the subtypes is also defined in the table. The resulting transformed model of the PIM in Figure 2 is shown in Figure 3, where the column `id` serves as the PRIMARY KEY and the column `type` serves as the discriminator.

As our assumption is that attribute multiplicities are `[1..1]` – as mentioned at the beginning of this section – all columns of a class should be NOT NULL in the table. This can be easily defined by the NOT NULL constraints for the columns of the superclass, as they have values even for the instances of the subclasses. However, the constraints for the subclasses' columns depend on the subclass of the instance, which the row represents – the other columns may contain NULL values. Moreover, all columns of a single subclass – not only a subset of them – should have a value.

[12]https://github.com/dresden-ocl

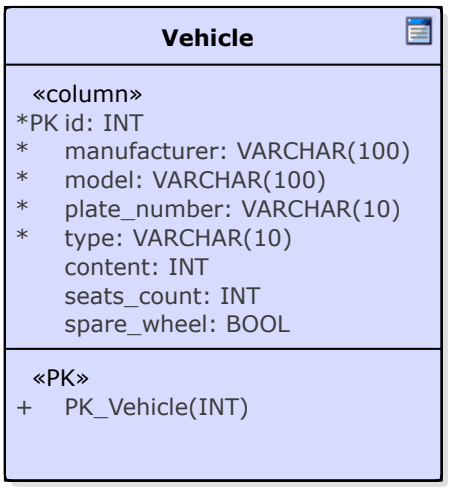


Fig. 3. PSM of vehicle types realized by a single table

As such constraints are not defined on the column level, they cannot be captured directly in the UML model. Instead, they must be defined as additional OCL invariants that are later transformed into their realization in a relational database (see subsection IV-C). For the example in Figure 3, these constraints can be defined for the individual subclasses as shown in Algorithm 1.

Furthermore, the constraints should be defined in respect to the meta-properties of the generalization set according to the following variants:

- `{complete, disjoint}`: For each row, all the columns of the superclass and all the columns of a single subclass must contain NOT NULL values, the other columns must contain NULL values.
- `{complete, overlapping}`: For each row, all the columns of the superclass and all the columns of at least a single subclass cannot contain NULL values.
- `{incomplete, disjoint}`: For each row, all the columns of the superclass must contain NOT NULL values. At the same time, all the columns of at most one of the subclasses may contain NOT NULL values, the other columns must contain NULL values.
- `{incomplete, overlapping}`: For each row, all the columns of the superclass must contain NOT NULL values. All the columns of any subclass may or may not contain NOT NULL values.

Such OCL invariant for the example shown in Figure 3 is shown in Algorithm 2. Because the generalization set is `{disjoint, complete}`, the `Vehicle` columns must contain a value – this is achieved by the NOT NULL constraints of the columns – and the columns of `Motorcycle` class must contain NOT NULL values while the `Car` class must contain NULL values, and vice versa.

The constraints discussed above become the more complicated the more attributes there are in the subclasses because

Algorithm 1 OCL invariants for the NOT NULL constraints of the subclasses

```
context v:Vehicle inv MotorcycleNotNull:
v.type = 'Motorcycle' implies v.content <> OclVoid

context v:Vehicle inv CarNotNull:
v.type = 'Car' implies v.seats_count <> OclVoid and v.spare_wheel <> OclVoid
```

Algorithm 2 OCL invariant for the `{disjoint, complete}` generalization set realized by a single table

```
context v:Vehicle inv MotorcycleOrCar:
def: validMotorcycle:Boolean =
        v.content <> OclVoid and v.seats_count = OclVoid and v.spare_wheel = OclVoid
def: validCar:Boolean =
        v.content = OclVoid and v.seats_count <> OclVoid and v.spare_wheel <> OclVoid
validMotorcycle xor validCar
```

of the exclusivity of the NOT NULL values. Therefore, we would recommend this variant of the transformation only in cases there are not many subclasses and their attributes.

2) Subclasses' tables: In this variant of the transformation, the tables are created only for the subclasses. The attributes of the superclass are transformed into columns in all the tables of all the subclasses. Therefore, each instance of a subclass is able to store also the values of the superclass's attributes along with their own in a single table. Because of this, all the NOT NULL constraints can be easily defined for all columns.

However, in this variant, it is more complicated to ensure the unique values for attributes of the superclass, as the data are distributed in several distinct tables.

Also, this variant cannot be used for an `incomplete` generalization set as it does not allow the storing of an instance of only the superclass. Even if the NOT NULL constraints on the subclasses' columns would not be defined, it would not be clear in which table to store the instance[13].

Moreover, this variant is not suitable for `overlapping` generalization sets either, as storing the data of multiple subclasses to their respective tables also duplicate the data of the superclass.

Therefore, based on the mentioned restrictions and complications, we would not recommend this variant of the transformation in any situation and we will not discuss it anymore.

3) Superclass and subclasses' tables: According to this variant, the superclass and all the subclasses are transformed each into their own table and the individual subclass's tables contain the FOREIGN KEY referring to the superclass's table. This direction is determined by the fact that an instance of the subclass is also an instance of the superclass. Therefore a record in the subclass's table requires exactly one record in the superclass's table – thus being related to `1..1` parent records. More details about determination of the FOREIGN KEY direction based on the multiplicities can be found in [28].

In this variant, the NOT NULL constraints are easier to define, as all columns in a table represent attributes of the same class and they can be expressed by simple NOT NULL constraints defined directly for each column. Still, an additional constraint must be defined for the meta-properties of the generalization set of the subclasses according to the following combinations:

- `{complete, disjoint}`: exactly one row from only one of the subclass's tables refers to the row in the superclass's table.
- `{complete, overlapping}`: at most one row from each of the subclass's tables refers to the row in the superclass's table, but at least one in total.
- `{incomplete, disjoint}`: at most one row from only one of the subclass's tables refers to the row in the superclass's table.
- `{incomplete, overlapping}`: at most one row from each of the subclass's tables refers to the row in the superclass's table.

The restriction of *at most one row from a table* can be realized by a UNIQUE KEY constraint on the FK column; the same column may also be part of the PRIMARY KEY constraint. However, the exclusivity must be checked by a special constraint, still.

In the running example, the subclasses `Motorcycle` and `Car` are transformed into their respective tables (see Figure 4). As their generalization set is `complete` and `disjoint`, the FK column is part of the PRIMARY KEY constraint to make it unique. Furthermore, the constraint shown in Algorithm 3 must be defined.

The other variants of the generalization meta-properties are not discussed here.

[13]Technically, it is possible to designate one of the subclass's tables for this purpose. However, we find this approach inconceptual.

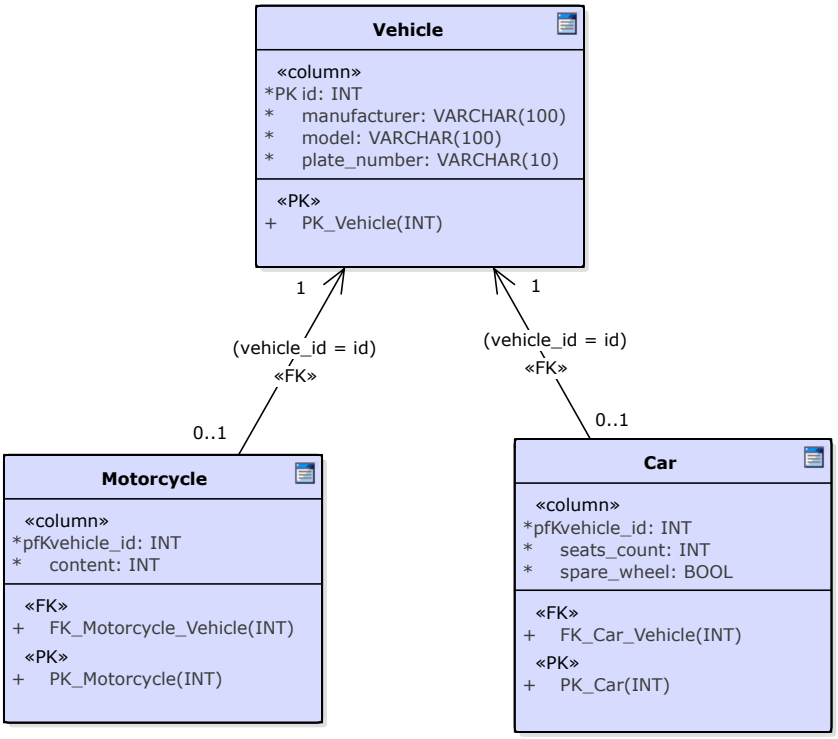


Fig. 4. PSM of vehicle types realized by a separate superclass's and subclesses' tables

Algorithm 3 OCL invariant for the `{disjoint, complete}` generalization set realized by a superclass's and subclasses' tables

```
context v:Vehicle inv MotorcycleOrCar:
def:validMotorcycle:Boolean = Motorcycle.allInstances()->exists(m|m.vehicle_id=v.id)
def:validCar:Boolean = Car.allInstances()->exists(c|c.vehicle_id = v.id)
validMotorcycle xor validCar
```

C. Transformation of PSM into ISM

The last step is the transformation of the PSM of a relational database into an ISM. This model consists of database scripts for the creation of the database tables, constraints and other constructs.

As we have the PSM of the relational database, the transformation is quite easy. Most of the current CASE tools such as Enterprise Architect or Visual Paradigm[14] can be used to generate SQL DDL scripts. These scripts usually include the CREATE commands for the tables, their columns, NOT NULL constraints and PRIMARY and FOREIGN KEY constraints.

However, the OCL invariants defined for the additional constraints require special transformation. Only a few tools currently seem to offer transformation of such constraints – e.g. DresdenOCL[15], OCLE[16] or USE[17].

[14]http://www.visual-paradigm.com/
[15]https://github.com/dresden-ocl
[16]http://lci.cs.ubbcluj.ro/ocle/
[17]http://sourceforge.net/projects/useocl/

Our approach to the realization of the OCL constraints derived from the OntoUML universal types is inspired by the approach for special multiplicity constraints discussed in [24]. Based on that approach, the following constructs may be used to prevent violating the derived constraints:

- *Database views* can be used to query only the valid data meeting the constraints. They do not slow down the DML operations, but they do not prevent inserting data violating the constraints.
- *Updatable database views with CHECK option* can be used to manipulate only the valid data, preventing to create invalid data by insert, update and delete operations. However, the use of such views is restricted by several constraints for the query expression.
- *CHECK constraints* can be used to check the values inserted to various columns of the table, but the common current database engines (e.g. Oracle 11g) do not support subqueries in the CHECK constraint expression and therefore they cannot be used for relational constraints.

- *Triggers* can be defined on the DML operations to prevent creating invalid data in the tables. In the triggers, complex queries and checks can be realized, and therefore they are capable to deal with almost every possible constraint. The constraint checks slow down each DML operation, however as shown in [24], the time increase is typically not substantial.

In [24], the research was focused on the realization of special multiplicity constraints. The same approach, however, may be used also for the realization of the constraints derived from the Rigid Sortal universal types and their generalization sets in OntoUML.

In the following sections, the transformation of the resulting PSMs from subsection IV-B is discussed using the approaches listed above (marked by italics).

1) Single table: For the generalization set transformed into a single database table, the constraint was defined as shown in Algorithm 2. Its realization using the *database view* is simple: query only such rows from the table that have either the `Motorcycle` columns or the `Car` columns filled with values. The resulting database view is shown in Algorithm 4. The WHERE condition filters out such vehicles that are neither motorcycle nor car as well as such vehicles that are both – and it exactly meets the `{complete, disjoint}` property of the generalization set.

Moreover, as the view definition meets the constraints for an *updatable view*, it can be defined WITH CHECK OPTION and used even for DML operations like inserts, updates and deletes. The WITH CHECK OPTION makes the database engine to check the view after each such operation executed on the view and prevents inserting a row that will not be accessible by the view or updating a row to make it inaccessible.

The same effect might be achieved also by defining a *CHECK constraint* which is checked after each operation on the table. The resulting CHECK constraint is shown in Algorithm 5. Using such constraint, it is not possible to create invalid data in the table and therefore the table can be used directly for querying the valid data.

As the CHECK constraint does not contain any subqueries, it is supported by the common database engines without any problems. The realization by triggers would achieve the same results but with more complex definition and slower evaluation. Therefore, it is not worth to use the *triggers* approach in this case.

2) Superclass's and subclasses' tables: For the generalization set transformed into database tables for the superclass and all the subclasses, the constraint was defined as shown in Algorithm 3.

The resulting *database views* for the realization of the OCL invariant are shown in Algorithm 6. The view `Valid_Vehicles` is used to query only such rows from the `Vehicle` table that have a row either in the `Motorcycle` or in the `Car` table referring to it. Therefore, using this view, we can access data about such vehicles that are either a motorcycle or a car, the vehicles having invalid data are hidden from the view.

To query the data of valid motorcycles, the view `Valid_Motorcycles` can be used that filters out invalid motorcycles using the `Valid_Vehicles` view. By analogy, the view `Valid_Cars` can be used to query only the valid cars.

All of these views are updatable – meeting the criteria for an *updatable view* – and therefore they can be defined WITH CHECK OPTION and used to manipulate with the vehicles to prevent creating vehicles without the motorcycle or car data. However, it would not be possible to insert data into any of the views as inserting into the `Valid_Vehicles` would violate the view condition while inserting into the `Valid_Motorcycles` or`Valid_Cars` would violate the FOREIGN KEY constraint. Therefore, the FOREIGN KEY constraint must be defined as *deferrable*, so it is checked at the end of the transaction and not at the time of execution of the command. Then, it is possible to insert data to the `Valid_Motorcycles` or `Valid_Cars` views, first referring to a not-existing vehicle and then to insert data into the `Valid_Vehicles` view.

However, the existence of such views does not prevent the manipulation with the data directly in the tables and thus violating the constraints. Therefore, another database construct should be used. A *CHECK constraint* might have been used as discussed in subsubsection IV-C1, however, as the constraint would contain subquery, it is not supported by the common current database engines [24]. Therefore, the CHECK constraint cannot be actually used in this situation.

Instead, according to the *triggers* approach, triggers might be defined for each of the tables for all the DML operations – insert, update, delete – to check that the operation will not violate the constraint. The following triggers would be needed:

- `BEFORE INSERT ON Vehicle`: This trigger would check there are car data or motorcycle data available in their respective tables for the vehicle. If violated, an error is raised and the operation is cancelled.
- `BEFORE UPDATE OR DELETE ON Motorcycle`: The trigger would check there are no vehicle data in the `Vehicle` table, to which the updated or deleted rows refer. If violated, an error is raised and the operation is cancelled.
- `BEFORE UPDATE OR DELETE ON Car`: The similar trigger should be defined as for the `Motorcycle` table.

Defining such triggers, along with the FOREIGN KEY constraints a PRIMARY KEY constraints, would prevent creating invalid data in the tables during any DML operation. However, the FOREIGN KEY constraints must be defined as *deferrable* – same as for the views – to allow inserting the subclasses' data before inserting the superclass's data or to delete the superclass's data before deleting the subclasses' data.

V. Discussion

As mentioned above, our approach to the realization of the constraints derived from the OntoUML Sortal universal types is based on the approach discussed in [24]. In this paper, the authors discuss possible ways to realize constraints for

Algorithm 4 Database view to query valid data from the combined `Vehicle` table

```
CREATE VIEW MotorcycleOrCar AS
SELECT * FROM Vehicle v WHERE
  (v.content IS NOT NULL AND v.seats_count IS NULL AND v.spare_wheel IS NULL)
  OR
  (v.content IS NULL AND v.seats_count IS NOT NULL AND v.spare_wheel IS NOT NULL)
WITH CHECK OPTION
```

Algorithm 5 CHECK constraint for the combined `Vehicle` table

```
ALTER TABLE Vehicle ADD CONSTRAINT MotorcycleOrCar CHECK
  (v.content IS NOT NULL AND v.seats_count IS NULL AND v.spare_wheel IS NULL)
  OR
  (v.content IS NULL AND v.seats_count IS NOT NULL AND v.spare_wheel IS NOT NULL)
```

Algorithm 6 Database views to query only valid data from the `Vehicle`, `Motorcycle` and `Car` tables

```
CREATE VIEW Valid_Vehicles AS
SELECT * FROM Vehicle v WHERE
  (EXISTS (SELECT 1 FROM Motorcycle m WHERE m.vehicle_id = v.id)
       AND NOT EXISTS (SELECT 1 FROM Car c WHERE c.vehicle_id = v.id))
  OR
  (NOT EXISTS (SELECT 1 FROM Motorcycle m WHERE m.vehicle_id = v.id)
       AND EXISTS (SELECT 1 FROM Car c WHERE c.vehicle_id = v.id))

CREATE VIEW Valid_Motorcycles AS
SELECT v.*, m.content FROM Valid_Vehicles v
  JOIN Motorcycle m ON (v.id = m.vehicle_id)

CREATE VIEW Valid_Cars AS
SELECT v.*, c.seats_count, c.spare_wheel FROM Valid_Vehicles v
  JOIN Car c ON (v.id = c.vehicle_id)
```

special multiplicity values using database views and triggers. The authors also provide results of experiments, proving that their realization guarantees database consistency in context of the multiplicity constraints with just a slight decrease in efficiency.

The OCL constraints derived from the OntoUML Sortal universal types have the same structure – they are based on multiplicities of related objects or their exclusivity. Therefore, also their realization using the views and triggers is very similar. Based on this, we can expect the same impact on the efficiency of the DML operations and queries. However, as our research is not yet fully concluded, experiments are yet to be done to prove that.

Also, in this paper, we focused on the most common situation of mandatory attributes (attribute multiplicity `[1..1]`). In case of optional attributes (minimal multiplicity `0`), some of the constraints will simplify – e.g. the NOT NULL constraints for individual columns representing the attributes of the subclasses (Algorithm 1 and Algorithm 2). On the other hand, collection attributes (attributes with the maximal multiplicity $*$) lead to the realization in the form of relations, references and FOREIGN KEYs.

VI. Conclusions

In this paper, we introduced our approach to the transformation of an OntoUML PIM of application data into an ISM of a relational database. This transformation is separated into three sequential steps: the transformation of an OntoUML PIM into a UML PIM, the transformation of the UML PIM into a PSM for relational database and the transformation of the PSM into an ISM of a relational database.

During these transformations, various options and additional constraints should be defined and realized to maintain the semantics defined by the OntoUML universal types. In this paper, we discussed the details of the transformation of Rigid Sortal universal types – Kinds and Subkinds and their generalization sets – discussing various possible realizations of the constraints derived from the semantics of these OntoUML

constructs. All the variants are described using a running example of a simple OntoUML PIM of vehicle types.

As for the future research, a similar work should be elaborated for the Non-sortal universal types – e.g. Category, Mixin, RoleMixin – and relational constructs – part-whole relations, Relators, etc. Also, combinations of multiple generalization sets of a single universal with various combinations of the meta-properties should be investigated. Finally, experiments should be carried out to study the finer points of individual variants of the constraints realization.

REFERENCES

[1] C. Ghezzi, M. Jazayeri, and D. Mandrioli, *Fundamentals of Software Engineering*, 2nd ed. Upper Saddle River, NJ, USA: Prentice Hall PTR, 2002. ISBN 0133056996

[2] S. J. Mellor, A. N. Clark, and T. Futagami, "Model-driven development," *IEEE Software*, vol. 20, no. 5, p. 14, Sep. 2003.

[3] G. Guizzardi, *Ontological Foundations for Structural Conceptual Models*. Enschede: University of Twente, 2005, vol. 015, no. CTIT Ph.D.-thesis series No. 05-74. ISBN 90-75176-81-3

[4] Z. Rybola and R. Pergl, "Towards OntoUML for Software Engineering: Introduction to the Transformation of OntoUML into Relational Databases," in *Enterprise and Organizational Modeling and Simulation*, ser. LNBIP. CAiSE 2016, Ljubljana, Slovenia: Springer, June 2016, in press.

[5] ——, "Towards OntoUML for Software Engineering: Transformation of Anti-Rigid Sortal Types into Relational Databases," in *Model and Data Engineering*, ser. LNCS, vol. 9893. MEDI 2016, Alméria, Spain: Springer, Sep 2016. doi: 10.1007/978-3-319-45547-1_1. ISBN 978-3-319-45546-4. ISSN 0302-9743 pp. 1–15.

[6] OMG, "MDA guide revision 2.0," http://www.omg.org/cgi-bin/doc?ormsc/14-06-01, Jun. 2014, accessed: 2016-03-10.

[7] J. Arlow and I. Neustadt, *UML 2.0 and the Unified Process: Practical Object-Oriented Analysis and Design (2nd Edition)*. Addison-Wesley Professional, 2005. ISBN 0321321278

[8] T. Stahl, M. Völter, J. Bettin, A. Haase, and S. Helsen, *Model-driven software development: technology, engineering, management*. John Wiley & Sons, 2013. ISBN 0-470-02570-0

[9] M. Brambilla, J. Cabot, and M. Wimmer, "Model-Driven Software Engineering in Practice," *Synthesis Lectures on Software Engineering*, vol. 1, no. 1, pp. 1–182, Sep. 2012. doi: 10.2200/S00441ED1V01Y201208SWE001

[10] A. R. da Silva, "Model-driven engineering: A survey supported by the unified conceptual model," *Computer Languages, Systems & Structures*, vol. 43, pp. 139 – 155, 2015. doi: 10.1016/j.cl.2015.06.001

[11] J. Whittle, J. Hutchinson, M. Rouncefield, H. Burden, and R. Heldal, "Industrial Adoption of Model-Driven Engineering: Are the Tools Really the Problem?" in *Model-Driven Engineering Languages and Systems*, ser. Lecture Notes in Computer Science. Springer Berlin Heidelberg, Sep. 2013, no. 8107, pp. 1–17. ISBN 978-3-642-41532-6, 978-3-642-41533-3. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-41533-3_1

[12] OMG, "UML 2.4.1," http://www.omg.org/spec/UML/2.4.1/, Aug. 2011, accessed: 2016-02-08.

[13] ——, "UML 2.5," http://www.omg.org/spec/UML/2.5/, Mar. 2015, accessed: 2016-02-08.

[14] G. Sparks, "Database Modeling in UML," accessed: 2016-02-02. [Online]. Available: http://www.eetimes.com/document.asp?doc_id=1255046

[15] OMG, "Object constraint language (OCL), version 2.4," http://www.omg.org/spec/OCL/2.4/, Feb. 2014, accessed: 2016-02-23.

[16] M. Richters and M. Gogolla, "OCL: syntax, semantics, and tools," in *Object Modeling with the OCL, The Rationale behind the Object Constraint Language*. Springer-Verlag, 2002. ISBN 3-540-43169-1 pp. 42–68.

[17] J. Cabot and E. Teniente, "Transformation techniques for OCL constraints," *Science of Computer Programming*, vol. 68, no. 3, pp. 179–195, Oct. 2007. doi: 10.1016/j.scico.2007.05.001

[18] G. Guizzardi, "The problem of transitivity of part-whole relations in conceptual modeling revisited," Amsterdam, The Netherlands, 2009.

[19] ——, "Agent roles, qua individuals and the counting problem," *Software Engineering of Multi-Agent Systems*, no. IV, 2006.

[20] B. Goncalves, G. Guizzardi, and J. G. Pereira Filho, "Using an ECG reference ontology for semantic interoperability of ECG data," *Special Issue on Ontologies for Clinical and Translational Research*, 2011. doi: 10.1016/j.jbi.2010.08.007

[21] P. P. F. Barcelos, G. Guizzardi, A. S. Garcia, and M. Monteiro, "Ontological evaluation of the ITU-T recommendation g.805," vol. 18. Cyprus: IEEE Press, 2011. doi: 10.1109/CTS.2011.5898926

[22] G. Guizzardi, G. Wagner, N. Guarino, and M. v. Sinderen, "An Ontologically Well-Founded Profile for UML Conceptual Models," in *Advanced Information Systems Engineering*, ser. Lecture Notes in Computer Science, A. Persson and J. Stirna, Eds. Springer Berlin Heidelberg, Jun. 2004, no. 3084, pp. 112–126. ISBN 978-3-540-22151-7, 978-3-540-25975-6. [Online]. Available: http://dx.doi.org/10.1007/978-3-540-25975-6_10

[23] Z. Rybola and K. Richta, "Transformation of Special Multiplicity Constraints - Comparison of Possible Realizations," in *Proceedings of the Federated Conference on Computer Science and Information Systems*, FedCSIS 2012, Wroclaw, Poland, Sep. 2012. ISBN 978-83-60810-51-4 pp. 1357–1364.

[24] ——, "Possible Realizations of Multiplicity Constraints," *Computer Science and Information Systems*, vol. 10, no. 4, pp. 1621–1646, Oct. 2013. doi: 10.2298/CSIS121210067R

[25] R. Pergl, T. P. Sales, and Z. Rybola, "Towards OntoUML for Software Engineering: From Domain Ontology to Implementation Model," in *Model and Data Engineering*, ser. LNCS, vol. 8216. Amantea, Italy: Springer, Sep. 2013. doi: 10.1007/978-3-642-41366-7. ISBN 978-3-642-41365-0 pp. 249–263.

[26] W. Kuskorn and S. Lekcharoen, "An Adaptive Translation of Class Diagram to Relational Database," in *International Conference on Information and Multimedia Technology, 2009. ICIMT '09*, Dec. 2009. doi: 10.1109/ICIMT.2009.56 pp. 144–148.

[27] P. Rob and C. Coronel, *Database Systems: Design, Implementation, and Management*, 2nd ed. Boyd & Fraser, 1995. ISBN 0-7895-0052-3

[28] Z. Rybola and K. Richta, "Transformation of Binary Relationship with Particular Multiplicity," in *DATESO 2011*, vol. 11. Písek, Czech Republic: Department of Computer Science, FEECS VSB - Technical University of Ostrava, Apr. 2011. ISBN 978-80-248-2391-1 pp. 25–38.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1593–1596
DOI: 10.15439/2016F82
ACSIS, Vol. 8. ISSN 2300-5963

Applying Mutation Testing for Assessing Test Suites Quality at Model Level

Joanna Strug
Faculty of Electrical and Computer Engineering, Cracow University of Technology

ul. Warszawska 24, 31-155 Krakow, Poland
Email: pestrug@cyf-kr.edu.pl

***Abstract*—Models are commonly used in software testing to select test suites. Application of mutation testing at a model level can contribute to reliable and early assessment of the quality of the test suites. It can also support selection of test suites achieving high fault detection rates. The main issue related to using mutation testing at the early development stage is to determine how reliably the quality of test suites can be measured at the model level. The research presented in this paper addresses this problem for object-oriented systems. It focuses on describing an experiment aiming at comparing results of applying mutation testing at a model level with results of applying this technique at an implementation level and presents and discusses the outcomes of the experiment. The paper presents also mutation operators applicable at the model level.**

I. Introduction

MODELS play an important role in developing object-oriented software systems. They are also commonly used by researchers and practitioners involved in software testing as a source of data for selecting tests [25]. As the main goal of software testing is to detect faults in a tested system, an adequate assessment of suites of tests provided by any test generation approach is essential. Mutation testing is a well established techniques that helps to assess the quality of test suites with regard to their ability to detect faults [7]. To assess test suite quality, by means of this technique, a number of faulty versions of the system (called mutants) is generated, by introducing small changes into the original system, and executed against tests from the suite. The ratio of the number of mutants detected (killed) by these tests over the total number of non-equivalent mutants (called a mutation score for a test suite) determines the test suite ability to detect faults. The higher the mutation score for a test suite, the better the suite is in detecting faults. Although mutation testing is an effective test suite assessment technique, its application area is mostly limited to implementation level and high computational costs still prevent it from becoming a practical approach.

Application of mutation testing at a model level could be a good alternative or at least a valuable addition to the current practice in assessing the quality of test suites derived from models. It would allow to use the most reliable assessment technique at the same early level and may also lower the costs of generation and execution of mutants, as system models are less complex than the implementations.

However, two issues should be considered before applying the approach in practice:

1) choice of models for representing a system, and
2) evaluation of the reliability of results of a model level test suite quality assessment.

The research presented in this paper concerns object-oriented systems, therefore UML/OCL class diagrams were used to describe the systems at the model level.

The second issue is essential, when increasing the level of abstraction at which mutation testing is applied. A model represents only certain aspects of a final system, thus it is not possible to predict all implementation level faults based only on faults that appear at a model level. Moreover, faults introduced into a model target only features of the model, not of the language used to implement the system, and hence application of mutation testing at a model level assesses a test suite ability to detect faults specific to that level. Even in context of object-oriented systems modeled with UML/OCL and implemented in an object-oriented language, there are significant differences between both description formalisms. Thus, it is unclear if test suite assessment results (provided in the form of a mutation score) obtained at model level are sufficiently reliable to be accepted as a measure of a test suite quality in terms of its ability to detect faults in a final, implemented system. To the author best knowledge, the problem has not been studied before. The paper presents an approach to model level, UML/OCL-based assessment of test suite quality and describes an experiment carried out to address the problem. It provides a description of mutation operators applicable to UML/OCL models, the procedures for conducting the experiment and its results.

II. Related Work

Mutation testing was originally introduced at the implementation level [7] and the majority of papers describing different aspects of mutation testing dealt with problems concerning implementation level mutation only (a survey of such papers can be found in [11]). However, application of mutation testing at model level seems to gain popularity [1], [2], [16], [18], [19], [23], [24]. Authors of the approaches have focused mainly on selecting tests, but some of them have addressed also the problem of assessing tests at the model level [24], [23] and discussed selected aspects related to the problem of

978-83-60810-90-3/$25.00

assessing tests quality at different levels [23]. Although the problem considered by authors of the work in [23] is partially related with the one studied in this work, their approach is not applicable in the context of object-oriented systems, because the formalisms used in the paper cannot support adequately object-oriented aspects of the systems.

Relevant to this research are mainly papers describing application of mutation testing to UML and OCL based models [1], [2], [14], [16] and papers providing information that can help to design mutation operators applicable to UML and OCL. The set of UML/OCL related mutation operators introduced by the author in [17] was developed based on the fault taxonomy for UML [9], traditional mutation operators, operators adapted from other formalisms [8], [15] and operators defined for specifications [5], [12] and contracts [12] and was supplemented with five new OCL-specific operators.

The work presented in this paper differs from other works concerning model or implementation level mutation testing, as it targets UML/OCL class diagrams (standard models in developing object-oriented systems) and attempts to find out if test suites ability to detect implementation level faults can be assessed reliable at the model level.

III. Experimental Evaluation of a Reliability of Model Level Test Suite Quality Assessment Results

The goal of the research was to determine how reliably one can assess test suite quality, in terms of its ability to detect real faults in an object-oriented software system, by assessing the test suite using mutation testing at a the model level. Empirical studies on mutation testing have provided evidences that application of the technique at the implementation level provides adequate measurement of test suite quality (in terms of its ability to detect real faults in a final, implemented system) [6], [13]. Thus, the implementation level measurements can be referred to to determine the reliability of the test suite quality assessment results obtained at the model level.

A. Experimental measures

Application of mutation testing provides a mutation score for a test suite. The mutation score is a quantitative measure of the suite ability to detect mutants. For the rest of the paper let T denote a test suite, $MS_{IL}(T)$ denote a mutation score calculated for T at the implementation level and $MS_{ML}(T)$ denote a mutation score calculated for T at the model level.

Implementation level mutation score $MS_{IL}(T)$ for a test suite T expresses its ability to detect implementation level mutants (and thus their ability to detect real faults) and is defined in the following way:

$$MS_{IL}(T) = \frac{MI_D}{MI_T}, where$$

- MI_D is the number of implementation level mutants detected by T,
- MI_T is the total number of non-equivalent, implementation level mutants.

Model level mutation score $MS_{ML}(T)$ for a test suite T expresses its ability to detect model level mutants and is defined in the following way:

$$MS_{ML}(T) = \frac{MM_D}{MM_T}, where$$

- MM_D is the number of model level mutants detected by T,
- MM_T is the total number of non-equivalent, model level mutants.

The reliability of a model level test suite assessment result is measured for a test suite T by comparing the value of $MS_{ML}(T)$ with the value of $MS_{IL}(T)$.

B. Mutation Operators

Mutation operators are defined as transformation rules that produce faulty versions (so called mutants) of a program or a model [7]. Each operator can produce a number of mutants by changing instances of some construction of the formalism to which the operator is applied.

Within the work generation of mutants was controlled by two sets of mutation operators. At the implementation level mutation operators designed for Java and implemented in mujava [15] were used, and at the model level operators modifying UML/OCL class diagrams were used. The second set was divided into two groups: class diagram related mutation operators, and OCL related mutation operators.

The first group consists of the following operators:

- Hiding attribute deletion (IHD) - deletes in a subclass an attribute having the same name and type as an attribute in a parent class,
- Hiding attribute insertion (IHI) - inserts in a subclass an attribute having the same name and type as an attribute in a parent class,
- Attribute multiplicity change (CAMC) - changes a multiplicity of an attribute,
- Operation arguments order change (OAO) - changes the order of arguments in an operation definition,
- Operation arguments type replacement (ADR) - changes a declared type of a method argument to the parent of the originally declared type,
- Overriding operation deletion (IOD) - deletes an overriding operation in a subclass,
- Generalization association deletion (GAD) - deletes a generalization association between two classes,
- Generalization association direction change (GDC) - changes a direction of a generalization association,
- Association type replacement (ATR) - replaces a type of an association with another type,
- Association end multiplicity change (EMC) - changes multiplicity of an association end to other one,
- Association end class replacement (ECR) - replaces an association end class with a parent class or a subclass,
- Association role swap (ARS) - swaps role names of two associations between the same two classes or their subclasses.

The second group consists of the following operators:

- Operand Replacement Operator (ORO) - replaces an operand with another one, applies also for components of a navigation path,
- Arithmetic Operator Replacement (AOR) - replaces a binary arithmetic operator with another one,
- Arithmetic Operator Insertion (AOI) - inserts an unary arithmetic operator,
- Arithmetic Operator Deletion (AOD) - deletes an unary arithmetic operator,
- Relational Operator Replacement (ROR) - replaces a relational operator with another one,
- Conditional Operator Replacement (COR) - replaces a conditional operator with another one, supports operators: and, or, xor,
- Conditional (unary) Operator Insertion (COI) - inserts an unary conditional operator (not),
- Conditional (unary) Operator Deletion (COD) - deletes an unary conditional operator (not),
- @pre Deletion (POD) - deletes @pre operator,
- @pre Insertion (POI) - inserts @pre operator,
- Collection Operation Replacement (OCR) - replaces an invocation of a collection operation with another one,
- Collection Operation Deletion (OCD) - deletes an invocation of a collection operation,
- Contextual Instance Replacement (CIR) - replaces a contextual instance with another one.

C. Experimental procedures

The experiment was divided into two stages (Fig. 1):

1) model level test suites quality assessment, and
2) implementation level test suites quality assessment.

The experiment was performed on six experimental test suites provided for two object-oriented systems. The experiment, at each stage, was carried out following the same scenario, but dealt with the systems at different levels of abstraction (i.e. models or implementations) and was supported by different tools.

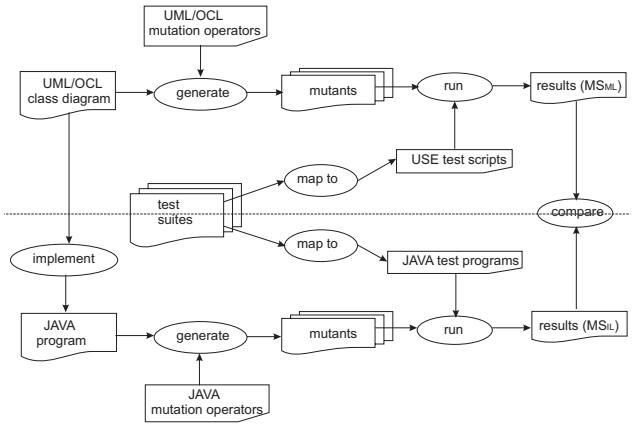


Fig. 1. An outline of the experiment

In the first stage each system was specified in a form of a UML/OCL class diagram, according to USE notation and each test suite was described as a USE command script [10]. Next, for each system, mutants of the system model were manually generated by applying mutation operators designed for UML/OCL models, and executed against tests from the test suites prepared for the system. The undetected mutants were then analyzed manually to remove the equivalent ones and for each test suite T its model level mutation score ($MS_{ML}(T)$) was calculated. At this stage the execution of the mutants was automated by use of USE (UML based Specification Environment) [10].

At the beginning of the second stage the UML/OCL models developed in the first stage were used to implement the systems in Java and the test suites for the systems were mapped into muJava test programs [15]. Next, for each system, its implementation was mutated by applying mutation operators defined for Java. The mutants were then executed against the test suites provided for the system and implementation level mutation score ($MS_{IL}(T)$) for each test suite T was calculated. This stage was fully automated due to the use of mujava [15].

Finally, for each test suite T the values of $MS_{ML}(T)$ and $MS_{IL}(T)$ were compared. Results of the experiment are presented and briefly discussed in Section III-D.

D. Experimental results and discussion

During the experiment each test suite T was assessed twice: first to get its model level mutation score ($MS_{ML}(T)$) and then to get its implementation level mutation score ($MS_{IL}(T)$). Table I shows the results of the experiment.

As it can be observed in Table I, for each test suite T the values of its $MS_{ML}(T)$ and $MS_{IL}(T)$ are very close to each other, in fact the results show that for each assessed test suite its model level mutation score differs from its implementation level mutation score by no more that 3%.

The results of the experiment let us to observe that:

- the model level mutation score of all test suites, but T4, was slightly (between 0.015 and 0.028) higher than their implementation level mutation score, and
- the difference between $MS_{ML}(T)$ and $MS_{ML}(T)$ for test suites that reached the model level mutation score over 0.80 remained at nearly constant level of about 0.02.

Both observations seem to suggest that the model level test suite assessment results may be seen as a reliable measurement of the test suite quality in general. The slightly higher fault detection rate observed for most of the test suites at the model level was to be expected, as the UML/OCL models do not define the processing of data as the implementations.

TABLE I
MUTATION SCORE FOR EXPERIMENTAL TEST SUITES

T	T1	T2	T3	T4	T5	T6
$MS_{ML}(T)$	0.78	0.87	0.90	0.76	0.83	0.84
$MS_{IL}(T)$	0.75	0.85	0.89	0.79	0.81	0.83

Nevertheless, the nearly constant difference between the model level and the implementation level results seems to indicate a regularity that would predict the quality of test suites based on the assessment results obtained at the model level alone.

The results obtained for test suites T1 and T4 shows that, for test suites attaining low model level mutation score, the test suite quality assessment performed at model level provides less predictable results than for the suites attaining higher model level mutation score. However, neither the overestimation of the quality of T1 nor the underestimation of the quality of T4 did not exceeded the 3% threshold. Moreover, a test designer, having developed a suite of such a low quality as T1 and T4, would most likely tend to improve it to achieve better score. Thus, it seems that the irregular behavior of T1 and T4 does not contradict the earlier conclusion regarding the reliability of model level test suite assessment results.

IV. Conclusions and Future Works

Mutation testing is an effective and reliable technique for assessing test suite quality with regard to their ability to detect faults specific to the given level, but a transferability of the assessment results between different levels of abstraction was not evaluated before. An experimental way to assess the relability of results obtained at model level was proposed in this paper. The results of the experiment let us to presume that for object-oriented systems, modeled in a form of UML/OCL class diagrams, the test suite's ability to reveal real faults can be reliably assessed at the model level. However, more experiments on larger systems should be carried out to verify the preliminary conclusions.

Future research concerning model level mutation testing should deal with the costs reduction problem. It seems that the migration to the model level alone lowers the number of generated and executed mutants, but other techniques should also be considered. The most efficient techniques should be the ones taking into account individual characteristics of modeled systems, such as proposed in [20], [21], [22].

Works on applying mutation testing at the model level should also include development of tools supporting generation of mutants. Availability of such tools would significantly increase the possibility of adapting such approach in practice.

References

[1] B. Aichernig and P. Salas, "Test case generation by ocl mutation and constraint solving," *in 5th International Conference on Quality Software,* Melbourne, 2005, pp. 64-71, http://dx.doi.org/10.1109/QSIC.2005.63.

[2] B. Aichernig, H. Brandl, E. Jobstl, W. Krenn, R. Schlick, S. Tiran, "Killing strategies for model-based mutation testing," *Software Testing, Verification and Reliability,* vol. 25(8), 2015, pp. 716-748, http://dx.doi.org/10.1002/stvr.1522.

[3] F. Belli, C. J. Budnik, A. Hollmann, T. Tuglular, W. E. Wong, "Model-based mutation testing - Approach and case studies," *Science of Computer Programming,* vol 120(1), 2016, pp. 25-48, http://dx.doi.org/10.1016/j.scico.2016.01.003.

[4] P. Black, V. Okun, Y. Yesha, "Mutation operators for specifications," *in 5th IEEE International Conference on Automated Software Engineering,* Grenoble, 2000, pp. 81-88, http://dx.doi.org/10.1109/ASE.2000.873653.

[5] A. Brucker, M. Krieger, B. Wolff, "A specification-based test case generation method for uml/ocl," *in International Conference on Models in Software Engineering,* Oslo, 2011, pp. 334-348, http://dx.doi.org/10.1007/978-3-642-21210-9-33.

[6] M. Daran and P. Thvenod-Fosse, "Software error analysis: A real case study involving real faults and mutations," *in ACM SIGSOFT international symposium on Software testing and analysis,* Mission Beach, CA, 1996, pp. 158-177, http://dx.doi.org/0.1145/229000.226313.

[7] R. A. DeMillo, R. J. Lipton, F. G. Sayward, "Hints on test data selection: Help for the practicing programmer," *Computer,* vol. 11(4), 1878, pp. 34-41, http://dx.doi.org/10.1109/C-M.1978.218136.

[8] A. Derezinska, "Object-oriented mutation to asses the quality of tests, *in 29th Euromicro Conference,* Belek, 2003, pp. 417-420, http://dx.doi.org/10.1109/EURMIC.2003.1231626.

[9] T. Dinh-Trong, S. Ghosh, R. France, B. Baudry, F. Fleurey, "A taxonomy of faults for uml designs," *in 2nd MoDeVa workshop - Model design and ValidationModel Design and Validation Workshop,* Montego Bay, 2005.

[10] M. Gogolla, F. Buttner, M. Richters, "Use: A uml-based specification environment for validating uml and ocl," *Science of Computer Programming,* vol. 69, 2007, pp. 27-34, http://dx.doi.org/10.1016/j.scico.2007.01.013.

[11] Y. Jia and M. Harman, "An analysis and survey of the development of mutation testing," *IEEE Transaction on Software Engineering,* vol. 37(5), 2011, pp. 649-678, http://dx.doi.org/10.1109/TSE.2010.62.

[12] Ying Jiang, Shan-Shan Hou, Jinhui Shan, Lu Zhang, Bing Xie, "An approach to testing black-box components using contract-based mutation," *Journal of Software Engineering and Knowledge Engineering,* vol. 18(1), 2008, pp. 93-117, http://dx.doi.org/10.1142/S0218194008003556

[13] R. Just, D. Jalali, L. Inozemtseva, M. D. Ernst, R. Holmes, G. Fraser, "Are Mutants a Valid Substitute for Real Faults in Software Testing?," *22nd ACM SIGSOFT International Symposium on Foundations of Software Engineering,* Hong Kong, 2014, pp. 654-665, http://dx.doi.org/10.1145/2635868.2635929.

[14] W. Krenn, R. Schlick, S. Tiran, B. Aichernig, E. Jobstl, H. Brandl, "MoMut::UML Model-Based Mutation Testing for UML," *in 8th International Conference on Software Testing, Verification and Validation,* Graz, 2015, pp. 1-8, http://dx.doi.org/10.1109/ICST.2015.7102627

[15] Yu-Seung Ma, A. J. Offutt, Yong-Rae Kwon, "Mujava: An automated class mutation system," *Software Testing, Verification and Reliability,* vol. 15(2), 2005, 97-133, http://dx.doi.org/10.1002/stvr.v15:2.

[16] R. Schlick, W. Herzner, E. Jobstl, "Fault-based generation of test cases from uml-models approach and some experiences," *in 30th International Conference on Computer safety, Reliability, and Security,* Naples, 2001, pp. 270-283, http://dx.doi.org/10.1007/978-3-642-24270-0-20.

[17] J. Strug, "Classification of mutation operators applied to design models," *Key Engineering Materials,* vol. 572, 2014, pp. 539-542, http://dx.doi.org/10.4028/www.scientific.net/KEM.572.539.

[18] J. Strug, "Mutation testing approach to negative testing," *Journal of Engineering,* vol. 2016, 2016, http://dx.doi:10.1155/2016/6589140.

[19] J. Strug, "Mutation testing approach to evaluation of design models," *Key Engineering Materials,* vol. 572, 2014, pp. 543-546, http://dx.doi.org/10.4028/www.scientific.net/KEM.572.543.

[20] J. Strug J and B. Strug, "Machine learning approach in mutation testing," *Testing Software and Systems,* vol. 7641 of LNCS, Springer, 2012, pp. 200-214, http://dx.doi.org/10.1007/978-3-642-34691-0-15.

[21] J. Strug J and B. Strug, "Using structural similarity to classify tests in mutation testing," *Applied Mechanics and Materials,* vol. 378, 2013, pp. 546-551, http://dx.doi.org/10.4028/www.scientific.net/AMM.378.546 .

[22] J. Strug and B. Strug, "Classifying mutants with decomposition kernel," LNCS, Springer Berlin Heidelberg, vol. 9692, 2016, pp. 644-654, http://dx.doi.org/10.1007/978-3-319-39378-0-55

[23] M. Trakhtenbrot, "New mutations for evaluation of specification and implementation levels of adequacy in testing of statecharts models," *in Testing: Academic and Industrial Conference Practice and Research Techniques,* Windsor, 2007, pp. 151-160, http://dx.doi.org/10.1109/TAIC.PART.2007.23.

[24] S. Weissleder and B. H. Schlingloff, "Quality of automatically generated test cases based on ocl expressions," *in 1st International Conference on Software Testing, Verification, and Validation,* Los Alamitos, 2008, pp. 517-520.

[25] M-F. Wendland, "Abstractions on Test Design Techniques," *in Federated Conference on Computer Science and Information Systems,* Warsaw, 2014, pp. 1575-1584, http://dx.doi.org/10.15439/2014F316.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1597–1602
DOI: 10.15439/2016F299
ACSIS, Vol. 8. ISSN 2300-5963

GRAD: A New Graph Drawing and Analysis Library

Renata Vaderna, Igor Dejanović, Gordana Milosavljević
Faculty of Technical Sciences, University of Novi Sad, Trg Dositeja Obradovića, Novi Sad, Serbia
Email: vrenata, igord, grist@uns.ac.rs

***Abstract*—Several important choices need to be made during the development of domain-specific languages, including the one regarding which concrete syntax to implement. There are several alternatives, with graphical and textual syntaxes being the most common ones. Having in mind that the developers and domain experts often have different preferences, supporting both is sometimes the best option. This means that models created using textual editors might need to be opened using separately developed graphical editors. Graphical elements corresponding to model elements must then be automatically created and positioned. Doing so in an aesthetically pleasing way requires usage of graph layout algorithms. Since implementing them is not an easy task, most developers have to rely on existing solutions. There are many Java libraries which have such capabilities, but they all have certain limitations and room for improvement, some of which are addressed in a new graph drawing and analysis library presented in this paper.**

I. Introduction

DOMAIN-SPECIFIC languages (DSLs) are computer languages specialized to a particular domain [1]. Development of such languages also includes the decision of how to present the concrete syntax to the users, who can be both developers and non-technical domain experts. There are several alternatives, with the most popular being the textual and graphical ones.

Each of these choices has its own advantages and disadvantages, so supporting both is the best solution at times. The textual concrete syntaxes can express any formal language, help in understanding all technical details of a DSL and are often preferred by the developers. On the other hand, end-users, who work in a non-technical domain, don't find them particularly appealing. These users generally prefer the graphical concrete syntax, which makes it possible to design DSL models using a completely functional graphical editor [2]. Graphical concrete syntaxes, if designed correctly, are intuitive and easy to understand. Having all of this in mind, it can be concluded that if a DSL needs to appeal to both developers and non-technical end-users, both textual and graphical concrete syntaxes should ideally be implemented. A similar conclusion was reached in [3], where the authors discussed the positive and negative experiences of using both the textual syntax, described in [4] and a graphical one to define the static structure of database applications.

Implementing both syntaxes leads to one problem: what happens if a part of the model is specified using the textual syntax and needs to be viewed and/or edited using the graphical editor? Graphical elements corresponding to previously described concepts need to be created automatically. These elements have additional visual properties, including positions which need to be calculated. This can be accomplished by applying a layout algorithm.

Implementing even the simplest of layout algorithms that would guarantee a somewhat pleasing arrangement of elements requires excessive knowledge of graph theory and can be rather time consuming. This is why the developers often rely on existing solutions. This paper focuses on libraries for the Java programming languages, but there are many similar ones for other languages like C/C++ and Python. The most popular open-source libraries offering the possibility of laying out elements of a diagram for Java projects include JUNG framework, JGrapX and Prefuse. All of these solutions put emphasis on visualization, providing their own visual components and thus strongly coupling layout capabilities with them. This makes the integration with separately developed graphical editors overly complex [5]. Furthermore, they only support a small number of different classes of layout algorithms, despite offering several algorithms belonging to the same classes. Even though the available algorithms can be used to lay out any diagram with acceptable results, it can be noticed that there is room for improvement. Certain classes of layout algorithms were designed with the goal of getting excellent results when applied to diagrams satisfying some special conditions (e.g. planar, straight-line, symmetric, rectangular). The mentioned libraries implement very few of them. Also, they don't offer a way of automatically choosing an appropriate algorithm based on properties of the diagram or on the wishes of the users regarding diagram aesthetics.

In order to address the mentioned issues, we are developing another graph drawing and analysis Java library, called GRAD (GRaph Analysis and Drawing) [6]. GRAD's main goals are to:

- offer a large number of different graph drawing algorithms, including some that haven't been implemented in Java yet
- provide a very quick and easy way to lay out elements of any existing graphical editor
- offer algorithms for graph analysis, which can later be used to automatically select a suitable layout algorithm
- enable the users to specify aesthetic criteria and automatically choose an appropriate layout algorithm based on their wishes

978-83-60810-90-3/$25.00

GRAD is not intended to be used as a visualization tool, but it also provides a simple graphical editor which can be used for familiarization with different algorithms.

The rest of the paper is structured as follows. Section 2 gives an overview of basic graph theory concepts, graph drawing aesthetic criteria, and different classes of graph drawing algorithms. Section 3 showcases some popular Java graph drawing and analysis libraries. Section 4 presents GRAD. Finally, section 5 concludes the paper and outlines future work.

II. Graph Drawing Aesthetics and an Overview of Graph Layout Algorithms

In the following section a short overview of the graph aesthetic criteria and the most popular classes of graph layout algorithms will be given. Firstly, the most important concepts which will be referenced later will be defined.

A. Basic graph drawing theory definitions

A graph (V, E) is an ordered pair consisting of a finite set V of vertices and a finite set E of edges, that is, pairs (u, v) of vertices [7]. If each edge is an unordered (ordered) pair of vertices, the graph is undirected (directed). A graph is simple if it doesn't contain any edges that join a vertex to itself or more than one edge connecting the same two vertices (multiple edges). A graph is said to be connected if there is a path from any vertex to any other vertex in the graph. A biconnected graph is a connected graph which has no vertices whose removal would disconnected it. Graphs which contain at least one cycle are called cyclic graphs, while the ones that do not are known as acyclic. A tree is a connected acyclic graph. Finally, a graph is planar if it can be drawn in a plane without graph edges crossing. A planar drawing partitions the plane into connected regions called faces.

The process of creating a drawing of a graph from the underlying structure is known as automatic graph layout. There is a great number of graph layout algorithms, with plenty of researchers still working on discovering new and enhancing existing ones. The quality of an algorithm is determined based on its computational efficiency as well as various aesthetic criteria. The following sections will give an overview of the mentioned criteria and the most popular layout methods.

B. Aesthetic Criteria

Many different quality measures or aesthetic criteria have been defined for graph drawings. Authors often optimize certain aesthetics claiming that the resulting drawing is therefore more understandable and more visually pleasing to a human observer. The most common criteria includes the following: Minimization of the number of edge crosses, maximization of the minimal angle between edges extending from a node, minimization of the total number of bends in polyline edges, even distribution of edges within a bounding box, appropriate lengths of edges, neither too short nor too long, similar length of edges, same flow of edges in directed graphs (as much as possible),orthogonality, and symmetry [8].

Some layout methods put emphasis on one of the measures trying to produce a drawing which, for example, has no edge crosses (planar drawing) or is maximally symmetric, while the other ones attempt to optimize as many as possible. The desired aesthetic criterion or criteria can be the deciding factor in choosing an appropriate layout algorithm. However, certain layout algorithms which insist on a particular aesthetic criterion might not be applicable to all graphs. Obviously, it is not possible to produce a drawing with no edge intersections of a graph which is not planar. Therefore, properties of the graph which is to be laid out can also be an important indicator of the best choice of the algorithm.

C. An Overview of Graph Layout Algorithms

There is a very large number of different classes of graph layout algorithms and the following paragraphs will present the most popular ones.

Tree drawing is one of the best studied areas of graph drawing. That is not surprising since automatic generation of drawings of trees finds many practical applications. Namely, a tree whose vertices represent entities and whose edges represent relationships is a typical data structure for modeling hierarchical information. There are various approaches to drawing trees and their detailed overview and comparison can be found in [9].

A *circular drawing* of a graph is its visualization where it is partitioned into clusters whose nodes are placed onto the circumference of an embedding circle. Each edge is drawn as a straight line. Simply placing nodes on a circumference of a circle might result in a drawing which is not particularly aesthetically pleasing due to a very large number of edge crossings, which is why there are techniques which also minimize this number when determining positions of the nodes [10].

Symmetric graph drawing algorithms aim to draw a graph with nontrivial symmetry, or, more ambitiously, with as much symmetry as possible. Some consider symmetry as one of the most important aesthetic criteria which clearly reveals the structure and properties of a graph. For example, graphs in textbooks on graph theory are normally drawn symmetrically and a symmetric drawing is in some cases preferred over a planar one.

Planar straight-line drawing algorithms rely on the fact that if a graph can be drawn with no crossings using edges of an arbitrary shape, then it can be drawn in the same way using only straight-line segments. Convex drawings are planar straight-line drawings where all faces are drawn as convex polygons. In [11] the author claims that the convex drawings of planar graphs make it possible for readers to easily and rapidly recognize structures of the graphs, such as adjacency of vertices.

Planar orthogonal and polyline drawing algorithms focus on angular resolution as the most important aesthetic criterion. Orthogonal drawings only use horizontal and vertical line segments for edges and are, therefore, often quite visually pleasing. A more specific type of orthogonal drawings are

rectangular drawings, which also make sure that each face is drawn as a rectangle. However, they have a pretty serious limitation of only being applicable to graphs which don't contain a vertex whose degree is higher than four. Polyline drawing are more general and don't have the mentioned disadvantage. They usually focus directly on sizes of the angles (which should not be smaller than some fixed threshold) and not on the type of edges.

Force-directed algorithms are among the most important and most flexible algorithms. Unlike many previously mentioned ones, which can only be applied if a graph is planar or satisfy some other specific conditions, force-directed algorithms can be used to calculate layouts of all simple undirected graphs. They only need the information contained within the structure of the graph itself.

Graphs drawn with these algorithms tend to be aesthetically pleasing, exhibit symmetries, and tend to produce crossing-free layouts for planar graphs. There are many force-driven algorithms, the most popular of which include the spring layout method of Eades [12], Kamada-Kawai [13] and Fruchterman-Reingold [14] methods.

Hierarchical drawing algorithms can be used when dealing with directed graphs (or digraphs) which represent hierarchies. These algorithms name uniform "flow" of edges as one of their main goals. More precisely, the edges should either go from left to right or top to bottom, depending on a particular application.

III. Related Work

There are quite a few libraries for graph analysis and visualization for Java. The next section will present the most popular ones, primarily focusing on their layout capabilities and mentioning implemented graph analysis algorithms which could be used to determine the best choice of the drawing algorithm.

It is important to mention that visualization tools which generate static drawings of graphs in a variety of output formats will not be taken into consideration since they are not suitable for this particular purpose. Furthermore, commercial solutions will not be considered since our focus is on open-source ones.

A. JUNG Framework

JUNG — the Java Universal Network/Graph Framework [15] is an open-source software library that provides a common and extendible language for modeling, analysis, and visualization of data that can be represented as a graph or network. JUNG framework is licensed under the permissive BSD license.

The current distribution of JUNG includes implementations of a number algorithms from graph theory, data mining, and social network analysis. However, most of them are of little importance to this research. Only Dijktra's shortest path and decomposition of a graph into biconnected components can be singled as potentially useful in the mentioned case.

JUNG framework offers implementations of several layout algorithms, some of which are quite complex. Most importantly, these include three tree layout algorithms and a number of force-directed ones. The tree layout algorithms include the following: an implementation of a level-based approach, radial tree method, and the balloon method. The radial tree method displays a tree structure in a way that expands outwards, radially. The balloon method positions vertices using associations with nested circles or "balloons". The force-directed algorithms are the already mentioned popular and flexible spring method, Kamada-Kawai and Fruchterman-Reingold, as well as an algorithm based on Bernd Meyer's self-organizing graph methods [16]. JUNG framework also contains a relatively basic circle layout drawing algorithm, which simply places vertices on a circumference of a circle of a given radius.

B. JGrapX

JGraphX is a Java Swing graph visualization library which is also licensed under the BSD license. JGraphX provides visualization and interaction with node-edge graphs, as well as a decent number of algorithms for graph analysis, such as graph traversal, forming the minimum spanning tree and Dijkstra's shortest path [17]. The minimum spanning tree is defined as the set of all vertices with minimal lengths that forms no cycles. Graph traversal includes deep-first search and bread-first search, both of which construct spanning trees with certain properties useful in other graph algorithms.

JGraphX provides various usable implementations of graph drawing algorithms. Similarly to JUNG framework, these include a tree and several force-directed layouts, but also a hierarchical one meant to be used if a graph is too complex to be laid out using the tree drawing algorithm. The tree layout in question is the compact tree layout, which improves the standard level-based approaches by trying to make the resulting drawing as compact as possible. Furthermore, JGraphX provides two force-directed layout algorithms: fast organic and organic. The fast organic method is best applied to smaller graphs with a more regular structure, but is supposed to be one of the faster force-directed layouts. The organic layout is one of the most complex algorithms implemented in JGraphX and is based on Davidson and Harel's simulated annealing layout [18].

JGraphX doesn't stop there and offers a very nicely implemented hierarchical layout. This implementation not only positions the vertices, but also routes the edges.

C. Prefuse

Prefuse is a software framework for creating dynamic visualization of both structured and unstructured data, that provides theoretically-motivated abstractions for the design of a wide range of visualization applications [19]. Like other mentioned libraries, Prefuse is also licensed under the BSD license.

Prefuse is bundled with a library which, among other actions, provides a host of layout and distortion techniques.

Available layout algorithms include random, circular, grid-based, forced-directed, and several tree ones.

The force-directed layout positions graph elements based on a physical simulation of interactive forces acting on bodies. The force simulator used to drive this layout can be set explicitly, allowing custom force-directed layouts to be created.

Tree layouts provided by Prefuse are previously mentioned balloon and radial algorithms, as well as an additional node-link tree layout, which lays out a rooted tree so that each depth level of the tree is on a shared line.

Based on the previous overview, it can be noticed that the mentioned libraries offer many layout algorithms which would be a good addition to any graphical editor in need for such features. However, all of them put heavy emphasis on data visualization and thus strongly couple it with graph drawing algorithms. Therefore, simply calling the desired algorithm and retrieving the results (positions of vertices and edges determined during the execution of the algorithm) requires an understanding of how a library works. GRAD aims to offer a solution to this problem.

Moreover, certain classes of graph drawing algorithms are substantially represented in these libraries (like tree and force-directed ones), while the other classes are barely or not present at all. For example, no symmetric, straight-line or orthogonal algorithms are available. Out library aims to remedy this and offer implementations of certain algorithms, that, according to our knowledge, have not been implemented in Java yet.

Finally, none of these libraries offer a way of automatically choosing an appropriate layout algorithm based on properties of the graph or according to the desired aesthetic criteria specified by the users. Implementing both of these features is among the goals of our solution. It can also be pointed out that none of the libraries provide many graph analysis algorithms which are of great significance to graph drawing.

IV. GRAD (Graph Analysis and Drawing Library)

In the following section different algorithms supported by our graph drawing and analysis library-GRAD will be shown. Additionally, possible ways of choosing appropriate drawing algorithms based on the properties of graphs will be discussed. It can be noted that GRAD can be used both to transform and existing drawing and to form a completely new one when nothing is known about the positions of the graph's vertices. The later is used in our open-source Kroki tool [21] for laying out imported class diagrams created by other modeling tools.

Like it was already mentioned, GRAD also provides a simple graphical editor which can be used to draw graphs we want to experiment with. This editor was used to create all examples of laid out graphs which will be shown in the upcoming sections.

A. Supported graph drawing algorithms

The most important objective of GRAD is to provide a large number of different graph drawing algorithms, both those which can only be applied if a graph has certain properties (e.g. is planar) and those that can be applied to all graphs with acceptable results. GRAD ports the best algorithms from the JUNG framework, JGraphX and Prefuse, and adds a number of new implementations of various graph drawing algorithms, not offered by any of the mentioned libraries. Summarily, the current version of GRAD includes several tree and force-directed drawing algorithms, a hierarchical, two straight-line, a circular which minimizes the number of edge crossings, symmetric, and a so-called box layout, which places elements in a table-like structure. The last four algorithms are GRAD's original implementations. The box layout positions a predefined number of vertices in one row, before continuing to the next row. Due to its simplicity, it will not be discussed in more detail.

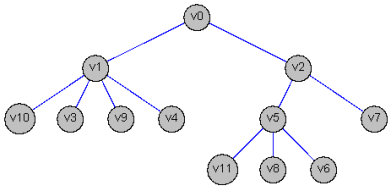


Fig. 1. Resulting drawing of applying the level-based tree drawing algorithm

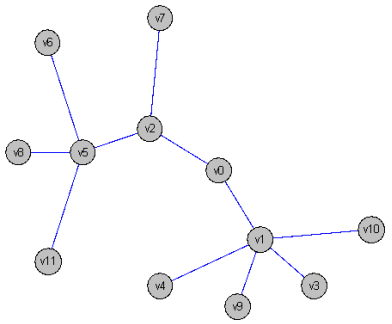


Fig. 2. Resulting drawing of applying the radial tree drawing algorithm

1) Tree and hierarchical drawing algorithms: Tree drawing algorithms included in GRAD consist of a level-based, radial, balloon and compact tree drawing algorithms, ported from the previously mentioned libraries. The best available implementation of a specific algorithm was selected. Fig. 1 shows the result of applying the level-based tree drawing algorithm to lay out a graph, while fig. 2 show the same graph laid out using the radial algorithm.

If a graph is too complex to be laid out using a tree layout and if it is important to emphasize the overall flow, a hierarchical layout can be used. GRAD ports this layout from JGraphX.

2) Force-directed graph drawing algorithms: Similarly to tree drawing algorithms, the force-directed ones were ported from the three mentioned libraries and the best one they had to offer were selected. They include spring, Fruchterman-Reingold, Kamada-Kawai and organic, and fast organic algorithms ported from JGraphX. Since the final results are relatively similar, only one of them will be shown. Fig. 3 show drawing of a graph laid out using the organic force-directed

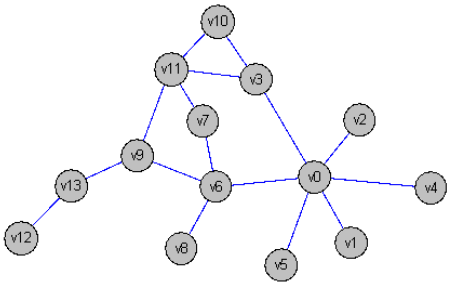


Fig. 3. Resulting drawing of applying organic force-directed drawing algorithm

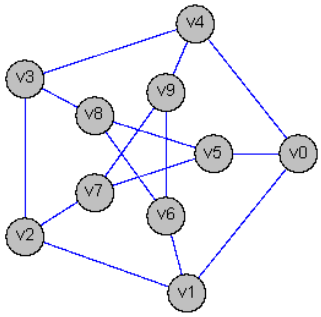


Fig. 4. Resulting drawing of applying the symmetric drawing algorithm

drawing algorithm.

Like it was already mentioned, force-directed algorithms tend to produce satisfactory drawings in most cases. However, the users might be looking for some specific aesthetic criteria, so GRAD doesn't stop here.

3) Symmetric graph drawing: Symmetric drawing algorithms are among the classes of drawing algorithms that the popular graph drawing libraries do not support. GRAD currently offers one such algorithm, with another one being implemented. The available symmetric layout algorithm is based on the work of Carr and Kocay [22], which, given a permutation (automorphism) and a graph, produces a drawing which displays the desired symmetry. The permutations can previously be discovered using an implementation of McKay's canonical graph labeling algorithm [23]. An example of a drawing computed by GRAD's symmetric graph drawing algorithm is shown in fig. 4. This drawing shows a very well-known view of the famous Peterson graph.

4) Straight-line drawings: Straight-line drawings are another class of drawing algorithms which are not provided by the most popular libraries. While not applicable to all graphs (they have to be planar, and in some cases, 2 or 3-connected), they can guarantee a crossing-free drawing.

The first of the provided methods is the one based on Tutte's or barycentric embedding [24]. Given a simple 3-connected planar graph, Tutte's theorem produces a crossing-free straight-line embedding whose outer face is a convex polygon. It is considered to be the first force-directed algorithm at the same time. This method is not difficult to implement, but is only recommended to be used on smaller graphs, with 100 or less vertices.

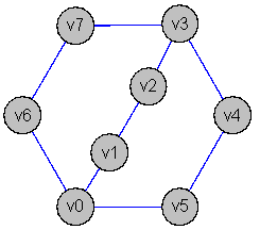


Fig. 5. Drawing of a graph on which Chiba's algorithm was applied

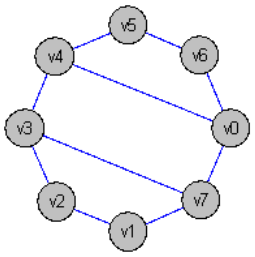


Fig. 6. A circular drawing of a graph

The second straight-line drawing algorithm is a much more complex one. The implementation is based on Chiba's linear algorithm for convex drawing of planar graphs [11], which firstly determines if a graph has a convex drawing and then draws one if the mentioned condition is satisfied. An example of this algorithm's application is shown in fig. 5.

5) Circular drawing: Practically all libraries which deal with graph drawing in some form, provide a circular drawing algorithm. Most of them, however, simply position vertices on a circumference of a circle. In addition to doing so as well, GRAD also implements an algorithm, described in [10], which determines the order in which vertices are placed so that the number of edge-crossings is as small as possible. An example is shown in 6.

It can be noticed that if, for example, vertices v_0 and v_7 switched places, the drawing wouldn't be planar.

Finally, it should be stressed that GRAD allows execution of drawing algorithms which were designed to be applied on simple graphs even if the given one is not simple. Upon execution of the desired algorithm in such case, GRAD detects multiple edges and loops and routes them in order to avoid overlapping of edges and correctly show those which connect one vertex to itself. Furthermore, GRAD provides a very simple way of calling any desired algorithm from a separately developed graphical editor, thoroughly explained in [5].

B. Graph analysis algorithms and choosing the appropriate graph drawing algorithm

Being a graph analysis library as well, GRAD provides a wide array of different algorithms of this sort. They can be used to reveal useful information regarding properties of a graph, which can later influence the decision of which drawing

algorithm to apply in the given situation. Some of these algorithms were of great importance to the implementations of the existing drawing algorithms and can be used to help the implementation of additional ones. Among others, GRAD provides several algorithms for planarity testing, algorithms for splitting of graph into biconnected components based on depth-first search, Hopcroft-Tarjan splitting into triconnected components [25], different algorithms for finding cycles of graphs and previously mentioned McKay's graph labeling algorithm.

By applying appropriate algorithms it can be determined if a graph is, for example, a tree, if it has a planar straight-line drawing or non-trivial automorphisms. Taking advantage of this fact, GRAD provides a way of automatically invoking an algorithm which might be best suited for the graph in question. If a graph is a tree or a forest, a tree drawing algorithm is used. If it is planar, convex drawing is performed. If no special properties are detected, a force-directed layout is applied. Also, if the graph is disjoint, an algorithm is chosen for each component independently and these drawings are later combined to get the final one. An example of this is a class diagram with several disjoint groups of classes where some represent hierarchies, while the other ones do not. This offers users who have no special preferences a simple way to lay out their graphs, even if they don't know anything about graphs and graph drawing.

Furthermore, every somewhat sophisticated drawing algorithm puts emphasis on one or more aesthetic criteria. By naming the criteria, the users can basically choose the algorithm without even knowing its name. In order to accomplish this, a DSL for describing the desired properties of the drawing is currently being developed.

V. Conclusion

This paper explained the need to automatically lay out diagram elements, gave and overview of different classes of graph drawing algorithms and the most popular Java libraries offering some of them. Some difficulties one might encounter when using layout capabilities of the mentioned libraries within a separately developed graphical editor, as well as certain areas of improvement were pointed out. For example, a developer of a DSL which needs to support both textual and graphical syntaxes might run into these issues. They were addressed in our new graph drawing and analysis library called GRAD.

GRAD provides a number of different graph layout algorithms and a quick and easy way of using them to position elements of a diagram in any graphical editor. In addition to porting the best layout algorithms provided by other open-source Java graph drawing libraries, it implements various other ones, not offered by the other mentioned libraries. These include symmetric and two straight-line algorithms, as well as an enhanced version of a circular one. Additionally, GRAD offers ways of automatically choosing appropriate algorithm or their combination to get the best possible result. GRAD is currently being used in our open-source Kroki tool for laying out imported class diagrams created by other modeling tools. These diagrams can contain over 600 classes.

Plans for future improvements of GRAD include:

- implementation of additional drawing algorithms, including a better symmetric and one or more orthogonal ones
- labeling algorithms which address automatic placement of text symbol labels
- a better way of letting user specify desired aesthetic criteria by developing a DSL.

References

[1] M. Mernik, J. Heering, and A. Sloane, "When and how to develop domain-specific languages," *ACM Computing Surveys (CSUR)*, vol. 37, no. 4, pp. 316–344, 2005.

[2] U. Zdun and M. Strembeck, "Architectural decisions for dsl design: Foundational decisions in dsl development," in *Proceedings of 14th European Conference on Pattern Languages of Programs*, Germany, 2009, pp. 1–37.

[3] I.Dejanović, M. Tumbas-Živanov, G. Milosavljević, and B. Perišić, "Comparison of textual and visual notations of dommlite domain-specific language," in *Proceedings of the Advances in Databases and Information Systems*, 2010, pp. 20–24.

[4] I.Dejanović, G. Milosavljević, B. Perišić, and M. Tumbas-Živanov, "Domain-specific language for defining static structure of database applications," *Computer Science and Information Systems*, vol. 7, p. 409–440, 2010. doi: 10.2298/CSIS090203002D

[5] R. Vaderna, G. Milosavljvić, and I. Dejanović, "Graph layout algorithms and libraries: Overview and improvement," in *ICIST 2015 5th International Conference on Information Society and Technology Proceedings*, 2015.

[6] "Graph analysis and drawing library," https://github.com/renatav/GraphDrawing, accessed: 2016-4-4.

[7] M. Patrignani, *Handbook of Graph Drawing and Visualization*. Chapman and Hall/CRC, 2007, ch. 1, pp. 1–42.

[8] H. Purchase, *Computer Graphics and Multimedia: Applications, Problems and Solutions*. Idea Group Publishing, 2004, ch. 8, pp. 110–144.

[9] A. Rusu, *Handbook of Graph Drawing and Visualization*. Chapman and Hall/CRC, 2007, ch. 5, pp. 155–192.

[10] J. Six and I. Tollis, *Handbook of Graph Drawing and Visualization*. Chapman and Hall/CRC, 2007, ch. 9, pp. 155–192.

[11] N. Chiba, T. Yamanouchi, and T. Nishizeki, *Progress in graph theory*. Academic Press, 1984, ch. 5, pp. 153–173.

[12] P. Eades, "A heuristic for graph drawing," *Congressus Numerantium*, vol. 42, p. 149–160, 1984.

[13] T. Kamada and S. Kawai, "An algorithm for drawing general undirected graphs," *Information Processing Letters*, vol. 31, pp. 7–15, April 1989.

[14] T. Fruchterman and E. Reingold, "Graph drawing by force-directed placement," *Software Practice and Experience*, vol. 21, p. 1129 – 1164, November 1991.

[15] "Jung framework," http://jung.sourceforge.net, accessed: 2016-4-4.

[16] B. Meyer, "Self-organizing graphs - a neural network perspective of graph layout," in *In Neural Computers, 393–406, ECKMILLER*. Springer, 1998, pp. 246–262.

[17] "Jgraphx," https://github.com/jgraph/jgraphx, accessed: 2016-4-4.

[18] R. Davidson and D. Harel, "Drawing graphs nicely using simulated annealing," *ACM Transactions on Graphics*, vol. 15, pp. 301–331, 1996.

[19] "Prefuse," http://prefuse.org, accessed: 2016-4-4.

[20] C. Buchheim, M. Juenge, and S. Leipert, "Improving walker's algorithm to run in linear time graph drawing," in *Proceedings of 10th International Graph Drawing Symposium*, Irvine, CA, USA, 2002.

[21] "Kroki mockup tool," http://www.kroki-mde.net, accessed: 2016-4-4.

[22] H. Carr and W. Kocay, "An algorithm for drawing a graph symmetrically," *Bulleting of the Institute of Combinatorics and its Applications*, vol. 27, pp. 19–25, 1997.

[23] B. McKay, "Practical graph isomorphism," in *Proceedings of 10th. Manitoba Conference on Numerical Mathematics and Computing*, 1980, pp. 45–87.

[24] W. Tutte, "How to draw a graph," in *Proceedings of the London Mathematical Society 13*, 1963, p. 743–767.

[25] J. Hopcroft and R. Tarjan, "Dividing a graph into triconnected components," *SIAM J. Computing*, vol. 2, p. 135–158, 1973.

4th Conference on Multimedia, Interaction, Design and Innovation

MIDI Conference provides an interdisciplinary forum for academics, designers and practitioners to discuss the challenges and opportunities for enriching human interaction with digital products and services.

The main focus of MIDI Conference is exploring design methods for creating novel human-system interaction, developing user interfaces and implementing innovations in user-centred development of advanced IT systems and on-line services.

TOPICS

Topics of interest include (but are not limited to) the following areas:

- interactive multimedia and multimodal interaction design
- novel interaction techniques, voice interfaces, interactive multimedia
- ubiquitous, multimodal, pervasive and mobile interaction, wearable computing
- novel information visualization and presentation techniques, Augmented/Virtual Reality
- design methods for usability, accessibility and outstanding user experience
- prototyping of user interfaces and interactive services
- human-centred design practices, methods and tools, user interface design
- unfolding trends in HCI research and practice, customer experience, Service Design
- advances in user-centred interaction design
- understanding people and interactions: theory, concepts, models and methods
- understanding people and interactions: contextual, ethnographical and field studies
- critique and evolution of methods, processes, theories and tools for human-computer interaction
- novel methodologies for conceptualization, design and evaluation of interactive products and services

EVENT CHAIRS

- **Marasek, Krzysztof,** Polish-Japanese Academy of Information Technology, Poland
- **Romanowski, Andrzej,** Lodz University of Technology, Poland
- **Sikorski, Marcin,** Polish-Japanese Academy of Information Technology, Poland

PROGRAM COMMITTEE

- **Christou, Georgios,** European University
- **Fernández Iglesias, Manuel Jose,** Vigo University, Spain
- **Fjeld, Morten,** Chalmers University of Technology
- **Forbrig, Peter,** University of Rostock
- **Guttormsen, Sissel,** University of Bern, Institute of Medical Education
- **Izso, Lajos,** Budapest University of Technology
- **Kaptelinin, Victor,** Umea University
- **Kołakowska, Agata,** Faculty of Electronics, Telecommunications And Informatics, Gdansk University of Technology, Poland
- **Landowska, Agnieszka,** Gdansk University of Technology, Poland
- **Manzke, Robert**
- **Markopoulos, Panos,** Eindhoven University of Technology
- **Marti, Patrizia,** University of Siena, Italy
- **Masoodian, Masood,** University of Waikato
- **Miler, Jakub,** Faculty of Electronics, Telecommunications And Informatics, Gdansk University of Technology, Poland
- **Obaid, Mohammad,** Koç University
- **Pribeanu, Costin,** National Institute for Research and Development in Informatics - ICI Bucuresti
- **Satalecka, Ewa,** Polish-Japanese Academy of Information Technology
- **Slavik, Pavel,** Czech Technical University
- **Stary, Christian,** Kepler University of Linz
- **Szwoch, Mariusz,** Faculty of Electronics, Telecommunications And Informatics, Gdansk University of Technology, Poland
- **Szwoch, Wioleta,** Faculty of Electronics, Telecommunications And Informatics, Gdansk University of Technology, Poland
- **Toro, Carlos,** Vicomtech
- **Vanderdonckt, Jean,** Université catholique de Louvain, Belgium
- **Visciola, Michele,** Experientia
- **Wieczorkowska, Alicja,** Polish-Japanese Academy of Information Technology, Poland
- **Windekilde, Iwona,** Aalborg University
- **Winkler, Marco,** University Paul Sabatier
- **Woźniak, Paweł W.,** Chalmers University of Technology, Sweden
- **Wróbel, Michał,** Gdańsk University of Technology, Poland
- **Ziegler, Juergen,** University of Duisburg-Essen

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1605–1612
DOI: 10.15439/2016F135
ACSIS, Vol. 8. ISSN 2300-5963

Automatically Generated Landmark-enhanced Navigation Instructions for Blind Pedestrians

Jan Balata*, Zdenek Mikovec*, Petr Bures†, Eva Mulickova‡
*Faculty of Electrical Engineering, Czech Technical University in Prague, Czech Republic
†Faculty of Transportation Sciences, Czech Technical University in Prague, Czech Republic
‡Central European Data Agency, a. s. (CEDA), Czech Republic
balatjan@fel.cvut.cz, xmikovec@fel.cvut.cz

***Abstract*—Visual impairment limits a person mainly in ability to move freely and independently. Even with many navigation aids and tools currently on the market, almost one third of the visually impaired do not travel independently without a guide, and human-prepared landmark-enhanced itineraries of the route are the most useful. We designed a system which based on a specific efficiently collected geographical data generates human-like landmark-enhanced navigation instructions. The studies we conducted (quantitative n = 16, qualitative n = 6) proved usability and efficiency of the system. Further we provide set of design recommendations to increase the usability of the system along with specific examples of usage with particular landmarks.**

I. Introduction

THE ability to travel independently is required for satisfactory level of quality of life and self confidence. Visual impairment limits mainly person's mobility and reduces travel-related activities [1]. This can lead to loss of work, friends, and hobbies and eventually to worsening psychical condition of a person. Even though visually impaired people undergo special training of navigation and orientation skills, 30 % of them never leave home alone without sighted guide [2], [3] and this number remains stable for last twenty five years.

The mobility of a person is mainly influenced by the efficiency of the wayfinding process [4]. This process consists of two parts. First, environment sensing such as avoiding obstacles and hazards. Second, the navigation to remote destination. Both parts of the wayfinding process can be supported by navigation aids, which assist the visually impaired pedestrian. The basic criteria for evaluation of navigation aids are safety, efficiency, and stress level, as defined by Armstrong [5].

Currently, this problem is solved by means of car navigation systems, which are not suitable for visually impaired, or better by assistance of orientation and mobility specialists. These specialists can prepare route itinerary to remote destination for visually impaired person for a particular route in advance and provide it in a form of a itinerary. This solution suffers from time requirements (it has to be prepared in advance) and rigidity (there is no option to change the route at user's will).

In our work we focused on developing efficient navigation aid which supports navigation to remote destination. We aimed to generate route itineraries similar to those prepared by orientation and mobility specialist. By using sophisticated data structures and algorithms we addressed the issue of time requirements (the itinerary is created immediately) and rigidity (user can select whichever origin and destination s/he want).

To compare our solution to state-of-art electronic navigation systems we developed two versions of the navigation system. The first version (*Landmark*), with itineraries enhanced by landmarks. The second version (*Metric*), simulating current metric-based navigation systems. For fair comparison of both conditions, *Metric* version also used pedestrian network for routing (sidewalks, crossings). Further, we provide the insights from qualitative study conducted on *Landmark* version.

II. Related Work

A. Pedestrian Navigation

Successful navigation and orientation in a space depends on building of spatial knowledge about the given environment. Siegel and White [6] define three levels of spatial knowledge. These levels are: landmark knowledge, route knowledge, and overview knowledge.

Relations between objects represent an overview knowledge. These relations may be represented by angles or distances between objects not necessarily located or related to the route itself.

The landmarks (representing landmark knowledge) represent the most frequently-used category of navigation cues used by pedestrians [7] (unlike distance, junctions or road type). Ross et al. [8] have shown that inclusion of landmarks within pedestrian route itinerary increased user confidence and reduced navigation errors. Findings of Ross et al. [8] extend also to voice-only navigation [9], where inclusion of landmarks was clearly more prefered by participants.

There are also some experimental designs of navigation systems, which rely primarily on landmarks to navigate users from origin A to destination B, e.g. Millonig and Schechtner [10].

In our system we aim to enhance metric-based navigation instructions with carefully selected landmarks suitable for navigation of visually impaired people.

978-83-60810-90-3/$25.00

B. Orientation and Navigation of Blind

In large spaces visually impaired pedestrians use different cognitive strategies from sighted ones, based on egocentric frames [11], [12]. Visually impaired people have to memorize large amount of information [13] in sequential order [11] while traveling. Fortunately a study by Raz et al. [14] has shown that congenitally blind people are better in both item memory and serial memory than sighted people. Their memory skills are outstanding namely for long sequences of information. Bradley and Dunlop [15] also discovered that blind people were significantly faster with verbal guidance from blind navigator than from sighted one.

Many navigation aids for visually impaired pedestrians have been developed. Some of the aids use special sensors to identify object along the route like cameras [16], or RFID based white canes [17]. Other navigation aids are based on concepts described in [18], and use some kind of position system (e.g. GPS) in combination with geographical information system (GIS) to navigate the pedestrian, e.g. Ariadne GPS, BlindSquare. Navigation aids based on special interaction techniques, e.g. an auditory display [4] or a tactile compass [19], has also been developed.

The navigation systems based on major GIS (e.g. OpenStreetMap) typically suffer from inappropriate level of detail (missing sidewalks, leading lines, slopes of sidewalk), ambiguity (inadequate description of pedestrian crossing so it cannot be located without sight), or they do not optimize routing algorithm to meet specific abilities of visually impaired people (e.g. inability to cross large open spaces).

In cooperation with major national data provider we focus on development of sustainable and scalable GIS with deeply modified data structure (see subsection Geographical Information Database), which allows us to generate specific landmarks for navigation of visually impaired people.

III. Landmark-enhanced Instructions Generation

To implement feasible solution to generate a landmark-enhanced navigation instructions, we need specially modified GIS which is capable of representing special features of the urban environment. Further we need algorithms, which use the landmarks and their parameters to generate route itinerary in a natural language. In our case each route itinerary is composed of navigation instructions for each segment of a route (typically part of a route from corner to corner, corner to crossing, etc.). For each navigation instruction we have chosen the following structure: environment description and action that should be performed by the blind pedestrian.

A. Geographical Information Database

Generally, we distinguish line, point and area features. Line features are tied to large part of pedestrian segment (e.g. geometry representation of a sidewalk in GIS) and represent their properties (like slope, surface quality) or phenomena along the segment (e.g. parking cars, railings). Point features describe phenomena that covers very small part of pedestrian segment (typically less than 3 m), they act as barriers or landmarks (e.g. crossing entry points, narrowings, steps, corners). Area features are landmarks extending over a certain area like traffic noise (busy streets, certain types of public transport, etc.).

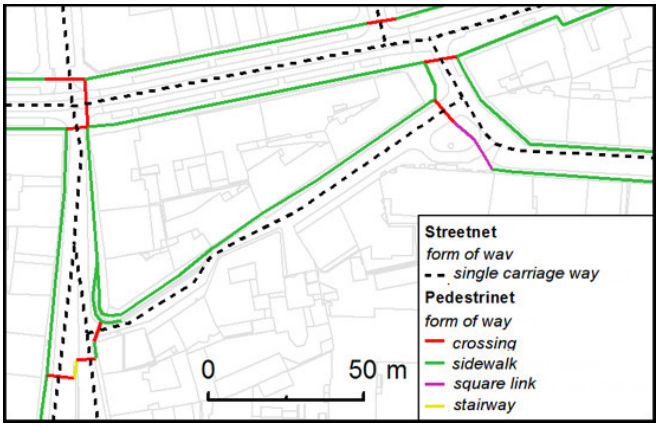


Fig. 1: Comparison of StreetNet and PedestriNet – background data: Digital city map - Map of town utilities (source: Open data).

The nature of those new features and the range of descriptive information reflect the needs of both visually impaired and wheelchair users. The complete data model was proposed in cooperation with Central European Data Agency, a.s. (CEDA), Faculty of Transportation Science and Faculty of Electrical Engineering (Czech Technical University in Prague) within project ROUTE4ALL. Data model was implemented as ArcGIS Geodatabase[1].

Pedestrian segment network. To be able to collect desired landmarks and their properties, the existence of the pedestrian segment network is essential. The national provider of digital vector road geodatabase StreetNet, thus proposed a new product PedestriNet that represents all paths designated for pedestrians. StreetNet is currently used for pedestrian navigation for users without disabilities. For disabled users it is not suitable due to simplified representation of crossings and sidewalks that does not reflect real topology of pedestrian network.

PedestriNet is characterized by very high positional accuracy - it represents footpaths in their real position in the level of Map of town utilities (see Fig 1) and maintain the topology of the pedestrian network. It covers special pedestrian segments like sidewalk, crossing, square link (a sidewalk crossing large open area of a quare) as well as pedestrian segments already represented in StreetNet database (e.g. pedestrian zone, walkway, gallery).

Geodatabase. All features of geodatabase include reference information to PedestriNet - ID of segment, orientation towards it (left/right/on). Thus they can be used for routing along the pedestrian network and generation of route itinerary.

[1] ArcGIS – http://www.arcgis.com/

To minimise data collection costs it is necessary to use existing data as much as possible. City municipalities administer Map of town utilities in high accuracy level that may be used to locate some features (e.g. stairways, corners) or to derive some properties (e.g. width of sidewalk). Unfortunately, these maps are mostly CAD drawings and therefore full automation of data processing is impossible. Further, Land-use map can be used to derive land-use of adjacent footpath area, digital terrain model can be used to calculate slope of footpath, etc. Another source of information may be actual video recordings. Field survey remains necessary to derive quantitative data for wheelchair users (e.g. height of curbs) and to verify collected data. Here, community reports may be of high importance.

Prototype of the geodatabase that is used for validation and testing covers area of about 1 km^2 / 52 km of pedestrian segments. This area took around 5 man/days to collect data and fill the geodatabase.

B. Navigation Instruction Structure

As described above, each navigation instruction is composed of environment description and action that should be performed by the blind pedestrian (similarly to [20]).

The environment description is generated from street names, addresses, corners, and crossings. The action is generated from geometry, street names, corners, slopes, land-use, and point features.

Table I shows how each navigation instruction is composed on a short 5 segment route itinerary with one crossing.

C. Algorithm

First a route is found by Dijkstra algorithm on PedestriNet graph. Then, for each segment, which is represented by vector data, we create a navigation instruction.

For generation of navigation instruction we find best matching sentence templates, which are selected based on a type of a segment, context (available metadata, adjacent segments, direction of the user, etc.), and priority (e.g. the template for corner is more prefered than the template for crossing) (see Listing 1).

An example for sentence template for environment description of a place *"You are at a beveled corner of streets Odboru and Karlovo namesti."* (see Listing 2).

In this way we created landmark-enhanced navigation instruction for blind pedestrians (*Landmark*), which were later compared with metric-based navigation instructions (*Metric*).

Listing 1: Selecting best sentence template.

```
ISituation actualSituation = new StartSituation();
ISituation endSituation = new EndSituation();

while(actualSituation != endSituaton) {
  ITemplate template = FindBestTemplate(actualSituation);
  ISituation actualSituation = template.Apply(actualSituation);
}

function FindBestTemplate(ISituation situation) {
  ITemplate[] availableTemplates = GetAvailableTemples();
  ITemplate bestTemplate = availableTemples
      .Where(template => template.Accepts(situation)
      .OrderByDescending(template => template.Priority)
  return bestTemplate
}
```

Listing 2: Application of a sentence template.

```
class CornerDescriptionTemplate : ITemplate {

  function Accepts(situation) {
    return IsAtCorner(situation) && situation.IsNotApplied(this)
  }

  function Apply(situation) {
    templateText
      = "You are located at a {0} corner of streets {1} and {2}"
    FillSituationVariables(template)
    return SituationWithAppliedTemplate(this)
  }

}
```

IV. Comparative Study

In our experiment we raised the question whether the error rate is lower for *Landmark* condition than for *Metric* condition, and whether measured completion time is lower for *Landmark* condition than for *Metric* condition.

Further we investigated subjective judgement of the participants about the level of safety, comprehension, and ambiguity of the generated itineraries, along with qualitative observations.

A. Participants

Sixteen visually impaired participants (10 female, 6 male) were recruited via e-mail leaflet sent to a group of our long term collaborators of our University. The participants in the experiment were aged from 23 to 66 years ($mean = 35.75$, $SD = 11.23$). Eleven participants had Category 5 visual impairment (no light perception); 5 participants had Category 4 visual impairment (light perception) [21]; 4 participants were late blind, 12 participants were congenitally blind. All of the participants were native Czech speakers.

B. Apparatus

Routes. For our experiment, we selected two routes in city center outdoor environment. Environments for this type of experiment are usually real environments [9], [15] rather than artificial (lab) environments, though exceptions are possible [22]. The location of the routes was in quiet area in the city center of Prague, Czech republic (see Fig. 2). Both of the routes were approximately 350 meters long. Both of the routes consisted of 7 segments and 8 decision points (points where the participant changes his/her direction). They had the same number of turns and pedestrian crossings. Thus we consider the routes to be the same similar to [23].

TABLE I: Main building blocks used for automated generation of route itineraries.

No.	**Environment description**	**Action**						
X/Y	*Corner / Street / Crossing*	*Direction*	*Action*	*Distance approximation*	*Slope*	*Endpoint*	*Landmarks*	*Land-use*
1/5	You are at address Karlovo namesti 293/13.	Turn to the left	and walk	approximately 150 meters	–	to round corner with street Odboru.	–	Keep buildings on your left hand side.
2/5	You are at round corner of streets Karlovo namesti and Odboru.	Continue straight	and cross street Odboru	–	–	to opposite corner	via crossing with light signalization and one-way traffic from right.	–
3/5	You are at beveled corner of streets Karlovo namesti and Odboru.	Turn to the left	and walk	approximately 100 meters	slightly downhill	to corner with street Myslikova	street bends to the right.	Keep buildings on you right hand side.
4/5	You are at corner of streets Odboru and Spalena.	Turn right	and walk	approximately 30 meters	–	to address Myslikova 282/26.	–	Keep buildings on you right hand side.
5/5	You are at destination. You are at address Myslikova 282/26.	–	–	–	–	–	–	–

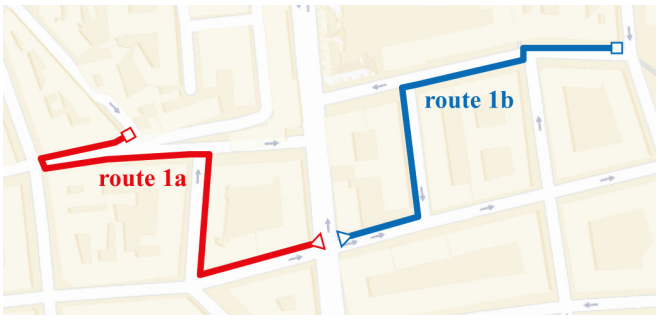


Fig. 2: Routes used in comparative study, the triangle depicts the beginning of a route, the square depicts destination of a route.

Equipment. The participant was equipped with a Nokia 6120 mobile phone with a lanyard which hung from his/her neck. In this way, the phone was protected from being dropped unintentionally, and the participant was able to release it and have an empty hand when needed, and s/he could also find it again quickly. The mobile phone was set to Czech language, and it was equipped with the MobileSpeak text-to-speech (TTS) screen reader application by CodeFactory.

Data collection. In each session, we recorded two video streams of the participant's activities. The first camera (GoPro Hero 3) recorded 1^{st} person view and was installed on a shoulder strap of the backpack that was carried by the participants during the session, while the second camera (Sony DLSR) recorded a 3^{rd} person view by the experimenter shadowing the participant.

C. Procedure

The experiment consisted of two walkthroughs of each route and it lasted around 1.5 hour. In the first walkthrough the experimenter guided the participants to the beginning of the first route, explained the purpose of the experiment to the participants, explained operation of the navigation application, and asked the participant to adjust the phone on a lanyard or to hold it in a hand, according to his/her own preference. The participant was asked to proceed as quickly and accurately as possible. The task was given as follows: *"You have a meeting in Hostel Emma (for route A; Cafe Amandine for route B). To reach the destination use the navigation application. Proceed as if you were alone, but we will be watching for your safety from a distance."* Than the participant started out.

After the first walkthrough the participant was returned to the start of the route and walked the route with the experimenter. The participants were retrospectively asked their about subjective judgement about level of safety ("Did the participant feel safe?"), comprehension ("Did the participant understand what to do?") and ambiguity ("Was the description of the environment ambiguous?") for each segment of a route (Likert scale 1-5 was used).

Then the experimenter took the participant to the start of the second route and proceeded same as on the first route. After the experiment the participant was debriefed and received their payments.

D. Design

The experiment was one factor (two levels) within subject design. The independent variable was itinerary quality (*Landmark*, *Metric*). The itinerary quality and route were balanced using a Latin square. The main measures were an error rate, defined as 0,1 if there was a navigation error (e.g. participant missed the turn, crossed the road on different place, etc.) on a particular segment of a route, and a completion time, calculated as time taken for traveling from start to destination of a route. For analysis of error rate and completion time we use confidence intervals (according to [24]).

E. Results and Discussion

The following subsections describe findings observed during the experiment. We collected the data from 16 sessions and based on the results we propose general design recommendations for creation of navigation instructions.

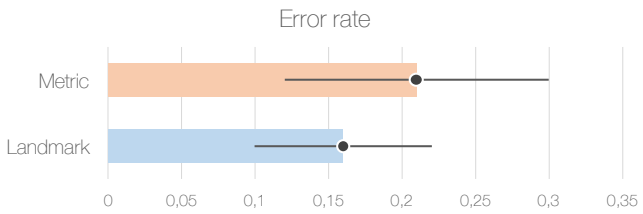


Fig. 3: Mean error rates with 95 % confidence intervals for Landmark and Metric condition (n = 16, lower is better).

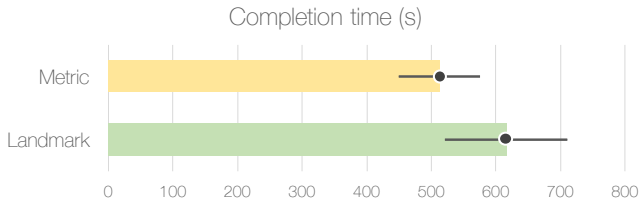


Fig. 4: Mean completion times with 95 % confidence intervals for Landmark and Metric condition (n = 16, lower is better).

Error rate. Fig. 3 provides evidence on error rates and 95% confidence intervals. It seems that error rate for *Landmark* ($mean = 0.16$, 95 % CI $[0.10, 0.22]$) was very similar as for *Metric* ($mean = 0.21$, 95 % CI $[0.12, 0.30]$) and the results are largely inconclusive concerning the difference between the test conditions, although with small favor for *Landmark*.

We could not decide whether the lack of difference in error rate was caused by random variables occuring during the experiment (see paragraph Random variables) or by selection of rather easy routes as described by participants.

Completion time. Fig. 4 provides evidence on completion times in seconds and 95 % confidence intervals. It seem that completion time for *Landmark* ($mean = 615.4$ $seconds$, 95 % CI $[520.89, 709.99]$) is 1.2× higher on average than for *Metric* ($mean = 513.00$ $seconds$, 95 % CI $[450.33,\ 575.67]$) and the results show that there is an effect of the test conditions on completion time. The completion times ranged from 267 to 917 seconds for *Landmark* test condition and it ranged from 292 to 757 seconds for *Metric* test condition.

It cannot be decided whether the difference in completion times for *Landmark* and for *Metric* was caused by longer text in the navigation instructions (participant waited longer time to listen it whole) or by occurrence of random variables (sometimes participants stopped because they didn't feel secure, see paragraph Random variables). It would be necessary to repeat the experiment in more controlled conditions, however it would affect external validity of the experiment.

Subjective judgment. During the second walkthrough we asked the participant about their subjective judgment of each segment of a route. The results suggest that comprehension is higher for *Landmark* (85 % of the participants strongly agree) than for *Metric* (65 % of the participants strongly agree). Ambiguity and safety were evaluated similarly for both test conditions (see Fig. 5).

Further we asked the participant for the comments on navigation instructions for both conditions. In *Landmark* condition, the participants highlighted information about traffic direction or information about land-use. However, in both test conditions they had problems with lack of the corner descriptions and lack of endpoints at crossings. In both test conditions, the participants often confused individual navigation instructions with each other as they started similarly. In *Metric* condition, the participant lacked information about endpoints and they were surprised by very precise distances (precision up to meters; in *Landmark* distances are rounded to tens of meters) reported by the system ("I cannot tell how far did I went precisely", P07, participant P01 laughed about it).

Random variables. Similarly to Rehrl et al. [9] our measurement were influenced by random variables (unpredictable urban environment). Further we classify the problems we observed into 4 categories.

Collision with objects – 28×. We observed that participants frequently collide with traffic signs, poles, beer gardens or parking-ticket machines. This seems as a common problems for the visually impaired pedestrians.

Interference of passerby people – 10×. During the experiment we observed the participant from a distance (shadowing method). Sometimes passerby people stopped the participants and offered them help (4×), however 6× they grabbed, dragged or guided the participant to some arbitrary chosen spot, which resulted in loss of orientation of the participant.

Disruption of the senses – 16×. As a hearing is one of primary orientation and navigation senses for visually impaired pedestrians, its disruption strongly affects the wayfinding process. We observed following sources of hearing disruption: garbage disposal trucks, road cleaning trucks, and rain. Another case was problem with finger sensitivity and operation of our Nokia smartphone.

Stress – 2×. One participant was anxious about the experiment even though we tried to calm him/her during briefing, it affected his/her performance (s/he proceeded much better towards the end of the experiment than in the beginning). Another case was when participant dropped his/her white cane (s/he was immediately assisted and handed the cane).

Allover we counted 32 common problems (collisions and offering a help of passersby people), 17 in *Metric* conditions and 15 in *Landmark*. Next we counted 24 serious problems (grabbing, dragging, guiding by passerby people, disruption of senses, stress), 14 in *Metric* condition, 10 in *Landmark* condition.

F. Recommendations for Design

Following design recommendations for navigation instruction creation were extracted from the findings collected during the experiment:

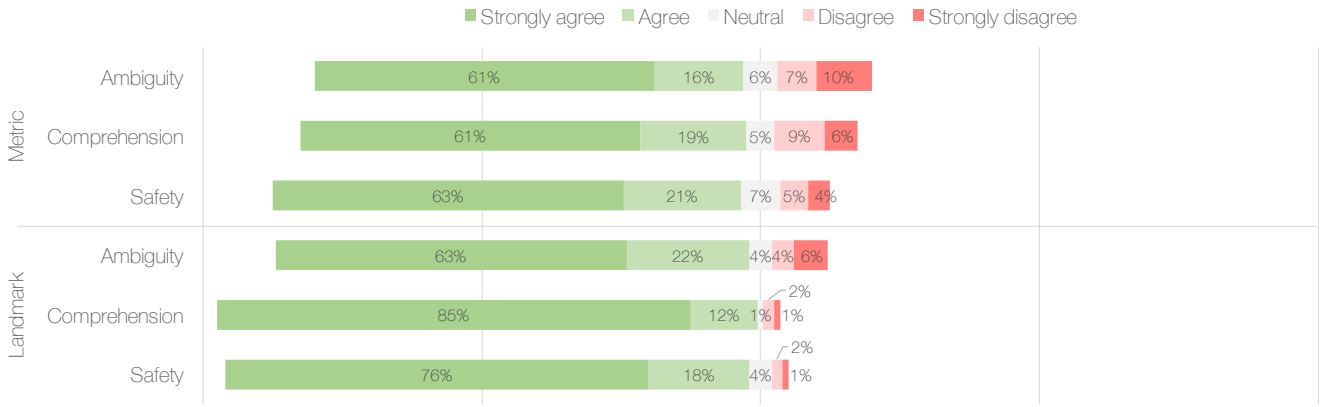


Fig. 5: Subjective judgements about level of safety, comprehension and ambiguity of navigation instructions, summarized over all segments, for both conditions (n = 16).

R1: *The shape of a corner should be added to the itinerary if the shape is different than sharp/plain.*

We observed, that participants were confused at corners, which were not sharp but bevelled or rounded.

R2: *The endpoint should be added for pedestrian crossings like the pavement on the other side of the street, the opposite corner, the opposite side of the street.*

The participants expected the information about the other side of the street when they used a pedestrian crossing.

R3: *Change the beginning of the navigation instructions by adding a sequential number and total number of navigation instruction.*

Some participants were confused when the beginning of the navigation instruction was the same one after another (i.e. crossing the street from one corner to the other corner).

R4: *Changing the naming conventions for pedestrian crossings without "zebra".*

One of the participants expressed concerns about usage of term "unmarked crossing" instead of "a place for crossing" for a place where there are lowered curbs but no zebra drawn on the street.

Others. Some of the participants mentioned that they would benefit from usage of GPS geofencing, which would notify them about next navigation instruction. Other participant found the *Landmark* version too detailed and s/he would prefer *Metric* version on a routes which s/he knew.

V. Qualitative Study

After the first experiment we implemented recommendations R1-R4 into the algorithms and geodatabase of our system. We further investigated automatically generated itineraries in a different, much more complicated urban environment (busy streets, park, passages).

A. Participants

Six visually impaired participants (3 female, 3 male) were recruited via e-mail leaflet sent to a group of our long term collaborators of our University. The participants in the experiment were aged from 30 to 68 years ($mean = 48.67$, $SD = 15.74$). Three participants had Category 5 visual impairment (no light perception; 3 participants had Category 4 visual impairment (light perception); 4 participants were late blind, 2 participants were congenitally blind. All of the participants were native Czech speakers. Originally we recruited 7 participants however participant P02 canceled the appointment.

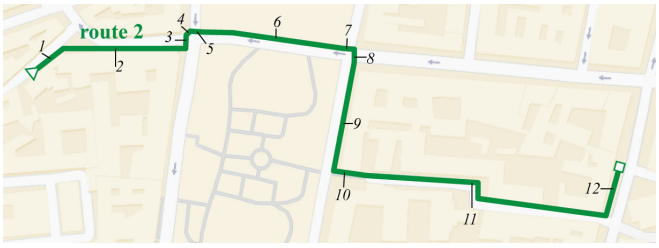


Fig. 6: Route used in qualitative study, the triangle depicts the beginning of a route, the square depicts destination of a route. Numbers represents segments numbers.

B. Apparatus

Route. For our experiment we selected a route in city center outdoor environment. The route went through busy square in a city center of Prague and ended in a quiet area (see Fig. 6). The route was 670 meters long and consisted of 12 segments and 13 decision points. There were 4 pedestrian crossings on a route.

Equipment and **Data collection.** The equipment and the data collection were the same as in the first Comparative Study (see IV).

C. Procedure

The experiment consisted of one walkthrough of the route and the whole session lasted about 45 minutes. At the beginning the experimenter guided the participant to the beginning of the route, explained the purpose of

the experiment and explained operation of the navigation application. The participants were asked to use think-aloud protocol. The task given to the participants was: *"You stand in front of your house on an address Na Zborenci 276/14. To reach the destination use the navigation application. Proceed as if you were alone, but we will be watching you for your safety from a distance, we will also assist you on pedestrian crossings if needed"* Then the participant started out.

When the participant reached the destination s/he was asked out his/her subjective judgment about level of comfort ("Was the navigation instruction comfortable for the participant?"), efficiency ("Did the participant think s/he proceeded efficiently?") and safety ("Did the participant feel safe?") (Likert scale 1-5 was used). The factors of subjective judgement were selected differently from the Comparative Study (see IV) as the subjective evaluation was done per route not per segment. After the experiment the participant was debriefed and received their payments.

D. Results and Discussion

All participants reached the destination successfully without any serious problems.

At the second decision point (bevelled corner) two participants missed the corner and described it as "unclear", P03, P04, however they found out that they missed the corner themselves after few meters.

At the second pedestrian crossing (fifth decision point) all participants turned right to face the pedestrian crossing even though they were not asked to. The next navigation instruction asked the participant to turn right, however this did not confused P07, P03, P01. On the other hand P06 and P05 crossed the pedestrian crossing immediately without selecting next navigation instruction, the did it on the other side of the street and they got confused by instruction to turn right and cross the street. P04 reported this as "weird" but was not confused by it.

At the sixth segment (pavement around greenery) all participants followed a curb on a left side along grass even though they were asked to have street on their right hand. P03 described that s/he used street as acoustic landmark on the right and followed curb on the left hand side.

Five participants found entrance to passage without any problems. Only P06 missed the passage and reported that s/he did not hear it. Participants P06 and P04 missed information about slope in a passage (it was steeply uphill).

The most problematic part of the route was segment 11. The problematic part was "the street is bend twice". Five participants reported that they were not sure they did perceive the second bent. P03 and P07 requested directions of the bends to be present in the itinerary. P01 did not have any problems with the bends.

Surprisingly all participants found destination in the middle of the block of buildings (30 meters from the corner) within 5 meter precision.

Subjective judgement. After the experiment we asked the participant about their subjective judgement of efficiency, comfort and safety during the walkthrough. Fig. 7 shows that 33 % of the participants strongly agreed on comfort, 50 % of the participants strongly agreed on efficiency and 33 % of the participants strongly agreed on safety of the navigation instructions. One participant disagreed on safety due to malfunction of acoustic signalisation on a crossing.

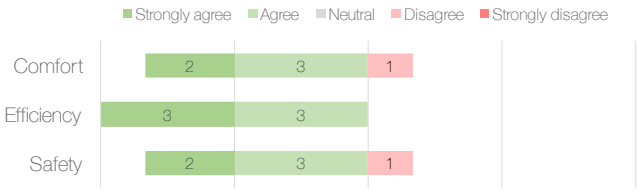


Fig. 7: Subjective judgements about level of safety, efficiency and comfort of navigation instructions (n = 6).

E. Recommendations for Design

Following design recommendations were extracted from the findings collected during the experiment:

R5: *If the corner is less than 90 degrees use "slightly right/left", if the corner is more than 90 degrees use "sharply right/left".*

Some participants did not recognize the corner between first and second segment of a route because the angle was not 90 degrees.

R6: *If the pavement does not follow the building on one side, mention land-use on both sides (i.e. "there is greenery on your left hand side, there is street on your right hand side").*

The participants followed curb along greenery and not the the street. They mentioned they would prefer knowing both sides of the pavement at this segment of the route.

R7: *If there are bends on a street mention them and add directions "first right, second left".*

The most problematic part of the route was where the street bends twice. Many participants did not perceive the second bend of the street.

R8: *Mention pavement slope in passages.*

The participants missed description of a pavement slope in a passage.

VI. Conclusions

We developed a method for automatic generation of landmark-enhanced navigation instruction for blind pedestrians, which is based on modification of GIS data structures and development of algorithm for routing and navigation instruction generation in natural language. We conducted a comparative study of landmark-enhanced navigation instruction (*Landmark*) with metric-based navigation instruction (*Metric*) with 16 visually impaired participants. Although previous studies [8] show that landmark-based navigation is better for pedestrian navigation, the measured results were inconclusive. However subjective evaluation suggested preference of landmark-enhanced condition. Moreover we provide a

set of design recommendations for creation of navigation instructions related mainly to corners and crossings.

Further we investigated improved landmark-enhanced method (*Landmark*) in qualitative study with 6 visually impaired participants resulting in additional recommendations related to leading lines or passages. Subjective evaluation suggested acceptance of the users in levels of safety, effectivity and comfort.

In the future, we will focus on integrating more landmarks such as recessed buildings or traffic sounds.

ACKNOWLEDGMENT

This research has been supported by the project Navigation of handicapped people funded by grant no. SGS16/236/OHK3/3T/13 (FIS 161 – 1611663C000) and by the Technology Agency of the Czech Republic through project Route4all (TA04031574).

REFERENCES

[1] R. G. Golledge, "Geography and the disabled: a survey with special reference to vision impaired and blind populations," *Tran. of the Inst. of British Geographers*, pp. 63–85, 1993. http://dx.doi.org/10.2307/623069

[2] D. Clark-Carter, A. Heyes, and C. Howarth, "The efficiency and walking speed of visually impaired people," *Ergonomics*, vol. 29, no. 6, pp. 779–789, 1986. http://dx.doi.org/10.1080/00140138608968314

[3] R. W. White and P. Grant, "Designing a visible city for visually impaired users," *Proc. of the 2009 Int. Conf. on Inclusive Design*, 2009.

[4] J. M. Loomis, R. G. Golledge, and R. L. Klatzky, "Navigation system for the blind: Auditory display modes and guidance," *Presence: Teleoperators and Virtual Environments*, vol. 7, no. 2, pp. 193–203, 1998. http://dx.doi.org/10.1162/105474698565677

[5] J. Armstrong, "Evaluation of man-machine systems in the mobility of the visually handicapped," *Human factors in health care*, pp. 331–343, 1975.

[6] A. W. Siegel and S. H. White, "The development of spatial representations of large-scale environments." *Adv. in child development and behavior*, vol. 10, p. 9, 1975. http://dx.doi.org/10.1016/S0065-2407(08)60007-5

[7] A. J. May, T. Ross, S. H. Bayer, and M. J. Tarkiainen, "Pedestrian navigation aids: information requirements and design implications," *Personal and Ubiquitous Computing*, vol. 7, no. 6, pp. 331–338, 2003. http://dx.doi.org/10.1007/s00779-003-0248-5

[8] T. Ross, A. May, and S. Thompson, "The use of landmarks in pedestrian navigation instructions and the effects of context," in *MobileHCI 2004*. Springer, 2004, pp. 300–304. http://dx.doi.org/10.1007/978-3-540-28637-0_26

[9] K. Rehrl, E. Häusler, and S. Leitinger, "Comparing the effectiveness of GPS-enhanced voice guidance for pedestrians with metric-and landmark-based instruction sets," in *Geographic information science*. Springer, 2010, pp. 189–203. http://dx.doi.org/10.1007/978-3-642-15300-6_14

[10] A. Millonig and K. Schechtner, "Developing landmark-based pedestrian-navigation systems," *ITSC 2007*, vol. 8, no. 1, pp. 43–49, 2007. http://dx.doi.org/10.1109/TITS.2006.889439

[11] S. Millar, *Understanding and representing space: Theory and evidence from studies with blind and sighted children*. Oxford University/Clarendon Press, 1994.

[12] S. Millar, *Space and sense*. Psychology Press, 2008.

[13] C. Thinus-Blanc and F. Gaunet, "Representation of space in blind persons: vision as a spatial sense?" *Psychological bulletin*, vol. 121, no. 1, p. 20, 1997. http://dx.doi.org/10.1037/0033-2909.121.1.20

[14] N. Raz, E. Striem, G. Pundak, T. Orlov, and E. Zohary, "Superior serial memory in the blind: a case of cognitive compensatory adjustment," *Current Biology*, vol. 17, no. 13, pp. 1129–1133, 2007. http://dx.doi.org/10.1016/j.cub.2007.05.060

[15] N. A. Bradley and M. D. Dunlop, "An experimental investigation into wayfinding directions for visually impaired people," *Personal and Ubiquitous Computing*, vol. 9, no. 6, pp. 395–403, 2005. http://dx.doi.org/10.1007/s00779-005-0350-y

[16] M. Bujacz, P. Baranski, M. Moranski, P. Strumillo, and A. Materka, "Remote guidance for the blind – a proposed teleassistance system and navigation trials," in *HSI 2008*. IEEE, 2008, pp. 888–892. http://dx.doi.org/10.1109/HSI.2008.4581561

[17] J. Faria, S. Lopes, H. Fernandes, P. Martins, and J. Barroso, "Electronic white cane for blind people navigation assistance," in *WAC 2010*. IEEE, 2010, pp. 1–7.

[18] H. Petrie, V. Johnson, T. Strothotte, R. Michel, A. Raab, and L. Reichert, "User-centred design in the development of a navigational aid for blind travellers," in *INTERACT 1997*. Springer, 1997, pp. 220–227. http://dx.doi.org/10.1007/978-0-387-35175-9_39

[19] M. Pielot, B. Poppinga, W. Heuten, and S. Boll, "A tactile compass for eyes-free pedestrian navigation," in *INTERACT 2011*. Springer, 2011, pp. 640–656. http://dx.doi.org/10.1007/978-3-642-23771-3_47

[20] J. Vystrčil, Z. Míkovec, and P. Slavík, "Naviterier-indoor navigation system for visually impaired," 2012.

[21] WHO, "ICD update and revision platform: change the definition of blindness," 2009. http://www.who.int/blindness/Change\%20the\%20Definition\%20of\%20Blindness.pdf

[22] V. R. Schinazi, "Spatial representation and low vision: Two studies on the content, accuracy and utility of mental representations," in *Int. Congress Series*, vol. 1282. Elsevier, 2005, pp. 1063–1067. http://dx.doi.org/10.1016/j.ics.2005.05.163

[23] T. Ishikawa, H. Fujiwara, O. Imai, and A. Okabe, "Wayfinding with a gps-based mobile navigation system: A comparison with maps and direct experience," *Journal of Environmental Psychology*, vol. 28, no. 1, pp. 74–82, 2008. http://dx.doi.org/10.1016/j.jenvp.2007.09.002

[24] P. Dragicevic, "HCI statistics without p-values," 2015.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1613–1619
DOI: 10.15439/2016F288
ACSIS, Vol. 8. ISSN 2300-5963

Design of Crowdsourcing System for Analysis of Gravitational Flow Using X-ray Visualization

Ibrahim Jelliti
International Faculty of Engineering
Lodz University of Technology
Email: 200036@edu.p.lodz.pl

Andrzej Romanowski
Institute of Applied Computer Science
Lodz University of Technology
Email: androm@iis.p.lodz.pl

Krzysztof Grudzien
Institute of Applied Computer Science
Lodz University of Technology
Email: kgrudzi@iis.p.lodz.pl

***Abstract*—The paper describes application of the crowdsourcing system to pre-process X-ray tomography images. In this paper, we show the analysis of the crowdsourcing system applied to process tomography imaging to investigate granular flow with aid of tracking particles. Applying crowdsourcing approach coupled with a proper system interface design enhances the performance of the workers and elevates the attitude. We show here how analysis of the proposed systems in terms of user adoption and performance. Proposed interface features are designed based on previous work evaluation in order to reduce the cognitive and physical workload demands.**

I. Introduction

CROWDSOURCING can be a delegation of work to be done, usually such that can be completed with Internet-connected computers [1]–[2], with an open call for people to contribute on an online task where the crowd here refers to an undefined, but aimed to be large, group of participants. Crowdsourcing systems coordinate those workloads to be completed by large groups of people to solve problems that a single individual could not achieve at the same scale, within the same budget or limited time, or a problem at high level of complexity. Another associated concept often revoked is the crowd intelligence that may result in either reaching the distinctive solutions to problems that are not so easily achievable for a single workers or is sometimes called power of averaging over large numbers of solutions to the same portion of work that leads to a perfect solution with no deviations form the correctly done work. Crowdsourcing system has been recently proved as an effective alternative to solve complex and mundane tasks [3] that can be solved by experts or as an alternative option may be distributed to several non-expert human operators that are willing to do the job or fragments of the job and therefore contribute to a joint solution of the problem. Sub-tasking the problem by the system typically lies to specific work flow depend on the problem, in dependent task, crowd system proves effectiveness in results while most of daily problems are more depend and much complex. Such task requires deep trained knowledge and experience that is difficult to formalize it in algorithms [4]. In crowdsourcing applications, contributors invited to accomplish system tasks using the human intelligence and capabilities [5]. During task submission, crowdsourcing serve as tool to gather information from the crowd intelligence. In the other side, the system relies on individual agent defined as expert serving for study and review the data gathered from the non-experts and processed by the core system. This paper focus on developed crowdsourcing system that contributes non-experts crowd for analysis if industrial tomographic images. The system introduce specific interface and features that work for enhance the crowd yield.

II. Related Work

Crowdsourcing systems raised as a sort of mediation between the way computer systems process the surrounding world's parameters in comparison to how humans sense it. Most of tasks submitted to a widely available crowdsourcing servers (www.mturk.com, www.crowdflower.com, www.crowdmed.com) can possibly be processed by computer systems. Unfortunatelly due to a number of reasons such as difficulty of tasks, uniqueness, complexity, problems with achieving high accuracy or just the economical settings give rise to the need to use different methods for processing these datasets. Taking into account this work, especially image processing for research purposes is widely noted in [13]–[16]. Another interesting example of this type may be referred to a case of possible submission of the own medical results to the medial community around the world or in search for a diagnosis [17]–[19].

In this section, we review different examples of crowdsourcing applications, which demonstrate advantages of crowdsourcing aspects. Reviewing related work allows better to understand the aim of the paper concerning design and extensions of functionality dedicated to industrial process investigations.

A. Crowdmapping

In 2014, NASA launch a space research project 'Be a Martian' [6] based on crowdsourcing approach. The project invite people to mark the craters on mars and size it using bubble-marking interface that compute after user identify a mark. Similarly, The 'The Milki Way project' [7] is a research project that looks through tens of thousands of images from the 'Spitzer' Space Telescope. At the project website, crowd workers mark in specific platform galaxies, star clusters and space objects. The project give volunteers to enter the unknown or unusual things so they identify and size or telling what they see in the infrared data. After gathering

978-83-60810-90-3/$25.00

all this information from crowds, it's easy to the system to classify and mining it.

B. Crowdsourcing in Journalism

Crowdsourcing is used in journalism to find story topics, information and sources. In recent years, a large amount of popular writing and relatively little academic research have been done in the areas of wisdom of the crowdsourcing, co-creation, and networked journalism. In 2009, the Guardian newspaper used crowdsourcing where readers (the crowd) were invited to investigate thousands of political documents that will help the newspaper to parse the data faster by CrowdSourcing. The online newspaper Huffington Post invited in 2009 their readers to compare the original stimulus bill from the US senate with the compromise. The volunteers asked to mark any the differences precisely identified examples of wasteful spending or corporate give ways that aren't stimulative.

C. Crowdfunding

Crowdfunding refers to funding a project or venture by raising monetary contributions from crowds online. Crowdfunding can be used in variety of purposes for instance Philanthropy and civic projects, Real estate crowdfunding, start-up investments, etc. At Kickstarter one of the famous platforms for crowdfunding, people gather funds for projects and ideas on various topics. In February 2012, a project named 'Double Fine Adventure' [8] collected over three million dollars to make a movie, in what was a record-breaking crowdfunded project. In Crowdfunding, individual investments or donated funds typically are small and the funding power is based on the number of participants and whether investments or donated crowd-funding is an ideally way to quickly fund a project by this it can be better than the traditional funding models.

In this work we propose crowdsourcing system is dedicated to image processing of industrial processes and presents progress presented in comparison to system described in the [3]. As data processed in the system constitutes the image sequences, each person has at his disposal the ability to use not only the spatial information (based only on one image), but also temporal information. The presented system provides additional information to the current image, which is taken from previous images. The time relationship presented on the images is an added feature in comparison to previously presented crowdsourcing systems and should increase effectiveness of work.

III. The Crowdsourcing System

The developed crowdsourcing system consists of several elements. Besides the data module, significant position is assigned to user interfaces, from customer (researcher/-expert) and crowd point of view, also additional module dedicates to improve comfort of users work and thus obtain better results. Before describing crowdsourcing system, we present a short characteristic of X-ray measurement system with a deep explanation in interpretation tomography and radiography images to understand better the problem of analysis experimental data.

A. X-ray tomography visualization

The described system and the proposed approach to image processing and analysis is new solution in terms of industrial process investigations as the standard methods not always give sufficient information [20] - [25]. The images of gravitational flow in silo models were obtained with aid of X-ray industrial tomography system [10] - [11]. Data recorded from X-ray system can be presented in form of 3D images (see fig. 1). However for this kind of data visualization is necessary to stop flow process, in analyzed case stopping of gravitational flow of solid, and rotated silo in space between the source and X-ray detector to have the whole set of projections. The 3D image is result of reconstruction procedure based on 2D radiography images (projections). But such approach, for flow phenomena investigation, forces stopping the flow what can be reason of flow characteristic changes - invasion in process. The better option is to apply the continuous record of data. This solution was used during experiments and results were saved in form of 2D radiography images, an example of single raw radiography is presented in figure 2. Quality of this visualization is lower than the case of 3D images (see fig. 1), what causes more complication in image processing and analysis procedure. The quality of radiography images are resulted of absorption of X-ray radiation by all material located on line between source of radiation and a single pixels on X-ray detector. In presented investigation was used 2D flat panel detector. The level of signal detected in a pixel, is due to absorption of X-ray radiation by material structure on path between source and detector pixel. The radiography image provides information about average value of absorption level, which is directly related with material concentration. More details about application of X-ray system to gravitational flow of solid can be find in [10].

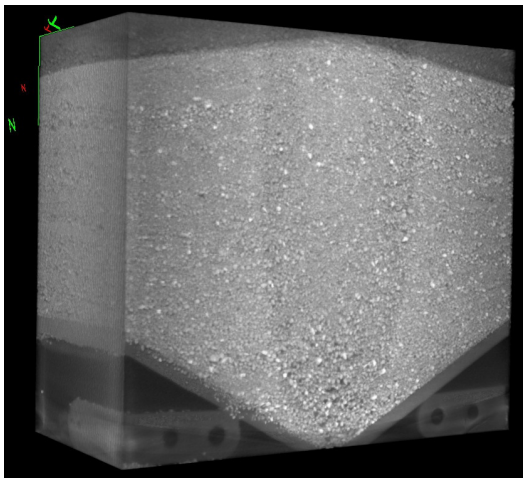


Fig. 1. 3D X-ray tomography of gravitational flow in silo

Fig. 2. RAW X-ray radiography image

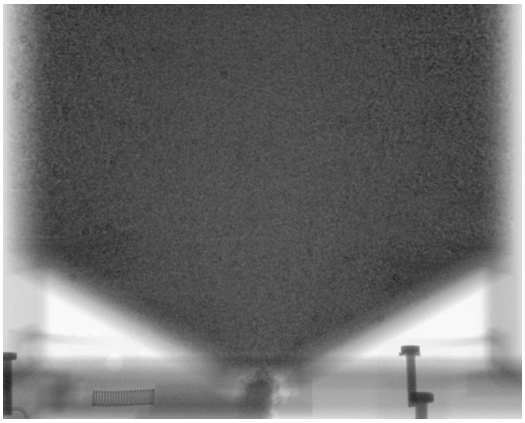


Fig. 3. Radiography image with stopped granular flow

After continuously recorded flow in sequence of 2D radiography images was applied correction of the tomographic images. The result of this processing is presented in figure 3, radiography image for stopped flow, and in figure 4, radiography image for continuously flow. Further improvement based on normalization procedure allows to better visualize center region of silo [11]. Figure 5 presents an example of a normalized radiography image.

As was mentioned the radiography images provide information about the material concentration distribution. On the presented images (see fig. 3-5) it's visible the different area of granular material with different level of concentration. Such flow phenomena is characteristic for funnel flow [9], where in the center of silo material is at lower concentration than at silo wall - stagnant zone appear. The distinguish of these two area is much easier to noticed in 3D tomography (see fig. 2) than in 2D radiography image (see fig. 3-5). Non moving material in silo allows to obtain better radiography image (see fig. 3) than for moving material (see fig. 4). But the influence of breaking flow on process causes different granular behavior than for free flow.

The main task for flow investigation was analyzed dynamic of gravitational flow in funnel area. In order to achieve the intended task the tracking particles, with higher X-ray absorption were added to granular material. The algorithm of image analyzes should provide trajectories of each single particles. It is not easy task to find all tracking particles, What is visible on figures (see. fig. 3-5). The results of automatic image processing algorithm doesn't provides full information about all particles [10]. The development of the crowdsourcing system should allow to find more particles in sequence of radiography images and complete knowledge about flow.

Fig. 4. Radiography image with free granular flow

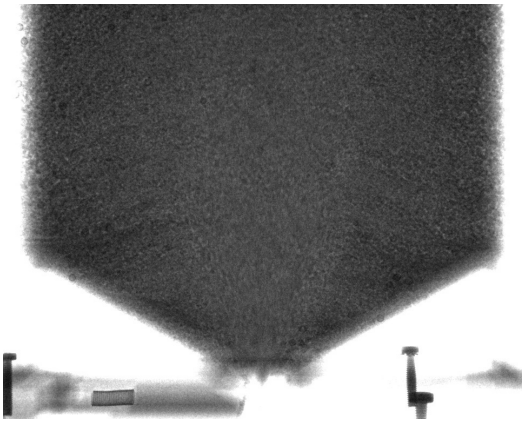


Fig. 5. Radiography image after normalization procedure

B. System workflow

The system is based on user oriented interface. First, the experts identify the task to accomplish and take series of X-ray images of the silos. Then, the obtained data will processed for correction and optimization.

The system workflow runs in three steps, first experts how prepare the x-ray data (in this paper there are radiography images) upload them using the system core into the database. Then the system core can generates tasks from the uploaded data whenever any person of the crowd access to the tasks webform. The system core here select set of data in the database that is not submitted yet by the person, from these

collection, tasks are created in form of 7 to 25 images per task in random way. The creation of task is happen at crowd workers side that means every person of the crowd has his own tasks display. At the third step, the worker chooses a specific task the core system task the task frames into specific interface for analysis of frames. Once worker launch his work submission the system core collect the frames mark data in structured form and save it into the database for later data mining.

C. System interface

Each workers' crowd member can browse the automatically generated task with information about the number of frames like presented on figure 6, then he could choose the one to accomplish. The system redirects the worker to an appropriate interface where all task frames are presented in slide show display together with variety of tools developed to enhance the task performance in time and quality meaning.

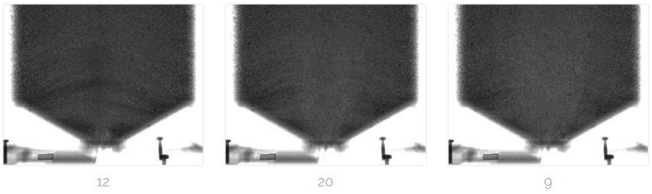


Fig. 6. Tasks interface

D. Task web interface

Once the worker chooses a task to accomplish automatically we be redirected to the another web-interface dedicated to analysis of the tomographic image. The new interface informs the worker at the very begin about the next interface component, features available to help him in doing the task and timing count for his work per frame in background. This interface is presented in figure 7. Such tool panel on the right-hand side is well known in other domains of crowdsourcing applications, but has not been reported for analysis of research tomography images yet.

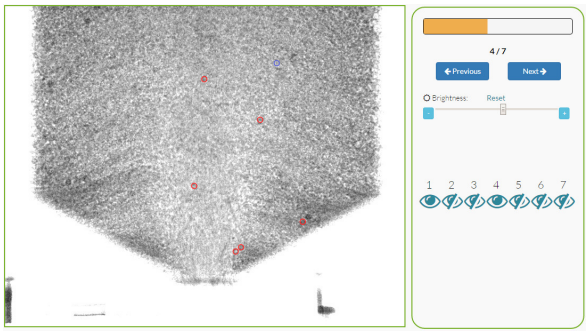


Fig. 7. System interface

E. System Features

1) Interface interactivity: In each frame, worker has to mark with double-click each particle that appears to him in the current frame.

- Double-click on mark will remove it from the frame.
- Single-click on mark will result to be selected.
- Moving the mouse on keep the mouse-key down will let the particle to move according the mouse coordinate in the frame.

2) Frame adjustment: Furthermore, the interface offer on the right-hand panel more features to adjust and ease the job of visual working on the tasks.

First a workload progress bar on the top right shows the progress of the job. Next, a frame caption below shows the index of current frame over the total number of processed frames. The user can browse between the frames easily while doing the task so he can learn from the changes in frames to identify the mark. The interface offer an image processing functions to adjust the current frame. The adjustment of frame brightness help the user to recognize better a particle inside the silo that was not well visible for him in the regularly displayed x-ray frame (as shown on figure 8).

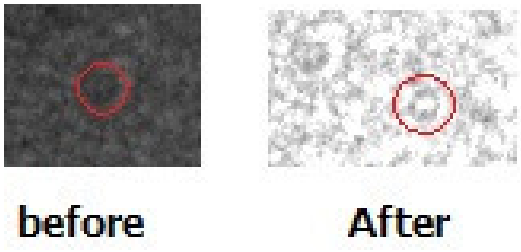


Fig. 8. Brightness adjustment effect on area of the frame.

3) Particles visualization: The system has an 'eye-bar' appear below the brightness control allows the worker to review his particle trace of previous frame on the current frame (reviewing by hide and show the others frame particles). this feature work for increasing the performance of the worker in the frame.

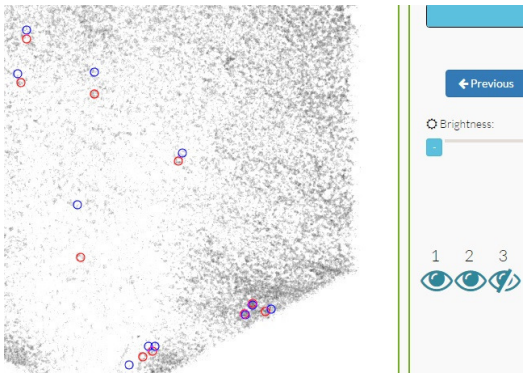


Fig. 9. Worker at frame 2 made mark in red and seeing the marks from previous frame in blue in the same frame.

IV. Pre-study & Results

System was firstly tested and evaluated by experts (2 people with skills in system design, X-ray image analysis, flow process analysis knowledge and previous experience with scientific images crowdsourcing systems). Experts prepared content of the test for the regular users' study. During the study and the evaluation on the system, we conducted tests with 7 non-expert workers completing the tasks. The workers were asked to accomplish task of 41 frames. Chosen frames were a fragment of a time series of X-ray consecutive radiography images and were specially selected to constitute a relatively difficult dataset in terms of marking trace particles. The results are archived by the system after every frame submitted by the workers. Besides the regular study goals, we also inquired workers about their experience in different aspects of using the system and this feedback help us in qualitative evaluation of the system performance.

A. System-side results

Since the main goal of the system is to enable correct identification of the trace particles the proposed design aimed at supporting to achieve this goal. Figure 10 shows an example of results where plots of different colors indicate the number of particles marked by the users for consecutive frames. These results were obtained only with possibility to change the brightness (no temporal hints). One can see a specific spread of counted values, however experts assessed these results as acceptable provided the quantitative analysis (visual inspection) shows that missed particles are not significant regarding the domain analysis. Including the temporal factor into the workflow, i.e. enabling users to use previous frames' results as well as possibility to browse frames back and forth significantly reduced the spread of particles count for different workers. The min difference between expert results and crowd study results decreased from 7% to 1% while max difference decreased from 24% to 9% and in general lead to equalize the results of the crowd, approaching the number expected by experts.

Figure 11 shows the marked particles pointed out by the crowd on the images. The particle easiest to find, located at silo wall, are marked by all workers. The main problem is the discrepancy visible in the funnel area of the flow (central zone of the image). However, when using the temporal options of the interface results tend to the similar homogeneity as for the by-the-walls zones that proves our concept of using additional information in order to ease the process of finding the particles and increasing the accuracy of the system.

In this situation, the system applies basic algorithm to compute real number of particles. In order to do that system considers a particle where most of users marked circles close to anothers' marks by applying an unsupervised clustering at high inner criteria measurements.

B. Human-side results

The other type of analysis we did concerns the crowd workers performance but this time it is associated with the

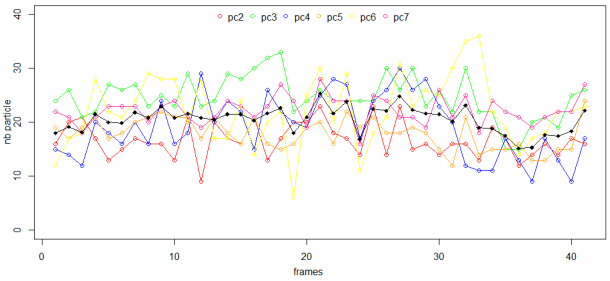


Fig. 10. Plot of numbers of particles marked by users on consecutive images.

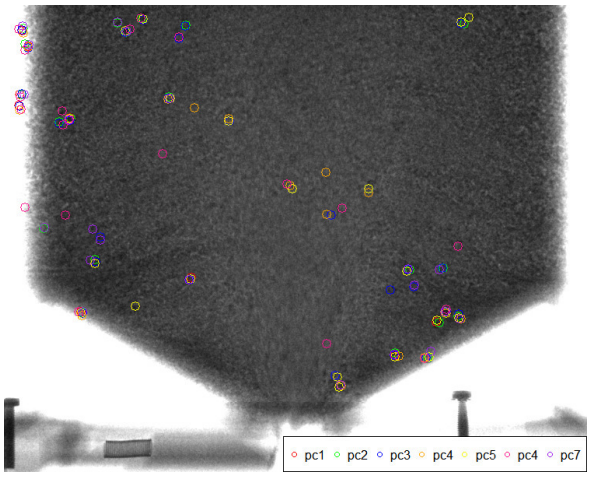


Fig. 11. Image showing particles marked by different users superimposed on a single frame.

other factors of humans' work disregarding the numerical results of particle counted. The analysis of these additional parameters is expected to give insight for further system development. One of the core modules is the time meter recording the period spent by the worker on each frame within the task. It may give some information about the difficulty of the task (or a given frame; especially if the particle count of the different workers for this frame has a large spread of values) but it require a longer study and more participants to derive specific conclusions about that. The system notifies the worker about timer counting per frame before launch the task. The timer feature count time per frame with possibility of browsing the collection back and forth and this enables to measure the performance of the system for the crowd in different frames. Figure 12 shows a heat map of a single experiment where each column represents time performance of a single worker in time from the bottom up to the top for all 41 frames. Yellowish colors indicate more time spent on a single frame while navy blue corresponds to shorter processing of a single frame. After experiments a survey was conducted. Workers answered 8 questions related to task demands and system features. Figure 13 shows a subjective performance in a perceived sense of time users felt themselves in completing the task.

Workers were also asked about the perceived workload in

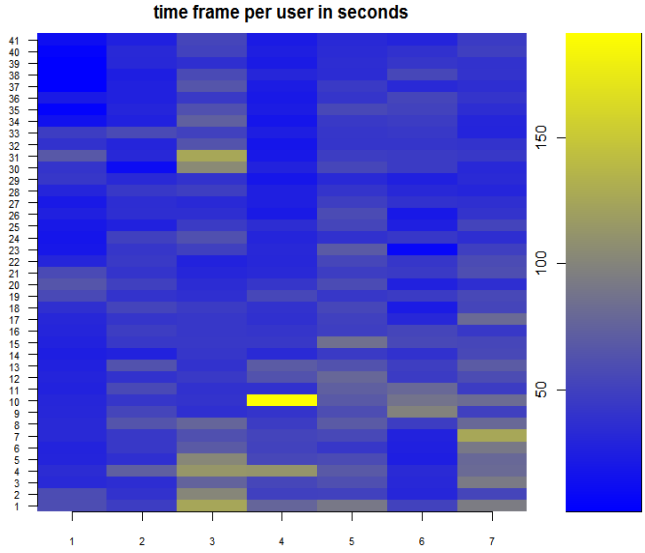


Fig. 12. Work-time per frame for all user

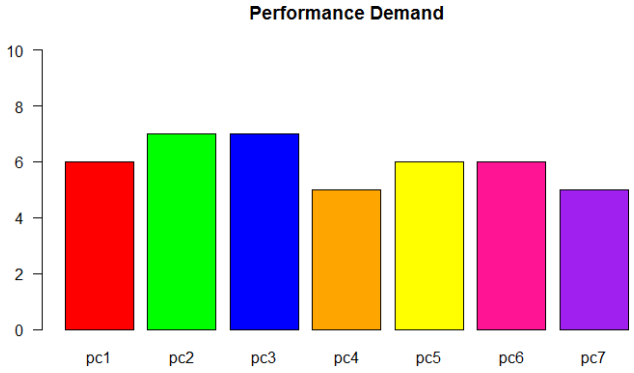


Fig. 13. Crowd workers perceived self-performance.

terms of effort, frustration level, learning, mental, temporal and physical aspects during accomplishing the task. Results are given on figure 14.

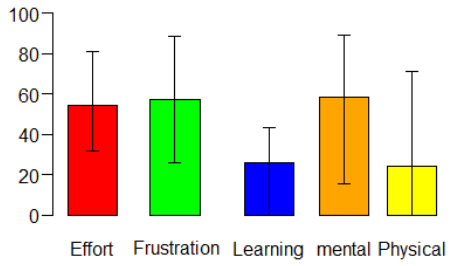


Fig. 14. Results on subjective crowd workers' feedback about perceived workload during the study.

The workers feedback reflect the computation advantage in realizing, the learning level is 20% as average and reach zero even. This indirectly means that the majority of workers do not browse the collection of frames to learn how to identify the particle. This is interesting since the flexibility of the interface was a very demanded feature by the experts. The conclusion for future is to take a deeper look into this aspect. One of the possibility is that the initial training of the users before the experiment/task has to be more detailed.

The level of physical demand is nearly the same this back to the efficacy of the features in the system interface. We have asked the crowd how much the interaction with mouse and brightness control was helpful in accomplishing the frame analysis. The results are promising as presented on figure 15.

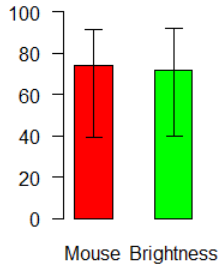


Fig. 15. Crowd workers feedback on system features.

V. Discussion & Future Work

When looking a the results for a system and humans as interpreted in the previous section we noticed a significant level of effort demand against the physical demand beside similar level of frustration. This a general and widely known problem concerning repetitive tasks and can be explained in a way that worker marks the particles in the frame and the same for the rest of collection even with the eye bar feature that reduces at him computing mark again and it may get higher harder if the choose a large collection to work. To enhance the system framework for better feedback results, we work on changing the manner how the work must perform to accomplish a task, so we may aim to decrease the effort by mouse clicking with automatically displaying the previous frame marks to the current and the worker will have just to drag the copy-particle to their new position. However, the previous experience of the experts reveal that this solution may be perceived as distraction for some workers. Therefore, flexibility of the interface in this aspect is required.

In order to cope with high complexity during the clustering of marked particles future work on parallelisation of trajectory calculation with the work of crowdworkers at the same time when he mark the particles may be considered.

Finally, we work on technique that tight the visual search space for the worker. We expect that the implication of this technique in the crowdsourcing system will result in decreasing the effort, mental, physical and temporal demand and at the same time hopefully increase the feasibility and scalability of the crowd system. Heading for these improvements will lead us to semi-automated crowd system that may be comparable with possible automatic algorithms in this complex problem. It seems to be interesting to utilize the eye tracking system to find out how worker use

crowd system [26]. Results will allow to improve the system from worker point of view. Similarly hands gesticulation recognition can be very useful to explore new direction of human computer interaction development in crowdsourcing based system [12].

VI. Conclusion

In this research study we analyzed the design aspects of crowdsourcing system developed for the investigation of gravitational flow with aid of X-ray images. The results are presented on the background of the distinct elements of system design. The additionally introduced features of brightness and especially a feature allowing to use the temporal information, i.e. browsing frames of the current task back and forth allowed to obtain more stable results; less diverse distribution of results delivered by different workers was observed. From the further research on the scientific crowdsourcing systems point of view, the presented heat map is a valuable tool for quantitative estimation of workload per frame since it enables analysis of time spent as well as another valuable information about how often workers used back and forth option. The designed system allows to generate dynamic tasks at worker-side. With the task interface features, we get satisfactory results from the experiments that build on the earlier published work for similar application. The crowd study on the system shows the system feasibility in solving the target problem and improvement in some aspects of the interface performance over the previous design for investigation of industrial flow processes.

VII. Acknowledgments

This work is partially funded by the European Commission under the Erasmus Mundus E-GOV-TN project (Open Government data in Tunisia for service innovation and transparency) -EMA2; Grant Agreement no. 2013-2434/001-001. Authors would like to acknowledge inspiring previous work conducted by all co-authors of [3].

References

[1] Brabham, D. C., "Crowdsourcing as a Model for Problem Solving: an introduction and cases", Convergence: The International Journal of Research into New Media Technologies 14(1), pp. 75–90, 2008.

[2] Howe J., *Crowdsourcing: why the power of the crowd is driving the future of business*, New York: Crown Business, 2008.

[3] Chen C., Wozniak P., Romanowski A., Obaid M., Jaworski T., Kucharski J., Grudzien K., Zhao S., Fjeld M., "Using Crowdsourcing for Scientific Analysis of Industrial Tomographic Images", ACM Transactions on Intelligent Systems and Technology (TIST),Vol. 7, 4, Article 52, 2016, DOI: http://dx.doi.org/10.1145/2897370, 26p.

[4] Stol, K.J. and Fitzgerald, B. Researching, "Crowdsourcing software development: perspectives and concerns", ICSE14 Workshop on Crowdsourcing in Software Engineering, pp. 7-10, 2014.

[5] G. Little, L. Chilton, M. Goldman and R.C. Miller, "Exploring iterative and parallel human computation processes", HCOMP '10 Proceedings of the ACM SIGKDD Workshop on Human Computation, pp. 68-76, 2010.

[6] NASA Citizen Science Lab: beamartian.jpl.nasa.gov/welcome.

[7] ww.milkywayproject.org the Milky Way project news: www.jpl.nasa.gov/news/news.cfm?release=2012-062

[8] www.kickstarter.com/projects/doublefine/double-fine-adventure.

[9] D. Schulze, *Powders and Bulk Solids: Behavior, Characterization, Storage and Flow*, Springer, Verlag, Berlin, GmbH Co. K., 512 p., 2008.

[10] K. Grudzien and M H. de la Torre Gonzalez, "Detection of tracer particles in tomography images for analysis of gravitational flow in silo", Image Processing and Communications, Vol. 18, pp. 11-22, 2013.

[11] L. Babout, K. Grudzien, E. Maire, and PJ. Withers, "Influence of wall roughness and packing density on stagnant zone formation during funnel flow discharge from a silo: An X-ray imaging study", Chemical Engineering Science, 97, pp. 210224, 2013.

[12] Półrola M., Wojciechowski A., "Real-Time Hand Pose Estimation Using Classifiers", Computer Vision and Graphics: International Conference, ICCVG 2012, Warsaw, Springer Berlin Heidelberg, pp. 573-580, 2012.

[13] Irshad H., Montaser-Kouhsari L., Waltz G., Bucur O., Nowak J.A., Dong F., Knoblauch N.W. and Beck A. H., âĂIJCrowdsourcing image annotation for nucleus detection and segmentation in computational pathology: evaluating experts, automated methods, and the crowdâĂİ, Pacific Symposium on Biocomputing, pp.294-305, 2015.

[14] Deepti Ghadiyaram, Alan C. Bovik, "Crowdsourced study of subjective image quality", 48th Asilomar Conference on Signals, Systems and Computers, pp. 84-88, IEEE, 2014

[15] Cabezas F., Carlier A., Salvador A., Giro-i-Nieto X., Charvillat V., "Quality Control in Crowdsourced Object Segmentation", IEEE International Conference on Image Processing (ICIP), Quebec City, Human-Computer Interaction, 2015, DOI: 10.1109/ICIP.2015.7351606.

[16] Flavio Ribeiro, Dinei Florencio ; Vitor Nascimento, "Crowdsourcing subjective image quality evaluation", 2011 18th IEEE International Conference on Image Processing, pp. 3097-3100, 10.1109/ICIP.2011.6116320.

[17] Max H Sims, Maria Fagnano, Jill S Halterman, Marc W Halterman, "Provider impressions of the use of a mobile crowdsourcing app in medical practice", Health Informatics Journal August 28, 2014.

[18] Meyer A.N.D; , Christopher A Longhurst, Hardeep Singh, "Crowdsourcing Diagnosis for Patients With Undiagnosed Illnesses: An Evaluation of CrowdMed", Journal Of Medical Internet Research, vol. 18, No 1, 2016.

[19] Liliya I. Besaleva, CrowdHelp: "A crowdsourcing application for improving disaster management", Global Humanitarian Technology Conference (GHTC), IEEE, pp. 185-190, 2013.

[20] Romanowski, A.; Grudzien, K.; Chaniecki, Z.; Wozniak, P., "Contextual processing of ECT measurement information towards detection of process emergency states", 13th International Conference on Hybrid Intelligent Systems (HIS), TUNISIA, IEEE, pp. 291 - 297, 2013.

[21] Romanowski, A.; Grudzien, K.; Sankowski, D.; Aykroyd, R.G.; Williams R.A., "Advanced statistical computing for capacitance tomography as a monitoring and control tool", 15th International Conference on Intelligent Systems Design and Applications (ISDA 2005), Wroclaw, Poland, pp. 49-54, 2005.

[22] Wajman, R.; Fiderek, P.; Fidos, H.; Nowakowski, J.; Sankowski, D.; Banasiak, R., "Metrological evaluation of 3D electrical capacitance tomography measurement system for two-phase flow fraction determination", Measurement Science and Technology, vol. 24, no. 6, pp. 1-11, 2013.

[23] Mosorow, W., "Flow Pattern Tracing for Mass Flow Rate Measurement in Pneumatic Conveying Using Twin Plane Electrical Capacitance Tomography", Particle and Particle Systems Characterization, vol. 25, no. 3, pp. 259-265, 2008.

[24] Sikora, J., Panczyk, M., Wieleba, P., "Hybrid boundary element method applied for diffusion tomography problems", in: Computer Vision in Robotics and Industrial Applications, (Series in Computer Vision); Eds. Nowakowski J., Sankowski D. - [Hackensack] New Jersey: World Scientific Publishing, pp. 197-229, 2014 .

[25] Kapusta, P., Majchrowicz, M., Sankowski, D., Jackowska-Strumiłło, L., Banasiak, R., "Distributed multi-node, multi-GPU, heterogeneous system for 3D image reconstruction in Electrical Capacitance Tomography - network performance and application analysis, Przegląd Elektrotechniczny, 89 (2 B), 2013, pp. 339-342.

[26] Wojciechowski A., Fornalczyk K., "Single web camera robust interactive eye-gaze tracking method", Bulletin of the Polish Academy of Sciences Technical Sciences, vol. 63, no. 4, pp. 879-886, 2015. DOI: 10.1515/bpasts-2015-0100.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1621–1626
DOI: 10.15439/2016F263
ACSIS, Vol. 8. ISSN 2300-5963

Towards detecting programmers' stress on the basis of keystroke dynamics

Agata Kołakowska
Gdańsk University of Technology
Faculty of Electronics, Telecommunications and Informatics
Email: agatakol@eti.pg.gda.pl

***Abstract*—The article describes the idea of detecting stress among programmers on the basis of keystroke dynamics. An experiment with a group of students of artificial intelligence classes was performed. Two samples of keystroke data were recorded for each case, the first while programming without stress, the second under time pressure. A number of timing and frequency parameters were calculated for each sample. Then statistical analysis was performed to evaluate the significance of keystroke parameters changes. It turned out that some of the defined features might be indicators of being stressed.**

I. Introduction

STRESS is nowadays present in most occupations. IT professionals is one of the groups that is exposed to stress the most. There are many reasons for such situation, e.g. deadlines, pressure from clients, high workload [1]. Some research has been made to show that emotions have impact on software developers' productivity [2]. Moreover, negative emotions such as stress may also influence employers' physical health and their mental state. Therefore it is worth detecting them and, if possible, reacting adequately to alleviate negative effects.

Affective computing, a domain intensively explored in recent times, meets the needs of the above problem. It "relates to, arises from, and influences emotions" [3]. Affective applications implement not only emotion recognition methods, but also interpret and react to the recognized affective states. One of possible areas of applying affective methods is software engineering, where emotion recognition may be used to improve software development process [4].

Various input channels may be considered while designing an emotion recognition tool, i.e. visual [5], depth [6], audio [7], textual [8], physiological [9], standard input devices [10]-[15] or multi-modal input [16]. Not all of them seem practical in any situation, e.g. physiological signals not only require specialized devices and disturb computer users, but these measurements are also disturbed by motions typical for human-computer interaction [17].

Analyzing keystroke dynamics and mouse movements is completely non-intrusive, as it does not require any special hardware and may be invisible for users. The aim of this study is to answer a question whether or not stress caused by time pressure influences programmers' keystroke dynamics.

II. Related work

There is a number of research studies on recognizing emotions on the basis of keystroke dynamics [10]. They deal with a number of problems, i.e. inducing emotions, collecting and labeling data samples, defining and calculating characteristic features and finally training and testing the models. Various solutions are designed to be applied for different emotional states. In some cases a number of emotions are recognized. Other works focus on detecting one selected emotional state.

In [11] for example an experiment on recognizing stress on the basis of keystroke and linguistic features has been presented. The stress was induced by giving some stressful tasks to the participants. Several machine learning techniques were applied to solve that task, i.e. SVM, k-NN, neural networks, decision tress and AdaBoost. It was possible to recognize cognitive and physical stress with accuracies 75% and 62.5% respectively. The authors also showed there was a strong relation between the emotional state and the use of backspace, delete, end, arrow keys and also the time per keystroke and pause length.

Another example of stress recognition was presented in [12]. In this case stress was only one of fifteen emotional states recognized during usual computer activities, e.g. using word processor, sending e-mails. Applying decision trees let recognize some of the emotions with high accuracies 77.4-87.8% (confidence, hesitance, nervousness, relaxation, sadness, tiredness) if the recognition was based on fixed texts. Stress was not one the best recognized emotions.

Detecting stress is an important issue in e-learning systems. A framework for stress detection system applied among Moodle students has been proposed in [13]. The authors of that idea not only chose to analyze keystrokes measured as frequency and intensity of keyboard usage, but

This work was supported in part by Polish-Norwegian Financial Mechanism Small Grant Scheme under the contract no Pol-Nor/209260/108/2015 as well as by DS Funds of ETI Faculty, Gdansk University of Technology.

also information from mouse, webcam, touch screen and accelerometer.

In [14] an intelligent tutoring system, which recognizes boredom and frustration on the basis of keystroke and mouse dynamics, is presented. The data in this study were gathered from students learning a programming language by performing programming tasks and then filling a short questionnaire about the level of the two mentioned emotional states. The obtained accuracies for boredom and frustration were over 83% and 74% respectively.

Another approach to stress detection has been presented in [15] where the possibility of using a pressure-sensitive keyboard and a capacitive mouse to discriminate between stressful and relaxed conditions was investigated. It turned out that under stress the typing pressure increased and more contact with the surface of mouse was observed.

This study focuses on the possibility of detecting stress among programmers by analyzing their keystroke dynamics. The essential assumption of the presented experiment was to perform it in a real-world stressful situation.

III. Experiment design and methodology

A. Research objective

A hypothesis stated in this research is that a programmer's keystroke dynamics change depending on whether or not he or she works under stress caused by time pressure. To verify this idea a proper experiment among programmers should be performed. This study might be treated as a preliminary experiment enabling to reveal all issues necessary to be taken into account before starting to cooperate with a group of programmers in their real life working environment.

B. Participants

Two groups of computer science students have been asked to take part in the experiment in order to collect data. They attended a course on artificial intelligence, where different tasks were solved using Matlab. Each group took part in three sessions performing three different tasks. Only those students, who agreed to participate in the experiment, were collecting the data.

C. Data collection procedure

To collect data coming from keyboard, an application running in background was used [18]. The students started the application themselves at the beginning of a session. The application did not disturb them in any way. Each session was organized in a similar way, as it always happened during that course. The fact, that some keystroke data was gathered, did not cause any changes in the standard way of class conducting, already experienced by the students. During the first part they were supposed to write a piece of code in Matlab script to implement a fragment of an artificial intelligence method. They were given instructions and prompts, both verbal and on the blackboard. They were given clues, when they asked for. Then another task from the same domain, was given to them. This time the students were supposed to solve it individually. They were given specified amount of time for this and they were evaluated at the end. All keystrokes were recorded by the application running in the background. Although the students knew about data recording, they did not know the details of the experiment, especially the stated hypothesis. They only knew the keystrokes would be analyzed later. This was to prevent from intentional change of keystroke behaviors.

16 students took part in the data collection phase, some of them in all three sessions, some in two, and some in one session only. Eventually, 36 data samples were collected, each consisting of two parts. The first part of a sample contained keystroke data from the first part of a session, when stress should not have appeared. The second subsample contained data from the second part of a session, i.e. when the students were working under time pressure.

D. Data preprocessing

The first stage of data processing was segmentation. The whole sequence of keystrokes was split into many shorter sequences depending on the presence of pauses. No one types continually. Every programmer types and stops for a while or a longer time. To identify the limits of typing sequences an idle threshold has been introduced. If the time between depressing a key and pressing the next one exceeded the idle threshold, then the split was made. The greater the value of the threshold the longer keystroke sequences were extracted. All timing characteristics described later in this section were calculated regarding the extracted partial sequences. The segmentation was performed for different values of the idle threshold: 0.3 s, 0.5 s, 0.7 s, 1 s. It lead to creating four sets of data to be analyzed.

E. Feature extraction

After segmenting the data, feature extraction procedure was performed. A number of parameters was calculated from raw data. They may be divided into the following groups: digraph features, trigraph features, special digraph features, frequency features and typing speed.

Digraph and trigraph features are timing characteristics for two-key and three-key sequences. They are all based on parameters commonly used in keystroke dynamics analysis, i.e. dwell time (the time a key is pressed), the time between releasing a key and pressing the next one, the duration of key sequences (the time between pressing the first and depressing the last key in a sequence) and the times between subsequent key presses. Moreover, the number of events for a digraph or trigraph was also calculated. These are the numbers of all key down an key up events in a graph, so it is usually 4 for a digraph and 6 for a trigraph. Sometimes, especially when a user types quickly, it happens that a user presses the next key before depressing one. In such cases additional events may appear between those coming from a graph and then the values for these attributes may differ from 4 or 6. A data

sample contains many digraphs and trigraphs. The parameters were calculated for all of them and then their mean values and standard deviations were saved. The detailed list of digraph and trigraph features is presented in Table I.

TABLE I.
PARAMETERS CALCULATED FROM RAW DATA

Feature subsets	Description (feature identifier)
12 digraph features (mean and standard deviation calculated for each parameter)	dwell time for the first key (di_01, di_02)
	dwell time for the second key (di_03, di_04)
	time between pressing the first and the second key in a digraph (di_05, di_06)
	time between depressing the first and pressing the second key (di_07, di_08)
	digraph duration (time between pressing the first and releasing the second key) (di_09, di_10)
	number of events for a digraph (di_11, di_12)
18 trigraph features (mean and standard deviation calculated for each parameter)	dwell time for the first key (tri_01, tri_02)
	dwell time for the second key (tri_03, tri_04)
	dwell time for the third key (tri_05, tri_06)
	time between pressing the first and the second key in a trigraph (tri_07, tri_08)
	time between pressing the second and the third key in a trigraph (tri_09, tri_10)
	time between depressing the first and pressing the second key (tri_11, tri_12)
	time between depressing the second and pressing the third key (tri_13, tri_14)
	trigraph duration (time between pressing the first and releasing the third key) (tri_15, tri_16)
	number of events for a trigraph (tri_17, tri_18)
10 special digraph features (mean and standard deviation calculated for the timing parameters)	time between pressing the first and the second key in a digraph starting from the left shift (di_L_01, di_L_02)
	duration of digraph starting from the left shift (time between pressing the first and releasing the second key) (di_L_03, di_L_04)
	percentage of times when the left shift starting a digraph is released before releasing the second key (di_L_05)
	time between pressing the first and the second key in a digraph starting from the right shift (di_R_01, di_R_02)
	duration of digraph starting from the right shift (time between pressing the first and releasing the second key) (di_R_03, di_R_04)
	percentage of times when the right shift starting a digraph is released before releasing the second key (di_R_05)
17 frequency features	frequency of using the following keys: enter (freq_ENTER), spacebar (freq_SPACE), tab (freq_TAB), backspace (freq_BCKSPC), delete (freq_DEL), up (freq_UP), down (freq_DOWN), left (freq_LEFT), right (freq_RIGHT), left shift (freq_LSHIFT), right shift (freq_RSHIFT), home (freq_HOME), end (freq_END), pgup (freq_PGUP), pgdn (freq_PGDN), percent (freq_PERC)
	number of capital letters to the total number of letters (freq_CAPS)
typing speed	average number of keystrokes per second (speed)

Some digraphs have been treated as special sequences in the case of this applications. These are digraphs containing either left or right shift key as the first one. Digits, operators, brackets, which require using shift to enter them, are common characters while programming. Therefore some digraph parameters were calculated for digraphs starting from the left and the right shift. The detailed list of these characteristics is presented in Table I.

Another group of features are frequency parameters. In contrast to digraphs and trigraphs they do not describe keystroke rhythm. Some of them may indicate the way users make corrections (backspace, delete), move across the code (pgup, pgdn, home, end, up, down, left, right), take care of programming style issues. For example the percent (%) symbol is used in Matlab to start a comment. The frequency was calculated as the number of a selected symbol to the total number of keystrokes. One of the frequency features was calculated in a different way, i.e. the number of capital letters to the total number of letters. The complete list of frequency features is shown in Table I.

Finally, the typing speed, which indicates the number of keystrokes per second, was calculated.

The total number of parameters calculated was 58, which is rather high. Some of the features are redundant. e.g. digraph duration is the sum of dwell times for the two keystrokes and the time between depressing the first and pressing the second key. However, in this study, this set of features is not going to be used for classification purposes for example. That is why dimension reduction is not priority. The aim of this research is to find out whether some of the parameter changes might be caused by stress. Therefore statistical analysis was performed for each of the proposed 58 features.

F. Statistical apparatus for data analysis

To verify the stated research hypothesis a statistical test should be applied. In this case dependent t-test was used. This is a proper test when the data from a person are gathered several times and their changes are investigated. The advantage of analyzing the changes, not the values themselves, is also the fact that in this way differences between the subjects may not be taken into account. In the case of this experiment the difference between the participants and between keyboard types are hidden by analyzing the values changes only.

The keystroke statistics of a person performing a task are measured twice, i.e. while coding without being evaluated and then while writing under time pressure a piece of code to be evaluated. The question is whether the changes of the values of keystroke parameters are significant. To answer this question the dependent t-test of the following form may be used:

$$t = \frac{\bar{d}}{s_d}\sqrt{n-1}$$

where $\bar{d}$ is the mean difference between the values of two measures obtained in two situations; s_d is the standard deviation of the differences; n is the number of degrees of freedom, i.e. the number of pairs of samples, for which the difference is calculated. Because of the fact that no assumption is made on the direction of the observed changes, i.e. keystroke parameters may either increase or decrease, the applied t-test should be two-tailed.

IV. Analysis of Results

A. Results

The values of t-test were calculated for all defined keystroke parameters and the corresponding levels of significance (p-values) have been presented in Table II. Only the features, for which t-statistic exceeded critical value for p=0.05 have been shown.

The results are divided into four sections obtained for different values of the idle threshold. Moreover, each section (table column) is divided into three parts depending on the observed level of significance of parameters' changes. The three subsections correspond to $p \leq 0.001$, $0.001 < p \leq 0.01$ and $0.01 < p \leq 0.05$ respectively. As it can be seen from Table II, about half of the parameters change significantly. The differences among the results obtained for different values of the idle threshold are not very clear. The number of significantly changed features is lower for the lowest threshold of 0.3 s. It can be noted that the subsets of parameters are quite similar. Most features appear in all four sections (columns). It is possible to indicate the parameters which seem to be the best indicators of keystroke dynamics changes. These parameters are potential candidates to be analyzed in a stress detection system.

TABLE II.
SIGNIFICANCE LEVEL FOR FEATURE CHANGES OBTAINED FOR FOUR SETS OF DATA GENERATED FOR DIFFERENT IDLE THRESHOLD VALUES

idle threshold = 0.3 s		**idle threshold = 0.5 s**		**idle threshold = 0.7 s**		**idle threshold = 1 s**	
Feature	**p-value**	**Feature**	**p-value**	**Feature**	**p-value**	**Feature**	**p-value**
di_09	0.00002	di_09	0.00000	di_09	0.00000	tri_09	0.00001
Speed	0.00029	tri_09	0.00000	tri_09	0.00003	di_07	0.00007
tri_09	0.00067	tri_10	0.00004	tri_13	0.00009	di_09	0.00013
di_07	0.00091	di_07	0.00005	di_07	0.00013	tri_13	0.00013
tri_13	0.00102	tri_12	0.00007	Speed	0.00033	di_03	0.00018
tri_15	0.00133	tri_13	0.00013	tri_03	0.00041	tri_03	0.00022
di_03	0.00140	Speed	0.00030	tri_15	0.00067	tri_15	0.00048
di_L_04	0.00155	tri_15	0.00031	di_03	0.00080	di_L_04	0.00067
tri_03	0.00215	di_03	0.00045	di_01	0.00117	tri_10	0.00086
di_L_01	0.00244	di_01	0.00085	tri_10	0.00140	di_L_01	0.00166
di_R_03	0.00559	di_L_01	0.00419	di_L_01	0.00309	speed	0.00253
tri_10	0.00684	di_12	0.00517	di_L_04	0.00329	di_L_02	0.00477
di_L_02	0.00713	di_L_04	0.00770	freq_BCKSPC	0.00850	tri_12	0.00563
di_R_04	0.00849	di_11	0.00849	di_11	0.00853	di_11	0.00738
freq_BCKSPC	0.00850	freq_BCKSPC	0.00850	freq_RIGHT	0.01230	freq_BCKSPC	0.00850
di_R_01	0.00954	di_10	0.01120	di_12	0.01284	di_12	0.00996
freq_RIGHT	0.01230	tri_03	0.01198	tri_12	0.01471	tri_07	0.01028
di_R_02	0.01236	freq_RIGHT	0.01230	di_R_04	0.01509	freq_RIGHT	0.01230
tri_12	0.01727	tri_07	0.01346	di_L_02	0.01558	freq_LSHIFT	0.01734
freq_LSHIFT	0.01734	tri_18	0.01449	di_R_02	0.01567	tri_16	0.01839
di_01	0.01898	freq_LSHIFT	0.01734	freq_LSHIFT	0.01734	freq_CAPS	0.02016
freq_CAPS	0.02035	di_R_04	0.02000	tri_07	0.01748	tri_01	0.02154
di_05	0.02939	freq_CAPS	0.02013	freq_CAPS	0.02020	di_08	0.02166
freq_ENTER	0.03468	di_08	0.02463	tri_17	0.02610	di_L_03	0.02285
di_12	0.03816	tri_17	0.03151	tri_14	0.02651	tri_18	0.02444
di_R_05	0.04214	di_R_03	0.03301	di_10	0.03150	tri_17	0.02740
tri_04	0.04584	di_R_02	0.03353	tri_16	0.03293	tri_14	0.03079
		freq_ENTER	0.03468	freq_ENTER	0.03468	di_10	0.03454
		tri_14	0.03511	di_08	0.03483	freq_ENTER	0.03468
		di_L_02	0.03945	di_L_03	0.03896	di_R_02	0.03804
		tri_01	0.04220	tri_18	0.03910	di_R_04	0.03905
		di_02	0.04960			di_01	0.04539

TABLE III.
NUMBER OF FEATURES CHANGED SIGNIFICANTLY

p-value	Idle threshold			
	0.3 s	**0.5 s**	**0.7 s**	**1 s**
Task 1				
$p \leq 0.001$	0	0	0	0
$0.001 < p \leq 0.01$	2	2	2	2
$0.01 < p \leq 0.05$	0	3	0	0
Task 2				
$p \leq 0.001$	9	14	10	9
$0.001 < p \leq 0.01$	11	9	7	7
$0.01 < p \leq 0.05$	13	7	11	11
Task 3				
$p \leq 0.001$	3	3	7	6
$0.001 < p \leq 0.01$	4	8	6	8
$0.01 < p \leq 0.05$	7	5	6	5

Most of the top parameters are digraph and trigraph characteristics, e.g. mean digraph duration (di_09), mean time between depressing the first and pressing the second key in a digraph (di_07), mean dwell time for the second key (di_03), mean time between pressing the second and the third key in a trigraph (tri_09), mean dwell time for the second key in a trigraph (tri_03), mean trigraph duration (tri_15). Typing speed also turned out to change significantly. Some parameters calculated for digraphs starting from the left shift, i.e. mean time between pressing the left shift and the subsequent key (di_L_01) and standard deviation for digraph duration (di_L_04). Regarding the frequency features, it may be observed that only five keys seem to be worth taking into account. These are: backspace, right arrow, left arrow, enter. It confirmed some observations made in other studies [11].

The mentioned observations have been made on the basis of data coming from different people. It is possible that analyzing the results individually would let draw different conclusions. First of all the idle threshold could be adjusted regarding one's typing speed. Then the subsets of significantly changed parameters could be also found individually. However, it would require gathering more samples from one person performing different tasks.

B. Limitations

Although significant changes for some of the defined characteristics have been observed, these results should be analyzed with precaution. The experiment was not free from limitations.

The first issue, which should be discussed, is the difference between the results obtained for different tasks. As it has been mentioned in Section III C, the 36 samples were collected while performing three different tasks. The number of samples form the three tasks was 15, 12 and 8 respectively. The significance of feature changes was also estimated for each task independently. Table III contains numbers of features found to change significantly in each case. Only a few parameter changes turned out to be significant in the case of task 1, whereas for other two tasks there were more of them. One of the reasons for this difference is the fact that the level of difficulty of the three tasks was not the same. The first one was the easiest although it required more coding than in the second case. The second task was more difficult but in this case the students were supposed to spend some time on designing before starting coding so the amount of code written was smaller. The third task was the most difficult and it required the highest number of code lines to be written. Moreover, in all three cases the students were precisely instructed during the first part of the lesson. The blackboard was used to explain the details of the problems being solved. In the case of the second and the third task more instructions were given on the blackboard and it was possible to make use of it by copying some of the lines to students' code. The amount of rewriting in the first task was much lower. During the second, i.e. the stressful, part of the lesson, no lines to be copied were given on the blackboard. Although few lines could be written by copying from the blackboard, keystroke dynamics with rewritten fragments might be different than without it.

It should be also noted that some of the features turned out to be useless, e.g. the frequency parameters for the pgup, pgdn keys. It was because of the specificity of the three given tasks, which did not require writing many pages of code. However, in real programming environments, these parameters could be worth calculating. Possible different behaviors are worth paying attention, e.g. either pressing pgdn/pgup quickly many times or keeping it pressed for some time to move down/up.

Another limitation of this study is neglecting the fact that people are not equally prone to stress and they react to stress in different ways. There are some factors such as marital status, age, gender, income, experience, which have been found as having influence on individual stress level [1]. Some of these factors (e.g. age, experience, income) were not present in the case of the presented experiment, due to the peculiar study group of students attending the same class, but some of them still remained.

Finally, it has to be highlighted, that although the stressful situation was induced in a way, there is no certainty that the participants were really stressed, because they were not asked for any self-assessment at the end of each session. The only way to make sure that the task given to the students really induced stress would be applying one of numerous questionnaires which cover a wide range of symptoms induced by stress and are used in the field of psychology [19]. Another interesting approach would be incorporating some physiological measurements, which could be indicators

of stress. However, this would require usage of special devices and coping with the problem of the sensitiveness of some biometric sensors to finger movements [16][17]. Thus it could not be implemented in an experiment performed in a real life situation as the one described.

I. Conclusions

The results of the presented survey may be treated as the preliminary ones, which could be useful in designing a deliberate experiment to be performed among programmers. Regarding the mentioned issues it can be noticed, that more factors should be taken into account to make sure on the influence of stress on programmers' keystroke dynamics. The presented results give some clues: the tasks performed with and without time pressure should be as similar in the sense of difficulty and length, as possible; effort should be made in order to ensure similar working conditions; the idle threshold should be adjusted individually depending on the typing speed. Finally, the experiment results should be also compared to the results of a proper psychological questionnaire.

Moreover, some other ideas could be explored. One of the most interesting ones is adding to the set of analyzed parameters the timing characteristics specially defined for a given programming language. It could be for example key words and also common sequences of symbols instead of calculating all digraph and trigraph parameters.

Another interesting idea is to incorporate information from mouse as well. There are some known studies on recognizing emotions from mouse movements [10]. Such analysis could be adapted to a given programming environment by tracking the way it is operated, e.g. using menus, moving across various windows etc.

The observed changes in keystroke dynamics are also worth investigating in another application, i.e. intelligent tutoring systems. Analyzing keystroke changes could be applied to detect specific situations, that might reduce the efficiency of the learning process.

References

[1] V. Sreecharan and M. Srinivasa Reddy, "A study on individual and interpersonal stress levels among software employees," *Int. Journal of Information Technology & Computer Sciences Perspectives*, vol. 2(4), pp. 711-716, 2013.

[2] M. R. Wróbel, "Emotions in the software development process," in *Proc. 6th International Conference on Human System Interaction*, Gdańsk, 2013, doi: 10.1109/HSI.2013.6577875.

[3] R. W. Picard, "Affective Computing", MIT Press, Cambridge, 1997.

[4] A. Kołakowska, A. Landowska, M. Szwoch, W. Szwoch, M. R. Wróbel, "Emotion Recognition and its Application in Software Engineering", in *Proc. 6th International Conference on Human System Interaction*, Gdańsk, 2013, doi: 10.1007/978-3-319-08491-6_5.

[5] S. V. Ioannou, A. T. Raouzaiou, V. A. Tzouvaras, T. P. Mailis, K. C. Karpouzis, S. D. Kollias, "Emotion recognition through facial expression analysis based on a neurofuzzy network," *Neural Networks*, vol. 18(4), pp. 423-435, 2005, doi: 10.1145/1980022.1980177.

[6] M. Szwoch, P. Pieniążek, "Facial Emotion Recognition Using Depth Data," in *Proc. 8th Int. Conf. Human System Interaction*, pp. 271-277, 2015, doi: 10.1109/HSI.2015.7170679.

[7] B. Schuller, M. Lang, G. Rigoll, "Multimodal emotion recognition in audiovisual communication", in *Proc. IEEE Int. Conference on Multimedia and Expo*, ICME, Lausanne, 2002, doi: 10.1109/ICME.2002.1035889.

[8] A. J. Gill, R. M. French, D. Gergle, J. Oberlander, "Identifying Emotional Characteristics from Short Blog Texts," in *Proc. 30th Annual Conference of the Cognitive Science Society*, 2008, pp. 2237-2242.

[9] W. Szwoch, "Using Physiological Signals for Emotion Recognition," in *Proc. 6th International Conference on Human System Interaction*, Gdańsk, 2013, doi: 10.1109/HSI.2013.6577880.

[10] A. Kołakowska, "A review of emotion recognition methods based on keystroke dynamics and mouse movements," in *Proc. 6th International Conference on Human System Interaction*, Gdańsk, 2013, doi: 10.1109/HSI.2013.6577879.

[11] L. M. Vizer, L. Zhou, A. Sears, "Automated stress detection using keystroke and linguistic features," *Int. Journal of Human-Computer Studies* 67, pp. 870-886, 2009, doi: 10.1016/j.ijhcs.2009.07.005.

[12] C. Epp, M. Lippold, R. L. Mandryk, "Identifying emotional states using keystroke dynamics," in *Proc. Conf. on Human Factors in Computing Systems*, Vancouver, pp 715-724, 2011, doi: 10.1145/1978942.1979046

[13] M. Rodrigues, P. Novais, F. Fdez-Riverola, "An approach to assess stress in e-learning students," in *Proc. 11th European Conf. e-Learning*, pp. 461-467, 2012.

[14] A. Hernandez-Aguila, M. Garcia-Valdez, and A. Mancilla, "Affective States in Software Programming: Classification of individuals based on their Keystroke and Mouse Dynamics," *Research in Computing Science* 87, 2014, pp. 27-34.

[15] J. Hernandez, P. Paredes, A. Roseway, M. Czerwinsky, "Under pressure: sensing stress of computer users," in *Proc. 14th Conf. Human Factors in Computing Systems*, pp. 51-60, 2014, doi: 10.1145/2556288.2557165.

[16] A. Landowska, "Emotion monitor – concept, construction and lessons learned," in *Proc. Federated Conference on Computer Science and Information Systems*, pp. 75-80, 2015, doi: 10.15439/2015F384.

[17] A. Landowska, "Emotion monitoring - verification of physiological characteristics measurement procedures," *Metrology and measurement systems*, vol. 21(4), pp. 381-388, 2014, doi: 10.2478/mms-2014-0049.

[18] T. Mankiewicz, "Emotion recognition based on keystroke dynamics" (in Polish: "Rozpoznawanie emocji na podstawie dynamiki pisania na klawiaturze"), Master's thesis, Gdańsk University of Technology, 2014.

[19] Centre for Studies on Human Stress, "How to measure stress in humans," Fernand-Seguin Research Centre of Louis H. Lafontaine Hospital, Quebec, Canada, 2007.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1627–1630
DOI: 10.15439/2016F491
ACSIS, Vol. 8. ISSN 2300-5963

APEOW: A Personal Persuasive Avatar for Encouraging Breaks in Office Work

Przemysław Kucharski[1], Piotr Łuczak [1], Izabela Perenc[1], Tomasz Jaworski[1], Andrzej Romanowski[1], Mohammad Obaid[2], Paweł W. Woźniak[3]

[1]Lodz University of Technology, Łódź, Poland {pkucharski, pluczak, izarenc, tjawors, androm, }@iis.p.lodz.pl
[2]KUAR, Media and Visual Arts, Koç University, Istanbul, Turkey, mobaid@ku.edu.tr
[3]Chalmers University of Technology, Gothenburg, Sweden, pawelw@chalmers.se

***Abstract*—Proper break taking during office work iss necessary to prevent musculoskeletal disorders and reduce the risk of heart disease. We present APOEW — an avatar for preventing continuous office work without taking breaks. APOEW is a system that uses a personalized robot avatar to encourage proper break behaviour during office work. The avatar signals the need for a break by stooping. The system was designed to be unobtrusive and blend well with the office environment. The avatars are customisable in order to enable users to design their work environment freely. We conducted a user study where we observed developers working in front of their computers next to the avatar. Preliminary results indicate it has no negative impact on the work environment and users are intrigued by the system. Moreover, a survey on attitude to our concept reveals interesting and positive feedback that will help to develop an APOEW system further.**

I. Introduction

THE SEDENTARY nature of office work is universally recognised as detrimental to health [4]. Employees are advised to take regular breaks as this results in benefits to cardiovascular health. Yet, workers often forget about the need for break for a variety of reasons. As digital artefacts have already pervaded the work environment and users often have multiple devices present on their desks, we propose introducing an additional object that would remind workers of taking regular breaks. APOEW (pronounced ape-oh) is a humanoid physical avatar robot that alters its posture to suggest a break is needed. We designed an avatar which can easily integrate with the work environment and attempt at persuading the user to take a break without disrupting the work. Being a humanoid robot, it visualises possible negative changes to the user's posture if a proper break regime is not followed. In the remainder of this paper, we present the past work that motivated the creation of APOEW and describe the design process of the avatar. We then provide details about the implementation of the system and report on a preliminary user study. This paper concludes with a discussion of future directions for the design of tools that support proper posture and break taking.

This work-in-progress paper contributes: (1) the design and implementation of a persuasive physical avatar aimed at promoting a proper break regime in office work and (2) initial design insights for future work on personal avatars in office spaces.

II. Related work

Personal avatars (both virtual and physical) have been used for a variety of purposes such as remote collaboration [10] or instant messaging [11]. Past research indicates that avatars can have a persuasive effect and it has been observed in virtual worlds [6]. Our work explores how physical avatars may influence users and promote a particular behaviour. We also build on the fuzzy avatar [3] concept (i.e. an avatar that does not provide direct messages, but merely offers hints) and thus explore ambiguity as a factor in persuasion.

Design interventions in workspaces have a long history in Human-Computer Interaction (HCI). Our work is particularly concerned with systems that attempt to change user behaviour in the work environment and measure office activity. Miro [1] informed office workers on the overall emotional climate in the building. The Clouds [12] attempted at persuading users to use the stairs instead of the lift. These works inspire our research as they prove that introducing new artefacts to office spaces can benefit the communities of office workers.

As we interpret APOEW as a system that addresses the domain of designing persuasive technology [2], our work is highly influenced by past achievements in this domain. Nakajima et al. [9] used persuasive technology successfully to promote healthy habits such as brushing teeth. Hong et al. [7] addressed posture issues while working with computers through a flower-shaped avatar, which provided real-time feedback. A similar approach can be found in [14], where an abstract shape of avatar is designed to reflect human behaviour. Morris et al. [8] investigated how proper break behaviour may be elicited by encouraging hands-free interactions during breaks. Haller et al. [5] used a sensor chair to provide posture feedback through digital and physical means. Also [13] addressed the problem of sedentary lifestyle, but with the use of an ambient display. APOEW is interestingly different from the above past work as it uses a partly ambiguous physical avatar and its feedback is based on cumulative data.

III. Design

Our design work was influenced by personal avatars described in past work as well as previous deployments of interactive technologies in office spaces. We imagined a scenario where an office worker could place a humanoid robot on their desk. The avatar would be personalisable in order

to make it blend well with the rest of the office space and add a personal touch to the desk. In spaces where desks are assigned dynamically, it could function as an easy way to make the space cosier. Continuous, accumulated avoiding of the necessary breaks would result in APOEW stooping. It would also use audio output in extreme cases when a break is really needed. While audio output is bound to cause a distraction, we reckoned that actively refusing to take break must be prevented using more direct means. This action would serve as a direct persuasive factor to suggest a break is much needed. Fig.1 presents the design of APOEW in the form of design sketches.

Our system also explores ambiguity as resource for design. We aimed for APOEW to provide persuasive cues that are not obvious and do not simply impose particular behaviours on the user. We decided to use a humanoid robot as we believed it would make the user reflect on the long-term effects of not taking regular breaks during office work. Customisable clothing would enable the user to make the desk space more personal. We aimed for making it possible for the user to have the avatar represent them. This would potentially trigger additional reflection potential. Furthermore, we enabled the user to notify the system of actually taking a break by pressing a button on the back of the robot. As the robot was designed to stand next to the computer screen, this would require some movement, perhaps prompting a break, even if the user intended to cheat the system. As our design goal was reflection and realising the importance of break taking, we opted for selfmonitoring instead of an elaborate activity sensing system.

IV. Implementation

APOEW was implemented iteratively and two major research prototypes were studied so far. First one was prototyped with the Lego Mindstorms NXT robotics set. There were two robots programmed using Java for the initial user study of the system. This prototype employed a single NXT controller, ´intelligent bricks´, based on 32-bit Atmel AT91SAM7S256 main microcontroller (with 256 KB flash memory and 64 KB of RAM). The robot was additionally equipped with a custom 'backpack' with the RaspberryPi and dedicated speakers, so that the robot could play sounds of higher quality than those provided by the NXT platform and produce synthesised speech. Three motors and two touch sensors were used to move the humanoid body, allow it to stoop and facilitate user input. Custom-made clothing was made for the robot using paper and cloth. Fig. 2 depicts the constructed and programmed humanoid robot next to a programmer at work.

A schedule for changing the avatars posture according to continuous work time was implemented. The robot would move slightly after 50 minutes without a break to signal that a work cycle was about to end. A more significant movement informed about a full cycle ending, after 55 minutes of work. After 65 minutes, the posture of the robot changed significantly and, after 75 minutes, sounds were produced. These effects were even stronger if the worker was skipping multiple breaks during a day.

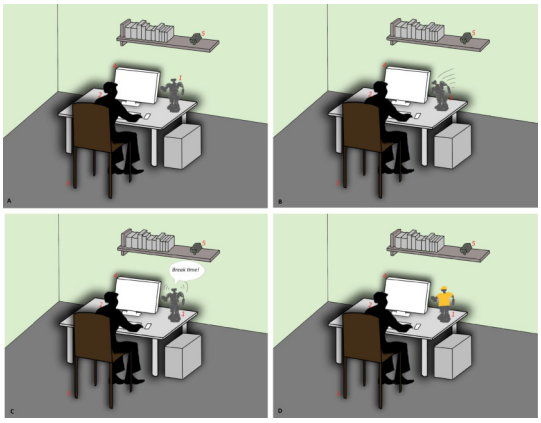


Fig. 1. Conceptual sketches for APOEW: (a) A personal avatar on an office desk. (b) APOEW suggests that a break is need by adjusting its posture. (c) A notification sound is played when the user excessively refuses to take a break. (d) A personalised version of APOEW featuring custom clothing.

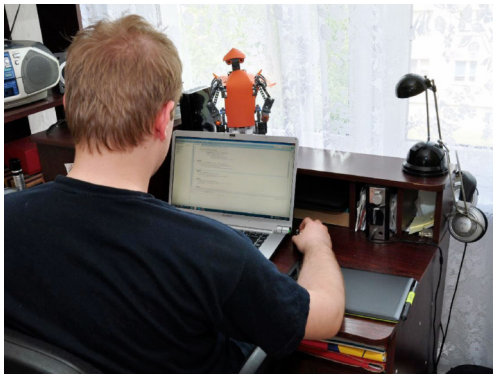


Fig. 2. The first implementation of APOEW. Note the custom clothing of the robot. APOEW is placed next to the computer screen so that interacting with the robot requires movement.

V. Initial evaluation

We conducted a preliminary user study to assess the effectiveness of APOEW, verify the design assumptions and investigate the user experience created by the avatar.

APOEW was deployed in a controlled office environment. Four office workers (two developers and two graphic designers) from a local IT company were recruited for the study that lasted for three days. As work behaviour data was being gathered, a detailed privacy policy was presented and explained to the participants. We monitored two workers at a time using a video camera. A total of 96 hours of video was recorded (4 workers x 3 days x 8 hours). The first day of the study was used as a sample workday for comparison purposes. APOEW was introduced to the office space on day two. After the conclusion of the study, we performed semistructured interviews with the participants.

Three researchers performed independent qualitative analysis of the video and interview material. A joint discussion session was then held and observations were made. The introduction of the avatars to the office space drew attention

Fig. 3. An in-situ picture of the preliminary user study. Two IT professionals are performing their regular work with two APOEW (circled in red) units standing next to their screens. Four users spent a total of eight days working with APOEW.

Fig. 4. The implementation of chair sensor during a second study - a Bioloid robot during the study on the left-hand side. Prototype sensors on the chair on the right-hand side.

of the participants. They were intrigued by APOEW and immediately started discussing its purpose. The participants quickly proceeded to personalize the robots on their desks, by adding features related to hobbies or favourite movies. The extra attention generated by APOEW suggests that our future work should concentrate on a long-term study of the system where the novelty effect of APOEW will be minimised. This is why this work-in-progress paper does not report quantitative data on the break taking behaviour of the participants. We also noted that the implementation of the prototype needs improvements — the noises generated by the servomotors may cause unintended distractions.

The next study was conducted using Bioloid robots and a prototype of a fully functional, yet not final, version of the system, i.e. we also prepared a set of sensors to be roughly mounted on top of chairs - on-top both of the seat and back of the chair. Such a design was to provide the automatic control of the system since the sensors were supposed to communicate the fact of the chair being released by the user and for how long, as well to trigger the APEOW procedure for persuasion if the chair is eventually not empty for a least 5 minutes. However, due to communication problems we were not able to conduct a proper study at that time. Fig. 4 shows the robot during the study with a Bioloid design (left-hand side of the picture) and a sensors prototype shown on the top: seat and back of an office chair. 2 Workers were monitored by 2 days each that gives in total 32 hours of this second design test.

In the interviews, users expressed their satisfaction at the

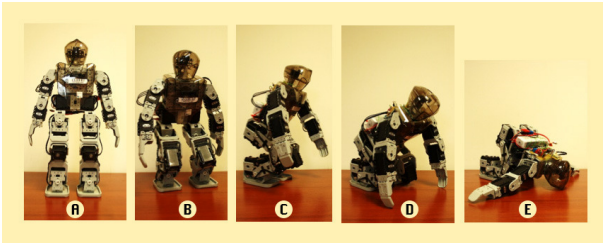


Fig. 5. The changes in the Avatar posture during the planned period.

fact the work environment became more attractive with the introduction of APOEW:

I really like this little fella. His bright clothing makes the office cosier.

The participants also wondered what would be the long-term effects of the system and suggested alternative uses or even peer pressure:

I'd like to see what happens when he stays here longer. I could also use this to see how in fact my friend is working or even pressure him to take a break through my avatar. And, I can dress him as Iron Man.

Overall, we can conclude that APOEW can potentially be integrated into an office environment and users are willing to consider the possibility of introducing personal avatars to their workspace. Privacy is a concern and we will aim to address that issue in future work. Further studies are required to measure the persuasive effect of APOEW.

In order to conduct further development of the project a survey was conducted to gain broader knowledge about the attitude towards the proposed concept. Fig. 5 shows the sequence of photos illustrating the consecutive phases of the APEOW persuasive movement.

N=198 people responded to our survey including 41.4% women and 58.6% men. 47% working, 27.3% working and studying, 23.2% studying and 2.5% unemployed or retired. 47.2% of workers are employed in big companies related to IT business (1000+ employees). Results of the survey revealed that 59.6% declare that they do or try to do periodic breaks from sitting work while most of people (between 68.9% to 88.6% depending on the issue inquired) shown great concern about possible negative results of refraining from conducting these breaks. Feedback shown that majority of respondents express their positive attitude towards the concept as the following data show: Idea assessment: 58% positive vs. 21.2% negative and 20.8% neutral. Anticipated efficiency: 41.4% positive vs 26.8% negative and 31.8% neutral. Possible interfering with the work expected: 31.8% confirm vs. 43.4% not agree and 23.8% neutral. Expected deterioration of the work conducted: 16.6% confirm vs. 67.2% not agree and 16.2% neutral. As one can see from these numbers the APEOW concept was warmly welcomed by the respondents with more then half supporting the main idea and only nearly one fifth being explicitly against it. One third of users see potential interference with the work conducted positive generally but at

the same time only about 17% foresee substantial disruption of the work and more than 67% not anticipating any significant issues with it. Another interesting results show possible ideas for how the proposed APEOW concept could work: Conviction of possibility to modify the look of the APEOW may strengthen its impact: Agree 36.4% vs 33.9% disagree and 29.7% not sure. Willing to personalize own APEOW: 60.6% by look: 38% using sound/voice: 27.2% adjusting the work-break intervals: 77.8% other changes in persuasive behaviour procedure, etc.: 61%. On the other hand, respondents think the following features would diminish the efficiency of the APEOW: personalisation: 11%, sound/voice emission: 73.2%.

Lastly, users answered questions related to anticipated adoption of the APEOW in the workplace and the results show that most of employers are expected to accept the proposed concept in their offices. 51% respondents declared it explicitly, 29.3% said they would probably accept it vs. 5% disapprove at all and 13.1% rather disapprove.

VI. Conclusions and future work

This paper presented the design, implementation and initial evaluation of APOEW; a persuasive physical personal avatar that helps uses maintain a proper work break regime. The avatar was designed to integrate well with an office environment and offer personalisation features. We explored how a humanoid robot can trigger personal reflection in users. An initial user study with four participants working for three days showed that the robot gained that attention of the users and may have had impact on the workers work regime. The results also indicate that feedback forms require further investigation as privacy concerns were raised.

In the near future, we aim to improve the prototype by conducting further studies. We also aim to develop new interactions with the environment, e.g. the robot could make use of user's eye movements [15] or hand movements [16]. First, we aim to compare different types of posture, personalisation options and other output parameters and their effectiveness on altering break regimes. Next, we plan to run a long-term in-the-wild study that will verify the effectiveness of our approach.

In the larger picture, given the rapid advancement of robotics, we believe that HCI should address the role personal physical avatars can play in future everyday spaces. Our work explores only a fragment of what personal avatars may offer. We wonder how systems similar to APOEW can be used not only as persuasive technology, but also whether they can improve workflows, communication or social relationships. We hope that our work will inspire further inquiries into how physical avatars can be integrated in the computing landscape.

VII. Acknowledgments

Authors of this paper would like to express their deepest appreciation for Tomasz Kosiński, who has greatly contributed to the work on this project.

References

[1] Kirsten Boehner, Rogerio DePaula, Paul Dourish, and Phoebe Sengers. 2005. Affect: from information to interaction. *Proceedings of the 4th decennial conference on Critical computing between sense and sensibility CC'05*, ACM Press, 59. http://doi.org/10.1145/1094562.1094570

[2] Bj Fogg. 2009. A behavior model for persuasive design. *Proceedings of the 4th International Conference on Persuasive Technology - Persuasive ´09: 1.* http://doi.org/10.1145/1541948.1541999

[3] William W. Gaver, Jacob Beaver, and Steve Benford. 2003. Ambiguity as a resource for design. *Proceedings of the conference on Human factors in computing systems - CHI ´03*, ACM Press, 233. http://doi.org/10.1145/642611.642653

[4] Fred Gerr, Michele Marcus, and Carolyn Monteilh. 2004. Epidemiology of musculoskeletal disorders among computer users: Lesson learned from the role of posture and keyboard use. it Journal of Electromyography and Kinesiology 14: 25-31. http://doi.org/10.1016/j.jelekin.2003.09.014

[5] Michael Haller, Christoph Richter, Peter Brandl, et al. 2011. Finding the right way for interrupting people improving their sitting posture. it Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 1-17. http://doi.org/10.1007/978-3-642-23771-3_1

[6] Austen L. Hayes, Amy C. Ulinski, and Larry F. Hodges. 2010. That avatar is looking at me! social inhibition in virtual worlds. it IVAâĂŹ10 Proceedings of the 10th international conference on Intelligent virtual agents, Springer-Verlag, 454-467.

[7] Jeong-ki Hong, Sunghyun Song, Jundong Cho, and Andrea Bianchi. 2015. Better Posture Awareness through Flower-Shaped Ambient Avatar. it Proceedings of the Ninth International Conference on Tangible, Embedded, and Embodied Interaction - TEI ´14, ACM Press, 337-340. http://doi.org/10.1145/2677199.2680575

[8] Dan Morris, A.J. Bernheim Brush, and Brian R. Meyers. 2008. SuperBreak: using interactivity to enhance ergonomic typing breaks. it Proceeding of the twenty-sixth annual CHI conference on Human factors in computing systems - CHI ´08, ACM Press, 1817. http://doi.org/10.1145/1357054.1357337

[9] Tatsuo Nakajima, Vili Lehdonvirta, Eiji Tokunaga, and Hiroaki Kimura. 2008. Reflecting human behavior to motivate desirable lifestyle. it Proceedings of the 7th ACM conference on Designing interactive systems - DIS ´08, ACM Press, 405-414. http://doi.org/10.1145/1394445.1394489

[10] Oyewole Oyekoya, William Steptoe, and Anthony Steed. 2012. SphereAvatar: a situated display to represent a remote collaborator. it Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI ´12, ACM Press, 2551. http://doi.org/10.1145/2207676.2208642

[11] Nadya Peek, David Pitman, and Richard The. 2009. Hangsters: tangible peripheral interactive avatars for instant messaging. it Proceedings of the 3rd International Conference on Tangible and Embedded Interaction - TEI ´09, ACM Press, 25. http://doi.org/10.1145/1517664.1517675

[12] Yvonne Rogers, William R. Hazlewood, Paul Marshall, Nick Dalton, and Susanna Hertrich. 2010. Ambient influence: can twinkly lights lure and abstract representations trigger behavioral change? it Proceedings of the 12th ACM international conference on Ubiquitous computing - Ubicomp ´10, ACM Press, 261. http://doi.org/10.1145/1864349.1864372

[13] Regan Mandryk, Kathrin M. Gerling and Kevin G. Stanley. 2014. Designing Games to Discourage Sedentary Behaviour. it Playful User Interfaces: Interfaces that Invite Social and Physical Interaction, Springer Singapore, 253. http://dx.doi.org/10.1007/978-981-4560-96-2_12

[14] Nassim Jafarinaimi, Jodi Forlizzi, Amy Hurst, and John Zimmerman. 2005. Breakaway: An Ambient Display Designed to Change Human Behavior. it CHI ´05 Extended Abstracts on Human Factors in Computing Systems, ACM Press, 1945. http://doi.acm.org/10.1145/1056808.1057063

[15] Adam Wojciechowski and K. Fornalczyk. 2015. Single web camera robust interactive eye-gaze tracking method. Bulletin of the Polish Academy of Sciences Technical Sciences, 63, no. 4, pp. 879-886, 2015DOI: 10.1515/bpasts-2015-0100

[16] Mateusz Półrola and Adam Wojciechowski. 2012. Real-time hand pose estimation using classifiers. *International Conference on Computer Vision and Graphics*, Springer Berlin Heidelberg.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1631–1640
DOI: 10.15439/2016F535
ACSIS, Vol. 8. ISSN 2300-5963

Limitations of Emotion Recognition in Software User Experience Evaluation Context

Agnieszka Landowska, Jakub Miler
Gdansk University of Technology, Narutowicza St. 11/12, 80-233, Gdansk, Poland
E-mail: {nailie, jakubm}@eti.pg.gda.pl

***Abstract*—This paper concerns how an affective-behavioural-cognitive approach applies to the evaluation of the software user experience. Although it may seem that affect recognition solutions are accurate in determining the user experience, there are several challenges in practice. This paper aims to explore the limitations of the automatic affect recognition applied in the usability context as well as to propose a set of criteria to select input channels for affect recognition. The results are revealed via a semi-experiment based on the case study of an educational game. As a result, a number of concerns were identified, providing a list of pros and cons for affective computing methods applied in the usability testing context. The lessons learned might be interesting for both researchers that develop emotion recognition algorithms and for practitioners, who apply them to diverse areas.**

I. Introduction

ADVANCES in mobile and ubiquitous technologies have made human-system interaction everyday practice in multiple aspects of life. As a result, natural interaction and positive experience of technology is receiving more and more attention. The traditional notion of software usability, as defined by the ISO 9241 standard, includes the effectiveness, efficiency and satisfaction with which specified users achieve specified goals in particular environments [1]. The term *user experience* goes beyond this definition, emphasising the affective component and forming a more holistic picture of human-system interaction [2].

Producers and the marketing/branding industry are interested in the affective aspect of user experience in terms of software products and create a demand for automatic emotion recognition techniques as a tool for getting a larger quantity of more objective data. However, the application of emotion recognition methods in UX testing is not so straightforward. An analysis of only the affective aspect of the user experience might be not enough to determine the issues of effectiveness, efficiency and satisfaction. Therefore Ahn and Picard [3] proposed the affective-behavioural-cognitive (ABC) framework that combines diverse techniques in order to evaluate the user experience in a more holistic manner. The framework was validated with an experiment on beverages [3] and an evaluation of web applications [4]. The ABC framework intends to address goals and interest of diverse stakeholders.

In our research, we develop and study educational games, where user satisfaction is a key affective component, although there are also typical usability issues involved. We turned to the ABC framework as a solution to evaluate the user experience of the software. However, we uncovered many challenges in the practical application of automatic emotion recognition methods. This paper presents the lessons learned and some adjustments in method, that might be useful for researchers who develop emotion recognition algorithms and for practitioners, who apply them in diverse contexts.

The main research questions addressed by this paper might be formulated as follows: *Which emotion recognition and affect representation techniques are applicable within the procedures of usability/user experience testing? What are the main limitations/challenges in their use? How to provide valuable information derived from affective analysis?*

This paper presents a semi-experiment based on a usability case study of an educational game and is organised as follows. Section 2 outlines the previous research on which we based our study. Section 3 includes the operationalisation of the variables and a study plan, while section 4 and 5 provide details of the study execution along with the results. Section 6 provides a summary of the results and a discussion, followed by some concluding remarks (section 7).

Although the authors are aware of the fact, that user experience is a broader term than usability [5], the paper sometimes uses these terms interchangeably, although in the broader (UX) sense.

II. Related work

Work that is mostly related to this research falls into two categories: (1) studies on emotion recognition based on different input channels and their comparison; (2) the use of affect elicitation techniques in user experience evaluation.

(1) There are numerous emotion recognition algorithms, that differ in terms of input information channels, output

This work was supported in part by the Polish-Norwegian Financial Mechanism Small Grant Scheme under the contract no Pol-Nor/209260/108/2015 as well as by DS Funds of the ETI Faculty, Gdansk University of Technology.

labels or representation model and classification method. The most frequently used emotion recognition methods that might be considered when designing an UX evaluation include: facial expression analysis [6], audio (voice) signal analysis in terms of modulation, textual input analysis, physiological signals as well as behavioural pattern analysis [7].

As literature on affective computing tools is very broad and has already been summarised several times, only a sample of papers on recognition methods are provided here. For a more extensive bibliography on affective computing methods, one may refer to Zeng et al. [8] or to Gunes, et al. [9]. The most important conclusions from a review of the literature related to emotion recognition that are the most relevant to this study might be formulated as follows:

(1) Emotion recognition techniques provide results in diverse models of emotion representation (from dimensional models through Ekman's six discrete basic emotions down to two-class classifiers) [10]; there is no common standard model for representing affect;

(2) No input channel is superior to any other in terms of the accuracy and granularity of emotion recognition [11]; a multimodal approach combining diverse input channels provides the most accurate results in most cases; for a multimodal approach, early or late fusion might be considered [12].

(3) Self-report of emotions, although subjective, is frequently used as a "ground truth" (another approach is manual tagging by qualified observers or physiological observations) [13].

The aforementioned results influenced the decisions made concerning the design of this study, especially that it is advisable to use more than one observation channel. The study design is reported in detail in section III.

(2) The second part of the literature review performed under this study was aimed at exploring how automatic affect elicitation techniques are applied in usability and/or user experience evaluation.

There are a few studies on fusing affect recognition and usability evaluation [2][4][12-17]. Most of them consider the usability of food or everyday items and evaluate the overall experience taking emotional factors into account. The most important paper related to this study introduces the Affective-Behavioural-Cognitive approach to UX evaluation [3]. Another is by Lew et al., providing an example of affect evaluation applied to quality assurance procedures for web applications [4].

Kołakowska et al. proposed involving affect recognition in usability evaluation and have suggested four different scenarios: a first impression test, task-based usability test, free interaction test and comparative test [14][15]. The main contribution of the study is the proposal of an emotional state set that might be important in usability evaluation scenarios: *frustration*, *empowerment*, *interest* (*excitement*), *boredom*, *disgust*, *engagement* and *discouragement*.

Partala and Kallinen suggested using the Positive and Negative Affect Schedule (PANAS) scale in self-reporting user experience [16].

Hazlet and Bendek described two studies that used facial electromyography (EMG) measures combined with verbal and performance measures to provide feedback on the user's emotional state. The multimodal approach used in this study was able to provide a measure of the desirability of features, a measure of emotional tension and mental effort expended while performing tasks [17]. There are also some studies of games that utilise the channels used in emotion recognition [18][19][20].

Zimmerman et al. proposed a new method for measuring mood based on the effects of affective processes on motor-behaviour and uses log-files from the mouse and keyboard as a proxy of the mood of the user [21].

There is one study on the limitations on affect recognition in the usability context and it proposed the following criteria: accuracy of emotion recognition, susceptibility to disturbance, independence of human will and interference with usability testing procedures. These criteria were used in an analysis of the recordings from a case study regarding usability evaluation, however were not put into practice [22].

Although some studies on blending affect recognition and usability testing exist, their practical applicability and interference of emotion recognition with IT user experience testing still requires more exploration.

III. Study Design and Research Methods

In order to verify the applicability of emotion recognition in the software UX evaluation context, a semi-experiment was conducted, based on a typical usability study of an educational game extended with user emotion recognition channels. The concept was to use multiple observation channels at the same time, but only those that do not significantly interfere with the typical usability evaluation procedure. Typical usability tests involve 5-10 participants, as this number is enough to reveal 75-90% of usability issues. We planned up to 10 participants for the experiment, as more participants are rarely involved in usability studies.

In order to conduct the study, we chose an educational game, still at the developmental stage, that would be suitable for performing usability evaluation. This choice influenced the participant group. The experiment had to include both the target group of the software under investigation (at least 5 people) and some participants outside the target group since the target group of the application was quite narrow, and we wanted to involve more age groups in the evaluation of emotion recognition techniques.

A. The software under research and UX evaluation goals

GraPM – an educational game about project management [23] was selected for the study. In this game the player assumes the role of a project manager and aims to complete a given product in a given time with some resources under

the uncertainty of some risks. Different product features have different business value, effort and impact on quality. Resources offer different productivity. Threats and opportunities appear and materialize randomly during the development process, requiring the player to take appropriate actions that he chooses from a given set. The effectiveness of particular actions is left for the player to discover during the game-play. Additionally, the satisfaction of the customer and the team must be monitored, as low ratings can result in the abandonment of the project and losing the game. GraPM involves both deterministic and random factors and requires considerable project optimisation to win the game.

The target group of the GraPM game includes two subgroups: (1) students wanting to develop their knowledge and skills in terms of project management; (2) players who enjoy strategy and management games.

The emotional activations that assist in achieving goals were identified as follows: (a) interest – the player should want to learn; (b) slight confusion – the player must be aware that he does not know everything; (c) joy – the player is pleased that he improves and learns; (d) sense of control – the player is content that he can fully control the project/game and win.

The emotional activations that hinder the achievement of goals were identified as follows: (a) fear – the player should not be afraid of learning; (b) strong confusion (frustration) – the player should not be lost and not know what to do; (c) anger – the player should not get angry that he does not understand the game and cannot win; (d) boredom – the game should not be too repeatable and unchallenging; (e) disregard – the player should not consider the game to be of no educational value.

The evaluation of the user experience of the GraPM game is expected to assess to what extent the user experience goals were achieved, with particular focus on learnability. The players should broaden their understanding of the aspects of project management as well as some principles of effective management such as planning, risk management and project supervision. In terms of the game mechanics, the user experience study is expected to provide observations on where the players encounter problems in manipulating the game, which will limit their ability to learn.

The affective extension of the usability study with the emotion recognition should provide additional information on which features of the game enhance learning and which hamper them. Overall, the extended usability study should allow conclusions to be drawn on how to develop the game to improve its educational efficiency, playability, and enjoyment.

B. ABC framework applied in the operationalisation of UX variables

The affective-behavioural-cognitive approach was used in the transformation of the UX study goals into a definition of the semi-experiment variables. We defined the following three general criteria for UX evaluation: understanding, engagement and enjoyment and the criteria were further operationalised into metrics.

Understanding means that the game is comprehensible for a player and this factor corresponds to the cognitive perception of the game mechanics and the game logic (C-cognitive aspect in the ABC approach). According to the information provided on the game, the mechanics understandability should be evaluated after the 2nd and 3rd game, while the understanding of the game logic should be assessed after the 4th and 5th game. Additionally, a learning curve might be derived based on the progress in consecutive game-play.

Engagement indicates that the game is engaging, that it attracts and maintains interest. This factor is a representation of an observable (B - behavioural) aspect of player-game interaction.

Enjoyment determines whether interaction with the game results in a growth in positive affect symptoms, which corresponds to the affective factor (A - affective aspect in the ABC approach).

The author's description of the game provides a list of desirable emotions: interest, slight confusion, joy and feeling of control and a list of undesirable emotional states: fear, strong confusion (frustration), anger, boredom and disregard.

The emotions were listed spontaneously without any guidance or presentation of affect representation models. This approach was chosen purposefully, as a presentation of the models might have influenced the choice. The emotions were mapped into the models provided by the algorithms chosen for the study and the mapping is described in section V.

In this paper we limit our report to the enjoyment factor, although all three aspects were measured and delivered to the game designers as the result of the study [24].

C. Experiment design

In this study we have used the semi-experiment as a research technique. It was a semi-experiment, as it was not possible to fully randomise the choice of subjects to sample. The experiment was based on a real case study, and a group of convenience was used instead of a randomised sample. However, the sample consisted of: representatives of the game target group (students) and some participants outside the group to represent some confounding variables (e.g. age, education and domain). We also set the group size limit (10), as more participants are rarely involved in UX evaluation and this limitation should be taken into account while assessing the affective factor of the game. In other words, one of the challenges was to determine whether 10 people is enough to provide valuable information on affect.

During the study, a limited the number of input channels were recorded (three). Audio (voice) signal analysis and textual input analysis were not considered for inclusion, because in the case study scenario, these channels of human-system interaction were not used. We decided to capture

video image for facial expressions analysis, to ask for self-report based on the PAD emotion representation model and to record physiological signals for reference (skin conductance). The use of other input channels (e.g. keystroke dynamics or mouse usage patterns analysis) are planned in future experiments.

The game-play (which was performed 5 times) was interspersed with questionnaires that measured: competence progress, self-report on emotions, usability questions, including System Usability Scale and questions on the subjective notion of camera and sensor disturbance.

D. Operationalisation of experiment variables

The main goal of the semi-experiment was to answer the research questions as specified in the introduction section — i.e. to determine which emotion recognition techniques are applicable and provide most value in the UX context.

This challenge was conceptualised using a Goal-Question-Metric technique.

GOAL: Analyse the emotion recognition solutions in order to characterise it with respect to applicability from the point view of experimenters relative to the user experience evaluation.

Q1: Is the procedure of software user experience compatible with emotion elicitation techniques?

Q2: Does application of such techniques hinder the process of usability evaluation?

Q3: Does the application of emotion elicitation techniques provide valuable information from the viewpoint of the UX evaluation goals?

These questions are mapped into the following three criteria: applicability, interference and affect-awareness gain.

Applicability represents the degree to which the emotion recognition techniques might be deployed into UX study and the criterion is divided into two factors: input channel availability and susceptibility to noise.

Input channel availability in the UX context was measured by the metric (AP1) time available/time of study ratio.

This study was not focused on the accuracy of classifiers, but rather on the interference (disturbance) introduced by the UX context, as the input channel might be unavailable or significantly noisy.

Susceptibility to noise was evaluated with different proxy metrics for diverse input channels and then qualified to the metric (AP2) level of susceptibility with a common scale of high-medium-low values. The proxy metric for the skin conductance input channel was the number of events that disrupt the channel per time unit (minute) and the events were defined as mouse clicks, which introduced movement artefacts to the EDA signal. The SC sensors (we used two) were placed on the base of the finger and on the wrist [25] and although not all mouse clicks introduced artefacts, most of them did.

For the video channel we used the quality of consecutive frames as the proxy metric.

The ***Interference*** factor measures the influence of emotion recognition application on the usability study. Changes in the usability study (introduced by emotion recognition) should not significantly influence the main goals of the usability study – i.e. gathering information on user effectiveness and learning with software. The factor was measured by 2 metrics: self-report on the subjective notion of camera (IN1) and sensor disturbance (IN2). In the self-report we used a 5-item scale from: 5 - very intrusive to 1 - not intrusive.

Affect-awareness gain is a factor that represents the value of introducing emotion recognition techniques into the software UX context. The criterion is divided in this study into three factors: (AA1) compatibility of the emotion classifier output with emotional states recognition requirements (the ones specified in advance); (AA2) consistency of multimodal observations; (AA3) subjective opinion of the customers of the extended UX study on the value provided by different information on the affective states of the user.

The compatibility metric (AA1) was evaluated for four emotion representation model types (6 basic emotions, arousal only, valence-arousal and PAD positiveness-arousal-dominance models). We used the following scale: 0 – no representation in the chosen model; 1 – could be represented, but might be confused with other emotions; 2 – could be easily and unambiguously mapped; 3 – directly available in the representation model.

Remarks on the consistency of multimodal observations (AA2) were introduced to this study but they will not be evaluated quantitatively. This criterion added, as we observed, huge discrepancies between emotional states estimated on diverse input channels. However, the evaluation of the discrepancy, its scale and analysis of its causes go far beyond the scope of this study. We decided to merely report it, as the differences might compromise the affect-awareness gain.

The results of the players' affect elicitation and analysis were presented to the game designer (the second author of this study) and were evaluated based on the criterion of value they bring to the understanding of the user experience with the game (AA3 metric). The value was measured on a 5-point scale ranging from: (5) very informative to (1) no affect-awareness gain. The designer evaluated the following: the perspectives offered in the presentation of the study results, the views used to visualise data and overall affect-awareness gain in understanding usability and the user experience.

IV. STUDY EXECUTION AND THE PRESENTATION OF THE UX RESULTS

The case study was carried out in 2016 at the Emotion Monitor Stand at Gdansk University of Technology [26]. The The following equipment was used: (1) for physiological signals tracking and analysis: Thought Technology ProComp Infiniti coder, compatible sensors,

Biograph Infiniti Physiology Suite software; (2) for video analysis: a standard Internet camera and video capture software from Logitech, for analysis of facial expressions, Noldus FaceReader was used; (3) for screen capture and user activity tracking and analysis – Morae Recorder, Observer and Manager were used. The three capturing sets were operated at three computer workstations.

We used 2 skin conductance sensors placed on: left-hand fingers and right-hand wrist (for right-handed participants). The locations of the sensors were chosen based on a previous study on the interference of mouse and keyboard usage movements with physiological signals from the fingers [25].

The camera was located above the monitor screen, centrally. Video capture was performed with a 29 FPS rate, 1280x720 resolution and saved as a mp4 file. The analysis of facial expressions was performed using Noldus FaceReader software, providing both Ekman's six basic emotion vectors for each frame as well as valence and arousal model time series.

Morae Recorder was used to capture the screen and gather questionnaire responses and Morae Manager was used to analyse the results.

The study was carried out in April and May 2016. The entire experiment involved 10 participants aged 23 to 43 (8 of them belonged to the game target group), 5 male and 5 female.

The results of the affective aspect of the UX study were reported to the game designer using three perspectives and seven views:

Perspective 1. All UX study participants – information on emotions was summarized for all UX study participants and all tasks performed. The perspective used the following views:

View #1. Declared emotional states versus desired and undesired emotional states

View #2: Recognized emotional states versus desired and undesired emotional states

Perspective 2. Single participant between-task analysis – provides information on fluctuation of emotional states between consecutive gameplays and uses following views:

View #3. Declared emotional states after each gameplay

View #4. Declared/recognized emotional states per gameplay

Perspective 3. Single participant single gameplay analysis – provides detailed information on how emotional state fluctuated during single gameplay, which might be combined with the events. This perspective uses following views:

View #5. Relative frequency of emotional states in Ekman's six basic emotions model;

View #6. Fluctuation of valence (positive/negative state);

View #7. Fluctuation of arousal (calm/active state).

Figure 1 provides sample analytical views on emotion as provided as a result of the UX study.

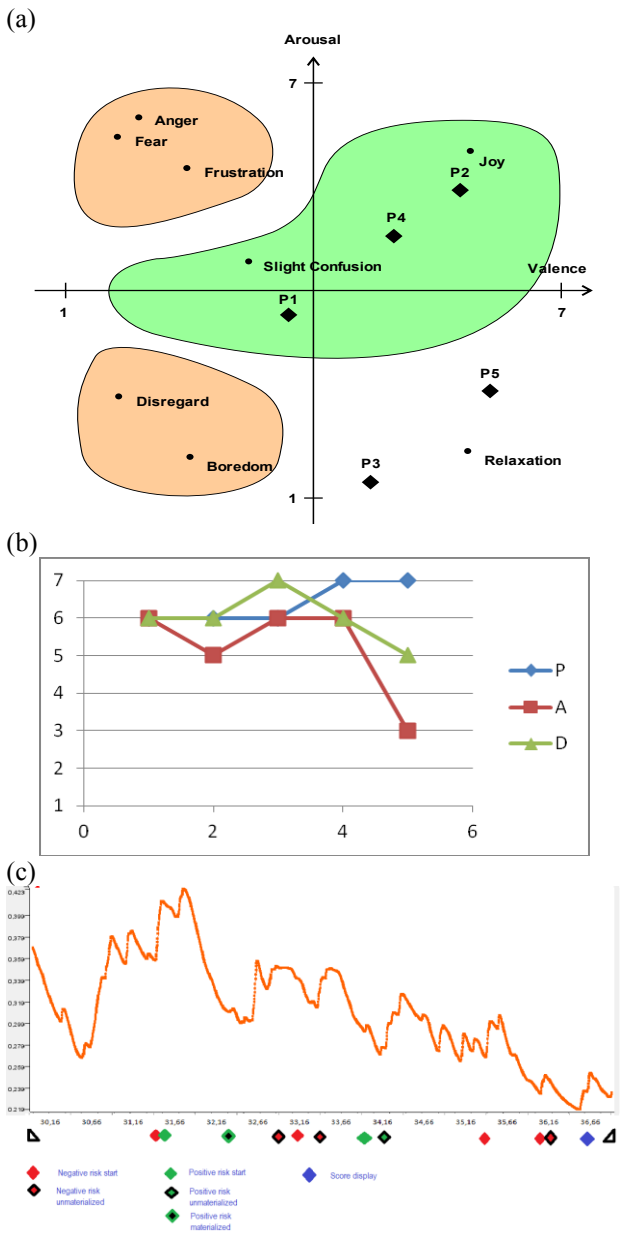


Fig. 1 Sample views from the UX emotion analysis report (a) view #1 (b) view #3 and (c) view #7

View #1 provided in Figure 1 (a) combines various information regarding emotional states. Firstly, it visualises desired and undesired emotional states using the valence-arousal model for emotion representation. The emotions listed by the game designers were clustered into three regions (the scope of the regions was a result of the preliminary mapping of emotional labels to the model and then discussion with the game designer). This preliminary clustering was used in a number of perspectives showing: reported emotional states versus desired/undesired (view #1), recognised emotional states versus desired/undesired (view #2) and one-player emotional state fluctuation from task to task (view #4).

The view provided in Figure 1(b) visualises the change in reported emotional state after consecutive game-plays. It uses the same scale as was used in the questionnaire (1 to 7).

The view provided in Figure 1(c) visualises the changes in the recognised level of arousal accompanied by event markers.

A detailed report on emotional states is not provided here; the views are provided in order to exemplify how the emotions were reported and to show the subject of the evaluation with the affect-awareness gain metrics.

V. Study results

The case study led to some qualitative and quantitative observations. First of all, most of the emotion recognition channels merge naturally with the software usability testing procedure — e.g. the video captured from the camera located near the screen as well as the filling-in of self-report questionnaires on affect. This section reports the results of the study and follows variables as defined in section III.D: applicability, interference and affect-awareness gain. The players participating in the study are coded P1 to P10.

A. Availability and noise susceptibility of the input channels

Applicability was represented with the metric (AP1) time available/time of study ratio and the metric (AP2) level of susceptibility to noise.

The results obtained for the AP1 metric are provided in Table I.

TABLE I.
METRICS OF AVAILABILITY OF THE INPUT CHANNELS

Participant	Self-report (AP1)	Video			Galvanic Skin Response (AP1)
		No of frames	Available frames	AP1	
P1	100%	56313	56313	100%	100%
P2	100%	67392	46323	69%	100%
P3	100%	97354	83601	86%	100%
P4	100%	69325	68512	99%	100%
P5	100%	90101	59275	66%	100%
P6	100%	*	*	*	100%
P7	100%	181135	76423	42%	50%**
P8	100%	56124	54753	98%	100%
P9	100%	70929	70747	99%	100%
P10	100%	69325	65451	94%	100%
Total	100%	na	na	77%	95%

* due to some disk error the video file was corrupted and unrecoverable

** one of the SC sensors detached during the recording session

The self-report on emotional state was merged with the usability study procedure and therefore all participants filled in all 5 questionnaires. Physiological signals were recorded and in one case (P7) only one of the sensors slipped off before the end of the recording. As sensor detachment is a random event, we might assume that such events might occur while hands are used in the human-computer interaction. Video availability varied among participants from 42% up to 100% and this is a result of individual movement patterns. Some participants were seated straight during most of the game-play time. Those with lower video channel availability displayed a number of behaviours that made facial expressions unavailable, e.g. moving sideways (outside the camera range), significant head movement, manipulating the hands in front of the face etc.

The susceptibility to noise (AP2) metrics are provided in table II.

TABLE II.
METRICS OF THE INPUT CHANNELS SUSCEPTIBILITY TO NOISE

Input channel	Proxy metric	Statistics		Susceptibility to noise (AP2)
		Avg	Min	
Video	Frame quality	0,87	0,70	23%
Skin conductance	Time between events causing artifacts [s]	7,79	1,81	80%
Self-report	Subjectivity	na	na	na

Video quality was relatively high and this factor might be improved by using a proper lighting rig as well as improvement in camera resolution. The channel, if available, is quite robust to noise.

One of the prerequisites in recording skin conductance is to restrict the movement of the particular body part that the sensor is attached to. As sensors for skin conductance are placed on the hands, a significant number of movement artefacts occur while using the keyboard and mouse in the UX evaluation procedure. If a keyboard and mouse must be used, the susceptibility to noise is quite high and the condition might be eliminated by moving the sensors to an off-hand location. Feet are mentioned as one of the possible SC locations, although the comfort of the participant might be compromised thereby.

B. Interference with UX procedure

After taking part in the UX evaluation procedure, each participant was asked about the subjective disturbance of the video observation. Camera presence was rated as 1 – non intrusive by 9 out of 10 participants, the other one rated the camera 2 on a 5-point disturbance scale, providing an average disturbance (IN1 metric) equal to 1.1.

We expected that using the camera and screen capture might cause the Hawthorne effect (people behave differently, while being observed), but we did not notice such symptoms.

Physiological measurements require the placement of sensors at the base of the fingers base and on the wrist. Sensor presence was rated as 1 (non -intrusive) by 5 out of 10 participants and 2 (slightly intrusive) by 5 of them Only one rated the sensors 3 on the scale of disturbance so, as a

result, the average disturbance of the sensors (IN2 metric) was equal to 1.6

The result indicates that both sensors and camera were non-invasive, on the whole, for the participants. Some of them even claimed, that they forgot that they were being recorded.

C. Affect-awareness gain

The first aspect of affect-awareness gain is the compatibility of the provided results with the desired/undesired emotions in terms of the affect representation model. The results from facial recognition according to the Noldus FaceReader might be obtained in the form of a 7-item vector of values within <0,1> range corresponding to: anger, joy, disgust, sadness, surprise, fear and neutral state. The results might be also exported to the valence-arousal model of emotions. Physiological signals mainly provide information on arousal and less on valence and therefore should be cautiously interpreted as a labeled emotional state.

The author's description of the game provided a list of desirable emotions: interest, slight confusion, joy and feeling of control and a list of undesirable emotional states: fear, strong confusion (frustration), anger, boredom and disregard. Some of them map directly into the models used by emotion recognition algorithms, and some others require a model of mapping. Table III shows how well the emotions map into diverse models for representation. We used the following scale: 0 – no representation in the chosen model; 1 – could be represented, but might be confused with other emotions; 2 – could be easily and unambiguously mapped; 3 – directly available in the representation model.

TABLE III.
DESIRED/UNDESIRED EMOTIONS AND THEIR MAPPING INTO EMOTION REPRESENTATION MODELS.

Emotion label	Model compatibility (AA1)			
	6 basic +neutral	Valence-Arousal	Arousal only	Valence-Arousal-Dominance
interest	0	1	0	1
slight confusion	2	2	1	2
joy	3	2	1	2
feeling of control	0	0	0	2
Fear	3	1	1	2
strong confusion (frustration)	2	1	1	2
anger	3	1	1	2
boredom	0	2	2	2
Disregard	2	1	1	2
Total	15	11	8	17

Out of the four specified desired emotional states, two are hard to map: interest as a more cognitive than affective state and the feeling of control, which is expressible only with the third dimension of the PAD model – dominance. Some of the desired states are directly available in Ekman's six basic emotions model that includes joy, fear, anger, disgust, surprise and sadness. However, boredom and feeling of control have no representation in this model. Therefore in this study we decided to use the PAD model of emotions. The affect self-report questionnaires were based on this model. For visualisation purpose only, we omitted dominance in some charts (view #1, #2 an #4).

If we use a PAD representation model, the dimensions must be obtained independently from the input channels. The video channel is able to provide valence and some estimation of the arousal, but not in terms of the dominance factor. Physiology-based emotion recognition contributes mainly to the arousal dimension. The only channel that provides the dominance is the self-report. This means that it is difficult to provide one value of emotion estimation reliability. Three independent metrics – one for each dimension should be provided instead.

During the study, we encountered huge discrepancies between the input channels — the self-report and the emotions recognised from the facial expressions were especially contradictory. Sample result showing the inconsistency is provided in figure 2. The figure presents the single-player all-task view #4. Consecutive game-plays were coded as G1 to G5. Diamonds shapes show the reported emotional states, while fuzzy circles depict the recognised ones. The middle of the circle is placed in an area close to the mean recognised state and the fuzziness corresponds to some fluctuations of the recognised emotional state around the average value.

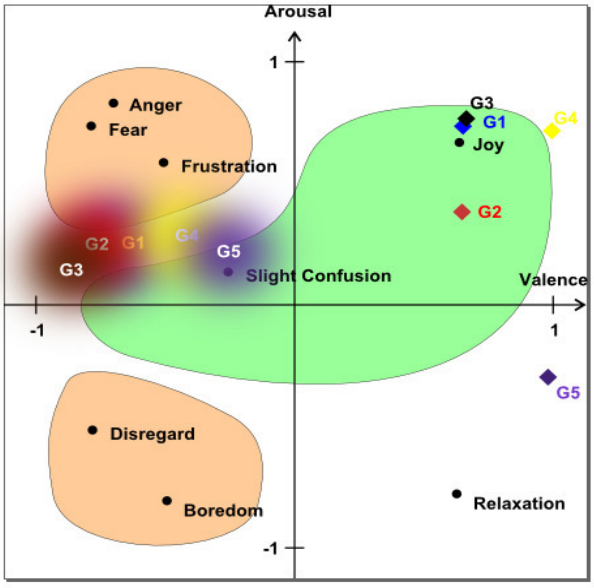


Fig. 2 Single player all-task view illustrating the inconsistency of reported and recognised emotional states

One might observe that the reported emotional states are mainly positive, with varying arousal. The recognised emotional states oscillated in the negative, high arousal quarter (but no extreme arousal levels were observed). The inconsistencies were observed for multiple users and game-plays and the systematic nature of this observation suggests some measurement or recognition error. The consistency of the multimodal observations (AA2) will not be evaluated quantitatively, as this goes far beyond the scope of this study. In trying to explain the reasons for the discrepancy, we watched some sample recordings. We did not notice signs of anger, which was recognised as a dominant emotion in the video. It seemed that perhaps the observable symptoms of concentration (lowered eyebrows) were mistakenly taken as signs of anger. One hypothesis is that the location of the camera above the screen was inappropriate. The inconsistency requires further research and we are planning more experiments to study it.

In this study we adapted to the inconsistency by reporting both recognised and reported emotional states in the UX emotional analysis reports.

The results of the players' affect elicitation and analysis were presented to the game designer (the second author of this study) and were evaluated. The results on the value they bring to the understanding of the user experience with the game (AA3 metric) are provided in table IV.

Some of the perspectives and views provide more information than others. Moreover, it seems that an application of the ABC approach provides more insight into user experience than usability, as was expected.

TABLE IV.
SUBJECTIVE MEASURE OF AFFECT-AWARENESS GAIN

Item type	Item name	Affect-awareness gain (AA3)
Perspective	All UX study participants	4
	Single participant between-task analysis	4
	Single participant single gameplay	2
View	View #1. Reported emotional states versus desired and undesired emotional states	5
	View #2: Recognized emotional states versus desired and undesired emotional states	3
	View #3. Declared emotional states after each gameplay	5
	View #4. Declared/recognized emotional states per gameplay	3
	View #5. Relative frequency of emotional states	2
	View #6. Fluctuation of valence	2
	View #7. Fluctuation of arousal	3
General	Usability understanding	2
	UX understanding	4

Apart from quantitative questions, the survey presented to the game designers included some open questions whereby multiple valuable observations and suggestions were provided:

(1) regarding perspectives: the single-player single-game-play perspective was considered as the least informative, as there were too few in-game events identified to make it interesting; a fourth perspective was proposed to report multiple player single game-play experience;

(2) regarding views: the emotions should be defined using the desired/undesired emotions as listed by the UX goals; some views should have a different scale; a new view should be added that indicates the fluctuation of emotions averaged among the users;

(3) regarding visualisation and reporting: legends and more detailed information should be provided to improve the understanding of the charts;

(4) regarding inconsistencies: the designer tends to believe in self-reporting rather than recognised emotional state, suggesting recognition error.

The most surprising comment was the question: "What is *neutral emotional state*?" This question raises the important issue concerning the proper understanding of the scales and models used for representing emotional states.

One of the expectations, which was not fulfilled in this report, was to provide a chart tagged with events, indicating that certain events caused anger and others caused joy, averaged among the participants. As emotional reactions are very individual, perhaps it would be easier to spot a single nervous system activation than to average the reaction to certain events among the users.

The most important information on affect-awareness gain is that despite the inconsistencies, reporting and visualisation imperfections, the designer claimed that he gained a valuable insight into user experience with the software (rated 4 on a scale from 1 to 5).

VI. SUMMARY OF RESULTS AND DISCUSSION

The main observations revealed through this case study might be summarised with the following statements:

1. It is possible to incorporate emotion elicitation techniques into UX procedures and it seems that emotion recognition has no negative impact on the usability evaluation;
2. Some input channels used in emotion recognition are hard to introduce (e.g. sentiment analysis from text), while others, especially video, self-reporting (and also keystroke dynamics) merge naturally into the UX context;
3. The accuracy of the emotion recognition techniques is compromised by: temporary unavailability of the input channels and susceptibility to noise;
4. Understanding the emotion representation models and mapping the desired and undesired affect to those

models is a preliminary step in providing valuable results.

The practical implications of this study on further applications of emotion elicitation in UX evaluation procedures include:

(1) The advisability to start with the definition of the desired and undesired emotional states and then to map them into one of the representation models for obtaining results.

(2) Selection of the emotion elicitation technique using a set of criteria: availability and susceptibility to noise, possible direct or indirect mapping of the desired/undesired emotional states to the algorithm's output.

(3) As all input channels are subject to temporal unavailability and noise, the challenge might be addressed using a multimodal approach;

(4) While using multiple observation channels, one must look for inconsistencies and the reasons behind them; in the case of discrepancies, perhaps manual tagging by a qualified psychologist might be considered.

(5) Present the results in the form of simple, standard views and provide detailed explanations (assume there is no obvious term regarding emotions).

The results obtained in this study might have some implications for the research on emotion recognition solutions and their integration. The following list of challenges have been identified during this study:

(1) The integration requires a common affect representation model or some mapping between the models. It is quite difficult to integrate and compare results based on labels — a discrete or continuous model might be considered instead.

(2) Emotion recognition algorithms still require improving accuracies and special attention should be given to wild-like conditions, in accordance with current trends in research.

(3) The temporary unavailability of the input channels might be bypassed if the algorithms provide some estimate of the quality of the result (although currently no algorithm does).

We are aware of the fact that the validity of this study has some limitations. We identified and addressed the following threats to its validity: (1) sample size – we engaged 10 users as the usability tests show that 5-10 users reveal 75-90% of the usability issues; (2) sample as a group of convenience – we selected the sample for the UX evaluation to ensure its diversity; (3) confounding variables – we performed the study in a strictly controlled environment, where we limited the possible influences of external factors; (4) subjective measurements – we operationalised most of the variables to objective metrics, the number of subjective self-reports is minimal; (5) observations are based on one case study only – more are planned in the future.

In the future research we would shift to a quantitative approach with results based on a couple of experiments/case studies of UX evaluation procedures based on the ABC framework.

VII. Conclusions

The study revealed three types of challenges: technical, organisational, and related to the cost/value ratio.

Technical challenges (e.g. the accuracy and disturbance robustness of emotion recognition algorithms) might be solved with the future evolution of the affective computing domain, as nowadays emotion recognition in-the-wild conditions is receiving more and more attention. Organisational issues might be eliminated with more experience and trying out different approaches (e.g. multiplying cameras, re-locating sensors). The main challenge remains to provide a reasonable cost/value ratio. Typical usability tests involve 5 to 10 participants, as this number allows 75-90% of usability issues to be revealed. The analysis of emotional states, even when employing automatic affect recognition, is labor-intensive. Multiplying input channels results in higher accuracies, but also introduces the challenge of integration, especially when the observations are contradictory.

Acknowledgment

Authors thank Dominika Makowiecka, who helped in the experiment setting and execution as well as colleagues from the Emotions in HCI Research Group at GUT (emorg.eu), who provided valuable comments on this study.

References

[1] ISO. 1998, Norm 9241: Ergonomics of human-system interaction.

[2] N. Bevan, 2009. What is the difference between the purpose of usability and user experience evaluation methods. In: Proceedings of the Workshop UXEM'09 (INTERACT'09), Uppsala, Sweden.

[3] H.I. Ahn, and R. Picard, 2014. Measuring Affective-Cognitive Experience and Predicting Market Success. IEEE Transactions on Affective Comp. 5(2):173-186, doi: 10.1109/TAFFC.2014.2330614

[4] P. Lew, L. Olsina, P. Becker and L. Zhang, 2012, An integrated strategy to systematically understand and manage quality in use for web applications. Requirements Engineering, 17(4): 299-330 doi: 10.1007/s00766-011-0128-x.

[5] W. Albert and T. Tullis. 2013. Measuring the user experience: collecting, analyzing, and presenting usability metrics. Morgan Kaufmann, USA.

[6] M. Szwoch and P. Pieniążek. 2015. Facial Emotion Recognition Using Depth Data, *The 8th Int. Conf. on Human System Interaction*, pp. 271-277, IEEE, doi: 10.1109/HSI.2015.7170679

[7] A. Kołakowska. 2015, Recognizing emotions on the basis of keystroke dynamics, *Proc. of the 8th International Conference on Human System Interaction*, Poland, doi: 10.1109/HSI.2015.7170682

[8] Z. Zeng, M. Pantic, G. Roisman, and T.S. Huang, 2009. A survey of affect recognition methods: Audio, visual, and spontaneous expressions. IEEE Transactions on Pattern Analysis and Machine Intelligence, , 31(1): 39-58, doi: 10.1109/TPAMI.2008.52.

[9] H.Gunes and B. Schuller, 2013. Categorical and dimensional affect analysis in continuous input: Current trends and future directions, *Image and Vision Computing*, 31:120-136, doi: 10.1016/j.imavis.2012.06.016

[10] A. Kołakowska, A. Landowska, M. Szwoch, W. Szwoch and M.R. Wrobel. 2015. Modeling emotions for affect-aware applications. In *Information Systems Development and Applications*, University of Gdańsk, Poland, pp. 55–69

[11] H. Gunes and M. Piccardi. 2005. Affect Recognition from Face and Body: Early Fusion versus Late Fusion, in *Proc. IEEE International Conference on Systems, Man and Cybernetics*, pp. 3437-3443. doi: 10.1109/ICSMC.2005.1571679

[12] I. Hupont, S. Ballano, S. Baldassarri and E. Cerezo. 2011. Scalable multimodal fusion for continuous affect sensing, *IEEE Workshop on Affective Computational Intelligence*, pp.1,8, 11-15, doi: 10.1109/WACI.2011.5953150

[13] J.N. Bailenson, E.D. Pontikakis, I.B. Mauss, J.J. Gross, M.E. Jabon, C.A.C. Hutcherson, C. Nass and O. John. 2008. Real-time classification of evoked emotions using facial feature tracking and physiological responses, International Journal of Human-Computer Studies, 66(5): 303-317, doi: doi:10.1016/j.ijhcs.2007.10.011.

[14] A. Kołakowska, A. Landowska, M. Szwoch, W. Szwoch, M.R. Wróbel. 2013. Emotion recognition and its application in software engineering, Proc. of 6th International Conference on Human-System Interaction, Poland, pp. 532 - 539, doi: 10.1109/HSI.2013.6577877

[15] A. Kołakowska, A. Landowska, M. Szwoch, W. Szwoch, M.R. Wróbel. 2014. Emotion recognition and its applications, Human-Computer Systems Interaction: Backgrounds and Applications 3. pp. 51-62, Springer. doi: 10.1007/978-3-319-08491-6_5

[16] T. Partala, A. Kallinen. 2012. *Understanding the Most Satisfying and Unsatisfying User Experiences: Emotions, Psychological Needs, and Context. Interacting with Computers,* 24(1):25–34. doi:10.1016/j.intcom.2011.10.001.

[17] R. Hazlett, J. Benedek 2007. *Measuring emotional valence to understand the user's experience of software*, Int. J. Human-Computer Studies, 65:306-314. doi:10.1016/j.ijhcs.2006.11.005

[18] A. Landowska, M.R. Wróbel 2015. Affective reactions to playing digital games, *8th International Conference on Human System Interaction,* IEEE, pp.264-270. doi: 10.1109/HSI.2015.7170678

[19] L. Chittaro, R. Sioni 2014. Affective Computing vs. Affective Placebo: Study of a Biofeedback-Controlled Game for Relaxation Training. International Journal of Human-Computer Studies, 72, 8–9, pp. 663–73. doi:10.1016/j.ijhcs.2014.01.007.

[20] W. Szwoch. 2015. Model of emotions for game players, *8th International Conference on Human System Interactions*, IEEE, pp.285-290. 10.1109/HSI.2015.7170681

[21] P. Zimmermann, P. Gomez, B. Danuser, S. Schar. 2006. Extending usability: putting affect into the user-experience, in *Proc. of Nordic Conf. on Human-Computer Interaction,* Oslo, pp 27-32.

[22] A. Landowska. 2015. Towards Emotion Acquisition in IT Usability Evaluation Context, *Proceedings of the Multimedia, Interaction, Design and Innovation*, 5, doi:10.1145/2814464.2814470

[23] GraPM website, http://grapm.html-5.me.

[24] J. Miler, A. Landowska. 2016. Designing effective educational games - a case study of a project management game, FedCSIS, Gdansk, Poland (accepted)

[25] A. Landowska, 2014. Emotion monitoring - verification of physiological characteristics measurement procedures, *Metrology and Measurement Systems Journal*, Vol XXI, 4:719-732. doi: 10.2478/mms-2014-0049

[26] A. Landowska, 2015. Emotion monitor-concept, construction and lessons learned, in *Proc. of Computer Science and Information Systems (FedCSIS)*, Łódź, Poland, pp.75-80. doi: 10.15439/2015F264

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1641–1645
DOI: 10.15439/2016F227
ACSIS, Vol. 8. ISSN 2300-5963

Virtual Sightseeing in Immersive 3D Visualization Lab

Jacek Lebiedź, Mariusz Szwoch
Gdańsk University of Technology
Faculty of Electronics, Telecommunications and Informatics
Department of Intelligent Interactive Systems
G. Narutowicza St. 11/12, 80-233 Gdańsk, Poland
Email: {jacekl, szwoch}@eti.pg.gda.pl

***Abstract*—This paper describes the modern Immersive 3D Visualization Lab (I3DVL) established at the Faculty of Electronics, Telecommunications and Informatics of the Gdańsk University of Technology (GUT) and its potential to prepare virtual tours and architectural visualizations on the example of the application allowing a virtual walk through the Coal Market in Gdańsk. The paper presents devices of this laboratory (CAVE, walk simulator etc.), describes methods of "immersing" a human in a virtual environment (city, building etc.) and discusses future possibilities for development (directions of research and limitations of today's hardware and software).**

I. Introduction

VIRTUAL reality (VR) is over 50 year old, now. The first devices like *virtual reality video arcade Sensorama Simulator* or *stereoscopic-television apparatus for individual use* (HMD Head-Mounted Display) were invented in the 1960s [1]. Initially, applications of VR devices were very restricted due to their high price and technological limitations, but now, fifty years later, such devices become very popular on the customer market (e.g. *Oculus Rift*, *HTC Vive*). Their relatively low price, acceptable reliability and passable level of immersion (despite screen-door effect) allow us to use them commonly for video games and other kinds of virtual reality experience.

The virtual reality CAVE is much younger than the HMD. The first *Cave Automatic Virtual Environment* (CAVE) came into being in the early nineties at the University of Illinois [1, 2]. Generally, the CAVE may be defined as a cuboidal chamber that has stereoscopic projection screens instead of the walls, the floor, and sometimes the ceiling. A human visitor, wearing only lightweight 3D glasses, is surrounded by a 3D virtual scene projected by projectors placed outside the CAVE. The three-dimensional impression is intensified by additional adjustment of the images forming the scene to the location of a human head. Visual effects are often supported by a 3D surround audio system. Therefore, the level of immersion is very high for CAVEs.

Dozens of CAVEs have been constructed within the last twenty five years [3, 5]. Some of them are rather simple and consist of only four screens, usually three walls and a floor (e.g. the first CAVE in Poland [6]). Other CAVEs are more sophisticated and contain more screens [12]. Four walls indicate that one of them has to be a gate. The complete six-faced cuboidal CAVEs with four walls, a floor and a ceiling are rather rare.

This work was supported in part by DS Funds of the Faculty of ETI of the Gdańsk University of Technology.

II. Immersive 3D Visualization Lab

The Immersive 3D Visualization Lab [7, 8, 9, 10] contains complete six-faced cubical CAVE made of thick square acrylic plates (Fig. 1). A spectator can see a 3D scene on each CAVE's face using a 120 Hz stereoscopy system in passive mode (spectrum channels separation by selective interference filters) or active one (separation in time with active shutter glasses [10]). The viewer's glasses have special markers that are tracked by four infrared cameras placed in the upper corners of the CAVE. Eight speakers located in the same corners and a subwoofer standing outside the CAVE provide surround audio system.

Fig. 1 The CAVE in the I3DVL (the gate of the CAVE is open)

Simulation participants may walk freely in the CAVE from wall to wall, as in any typical CAVE. However, unlimited virtual wandering is also possible using a handheld controller, called wand or fly-stick. Unlike other common solutions, the CAVE in the I3DVL can use a spherical walk simulator as an additional movement controller (Fig. 2). The spherical walk simulator [4, 11, 13] has a form of an openwork sphere, that freely rotates on rollers with a small friction. One can treat it like a human size omnidirectional "hamster wheel". A user may walk

inside the sphere watching, through its rotating surface, the images on the CAVE screens that change according to the direction and speed of the sphere revolutions. This solution allows the user to march on foot through the virtual world projected on the CAVE screens without any space limitations. There are no other CAVEs in the world with such possibility.

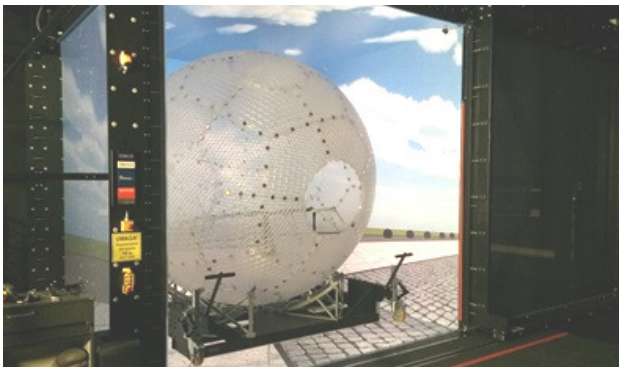


Fig. 2 The spherical walk simulator inside the CAVE in the I3DVL (the gate of the CAVE is open)

The CAVE in the I3DVL can be used with the spherical walk simulator, but both devices can also work separately: the CAVE in the typical configuration with a handheld controller and the spherical walk simulator with the HMD. The specially designed and constructed transport trolley allows to roll the spherical walk simulator in and out (Fig. 2) without a risk of collision with the CAVE's screens. The diameter of the spherical walk simulator is 3.05 m, and the edge of the cubical CAVE is 3.4 m long. Hence, there is only about 17 cm between the surface of the 250 kg sphere and the wall of the CAVE from each side.

Every wall-screen obtains two images from two WUXGA (1920 × 1200) projectors. The final image, combined by blending, has a resolution 1920 × 1920. Thus, all six screens require twelve projectors driven by dozen computers. Two additional computers control virtual scene dynamics, real object tracking, sound generation, and video surveillance.

III. Software Platforms

The Immersive 3D Visualization Lab can work in various software environments that support multicomputer (cluster) process synchronization. Each computer can operate under control of Windows or Linux operating systems, and VR programmers may use both of them. At present, the most popular software platform in the I3DVL is Unity – the environment dedicated to video game developers. Low-level graphics programming in C++/C# and OpenGL or Direct X is also used, particularly for experiments with global illumination rendering methods like raytracing and radiosity. Use of other tools for game development (e.g. Unreal Engine) or virtual reality applications (e.g. Bohemia VBS3, Presagis Vega Prime) is possible, too.

The computational power of I3DVL computers might not be sufficient for some complex tasks, such as sophisticated rendering, nontrivial physics calculations, advanced artificial intelligence or prediction of user behavior. In such cases, in order to provide efficient real-time processing, the power of high performance cluster Tryton from the Academic Computer Center in Gdańsk (CI TASK), connected via the fast optical fiber InfiniBand, can be used [14]. This connection allows treating the computers of the I3DVL and the nodes of the cluster Tryton as one uniform cluster.

IV. The Coal Market virtual sightseeing

The Coal Market is one of the historical squares of Gdańsk with many landmarked buildings, e.g. the Great Armoury (Fig. 3), the Straw Tower, the Court of the Brothership of St. George (Fig. 4), the Torture House, the Prison Tower, the Golden Gate (Fig. 4), the Upland Gate, and several cultural institutions. Regrettably, the contemporary buildings at the Coal Market's western frontage contrast with the historic and cultural character of the Market. Therefore, the City Council of Gdańsk announced a competition for new urban-architectural concepts of this square. Over 10 teams from Faculty of Architecture at GUT took part in this contest and, finally, three architectural projects were awarded (Fig. 5-7).

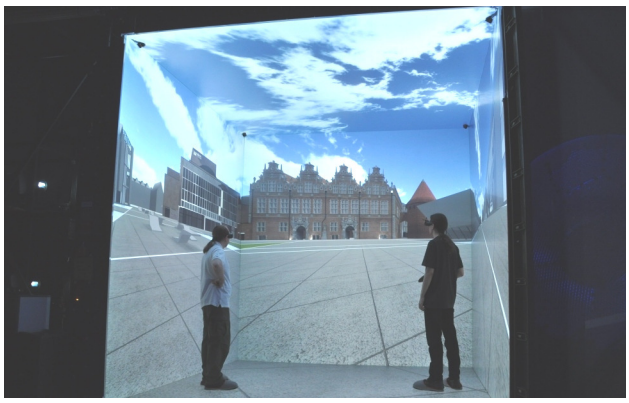


Fig. 3 The virtual Great Armoury

Fig. 4 The Court of the Brothership of St. George and the Golden Gate

This contest provided a good opportunity to propose a new approach to visualize the awarded projects not only as a limited set of static architectonic sketches but as dynamic, interactive real time visualization in the CAVE environment. Spectators, e.g. architects or council members, could observe the modeled urban area from practically any

Fig. 5 Visualization of the western frontage of the Coal Market (project by E. Kowalik *et al.*)

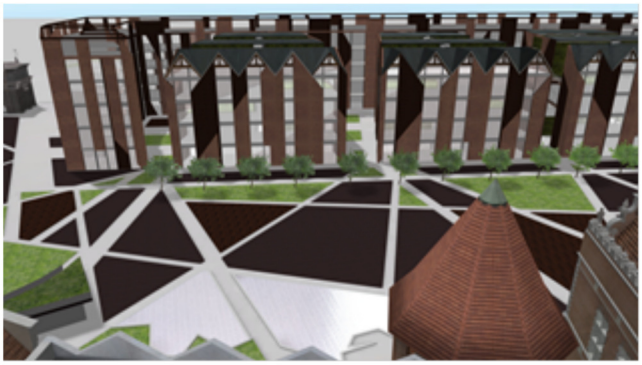


Fig. 6 Visualization of the western frontage of the Coal Market (project by M. Chrzanowska *et al.*)

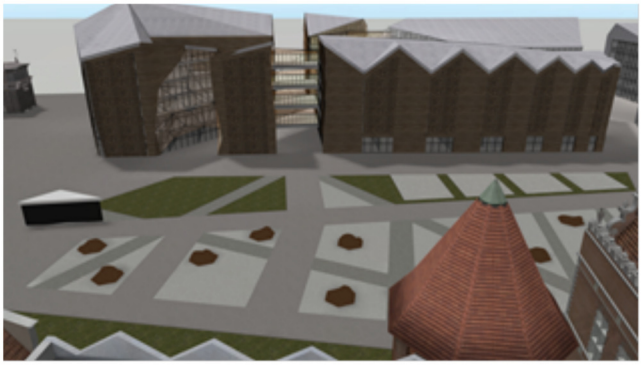


Fig. 7 Visualization of the western frontage of the Coal Market (project by Ł. Plata *et al.*)

viewpoint, and freely walk along virtual houses and buildings with true 360° 3D surrounding view.

Thus, the goal of the project described in this paper was to fully model the Coal Market urban area with three different variants of its western frontage buildings. Additionally, the requirements specification assumed automatic and free walk functionality, and capability of switching between several predefined viewpoints and frontage versions.

Although all models from competition were available in electronic form, two severe reasons prevented their direct usage in the CAVE visualization. Firstly, all adjacent buildings were modelled as single, large 3D objects. While such approach may be used for not real-time architectural visualizations, it prohibits the use of any optimization algorithms in modern 3D graphics engines, such as Level of Detail (LOD) techniques, occlusion culling, and other. Secondly, visualization of all unchanged historical buildings was based on very simplified low-poly models with no textures nor normal vectors but with some geometric errors instead. Although these imperfections were hardly noticeable in static architectural visualization, where these buildings provided the background for contest models, they were unacceptable for interactive CAVE visualization, where all buildings at the Coal Market were equally important.

For these reasons the project's decision was made to model all the Coal Market's buildings from scratch as separate 3D objects. Such models should have significantly reduced complexity while still preserving visualization quality. According to these assumptions, two architect students created new 3D models both for three contest propositions, as well as for ten buildings located in other untouched frontages of the square.

Based on provided models three 3D scenes of the Coal Market were designed in Unity environment with three different variants of the square's western frontage (Fig. 5-7). Unfortunately, the models prepared by the architect students were too detailed, with hundreds of thousands of vertices, even for not significant architectonic details, such as ornate wrought-iron gate's lattice (Fig. 8), dome-like roof details or roof sculptures. Initially, the whole scene consisted of over 5 million vertices, which occurred too complex for real-time visualization, resulting in low refresh ratio lower than 10 fps. Thus, most models had to be simplified by dramatic reduction of the number of their vertices with minimal loss of quality. This reduction was performed by an IT student with 3D modelling skills. Sample effects of performed complexity reduction are presented in Fig. 8. The original 3D model of wrought-iron lattice (a) consisted of nearly 670 thousand vertices and 1.2 million triangles, while the simplified one (b) is built from only 25 vertices and one partially-transparent mid-resolution texture. Original high complexity of the model (c) is not visible during standard CAVE visualization (d).

Finally, the scene complexity were reduced by a factor of 10 resulting in about 500 000 vertices in total. This reduction, accompanied by well-known optimization techniques such as LOD, occlusion culling and light mapping, allowed for effective real-time scene visualization in CAVE environment at refresh rate of about 120 fps.

Additional feedback from architects pointed the way of visualization's improvement by adding simplified building models that would surround the modelled Coal Market and provide some visual background both for a bird's-eye view of the square as well as for wandering around its neighborhood. Such buildings were added based on the simplified building models consisted of only 10 vertices each, which did not increase the overall scene complexity in practice. The results are presented in Fig. 9 where additional buildings are visible in the background.

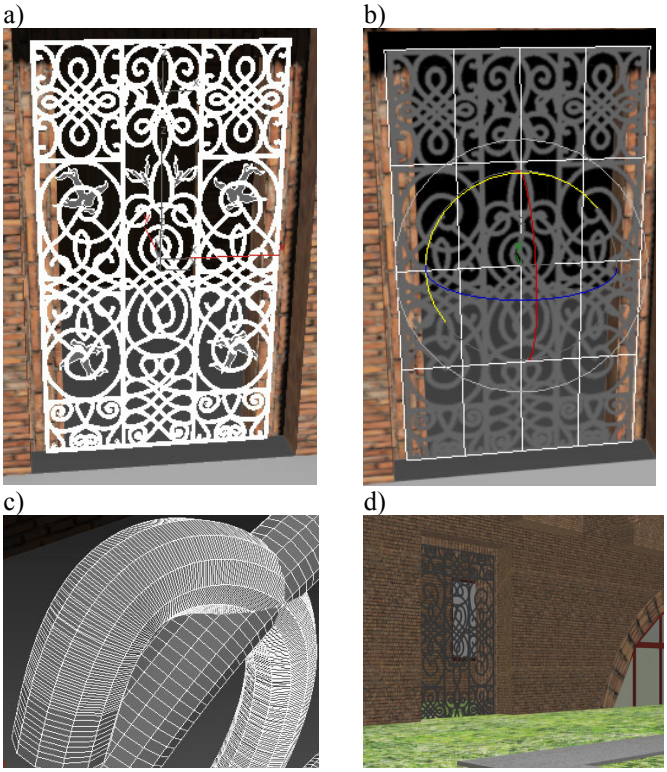


Fig. 8 Sample model of wrought-iron gate's lattice: a) highly detailed original (white color is used for vertices and edges), b) simplified flat model, c) magnified detail of the original model, d) final appearance in 3D visualization

V. VIRTUAL SIGHTSEEING

Visualization of the Coal Market became the first application demonstrating I3DVL capabilities. It was presented to many visitors who confirmed high usability of the CAVE environment in the field of architectural visualizations. Possibility of almost free interaction within a 3D scene and high immersion level, which is guaranteed by the surrounding view, allow for better scene perception. For example, most interaction participants enjoyed the possibility of visiting every nook and cranny of the buildings and courtyards. This possibility of visual testing of all inner passages and communication courses occurred as vital as the general view of building frontage.

Also, the possibility of immediate switching between different scene versions makes their comparison easier than at traditional exhibition as it gives the same visual context. Additional application feature allows for real-time switching between daylight and night scenes (Fig. 10, 11), and thereby the visitors can verify how the place would look like by night with artificial illumination. Yet another kind of experience is the possibility of bird's eye view from highly located viewpoints (Fig. 5-7,9) which is especially realistic inside the CAVE.

A wide multitude of presentations for different peoples have allowed for many interesting observations. For example, some visitors have serious problems with movement control in the CAVE. For such persons the automatic walk mode occurred a real salvation, allowing for simple looking around while being moved forward along the predefined path as inside a sightseeing trolleys or bus. It was also confirmed that some people do have some problems with 3D image perception using 3D glasses. Fortunately, most of them reported that the problem is far less than with the HMDs such as Oculus Rift. This surely results from much higher image quality as well as from better user orientation inside the CAVE. Moreover, a man inside the CAVE can see his own body and other participants of the simulation. The depth of immersion in the CAVE over the HMD seems to be unquestionable, but conclusive proving this general thesis requires further research comparing the reactions of people using different VR devices and various scenarios by means of questionnaires, behavioral observation and biological measurements (as heart rate, blood pressure and electrical activity of the brain).

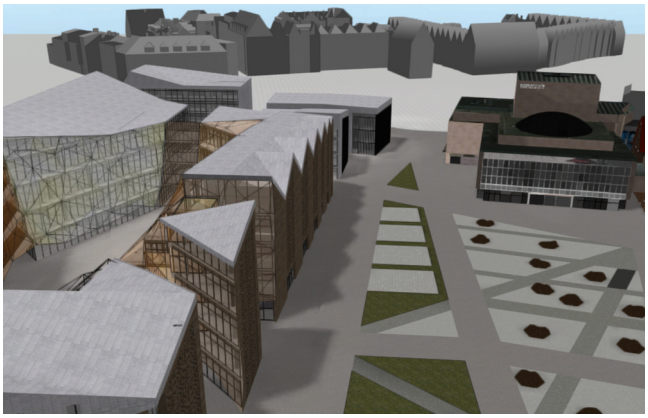


Fig. 9 The bird's-eye view of the virtual Coal Market with simplified buildings in the background

The project realization allows also for gathering invaluable experience and clues for further application development for I3DVL's CAVE environment. For example, many inbuilt Unity mechanisms, such as scene switching or interface elements, had to be redesigned due to the client-server architecture of the CAVE and the need of the network synchronization between I3DVL's computers. Moreover, such a presentation of 3D image requires higher rendering rate than for a PC demonstration.

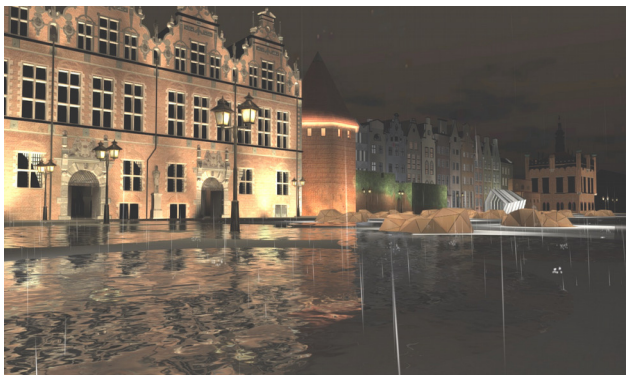


Fig. 10 A night scene of the Coal Market with rain simulation

Fig. 11 A night scene of the Coal Market with crowd simulation

VI. Conclusion and Future work

Virtual sightseeing can be useful not only for virtual tourism, but also for virtual prototyping, allowing assessment of designed objects, like urban areas, squares, streets, buildings with their interiors, and even crafts, vehicles, machines etc. Believable visualization can help in decision-making without the use of spatial imagination. Just take a stroll through the virtual scene and decide.

Future works are focused on using of additional computational power, provided by the high performance Tryton cluster at CI TASK, to expand the Coal Market scene with crowd, snow and rain simulations. The first experiments with developed scalable particle system for cluster-aided visualization proved successful in weather phenomena and crowd simulation (Fig. 10, 11), though many problems of latency-critical processing have to be solved, yet.

Acknowledgment

Authors would like to thank Małgorzata Chrzanowska and Łukasz Plata for preparing 3D architectural models, and Adrianna Szwoch for 3D scene modelling and scripting.

References

[1] G. Burdea, P. Coiffet: *Virtual Reality Technology*, 2nd Ed., Wiley, New York 2003.

[2] C. Cruz-Neira, D. J. Sandin, T. A. DeFanti: The CAVE: A Virtual Reality Theater. *HPCCV Publications*, 2, 1992.

[3] T. A. De Fanti et al.: The future of the CAVE. *Central European Journal of Engineering*. November 2010.

[4] J. Gantenberg, K. Schill, C. Zetzsche: Exploring Virtual Worlds in a Computerised Hamster Wheel. *German Research* 1/2012.

[5] Iowa State University News Service: *The most realistic virtual reality room in the world*. 2006.

[6] J. Jurkojć, P. Wodarski, R. Michnik, M. Gzik, A. Bieniek: The influence of visual parameters of a scenery on the ability to keep the balance in the virtual reality. *13th International Symposium on 3D Analysis of Human Movement*, Lausanne 2014, pp. 88-91.

[7] J. Lebiedź, J. Łubiński, A. Mazikowski: An Immersive 3D Visualization Laboratory Concept. *Proc. of the 2nd Int. Conf. on Information Technology – Information Technologies*, Vol. 18, 2010, pp. 117-120.

[8] J. Lebiedź, A. Mazikowski: Launch of the Immersive 3D Visualization Laboratory. *Szybkobieżne Pojazdy Gąsienicowe*, vol. 34, nr 1, 2014, pp. 49-56.

[9] J. Lebiedź, A. Mazikowski: Innovative Solutions for Immersive 3D Visualization Laboratory. *22nd International Conference on Computer Graphics, Visualization and Computer Vision WSCG 2014 – Communication Papers Proceedings* (ed. Vaclav Skala), Plzeň 2014, pp. 315-319.

[10] A. Mazikowski, J. Lebiedź: Image projection in Immersive 3D Visualization Laboratory. *18th International Conference in Knowledge Based and Intelligent Information and Engineering Systems* KES 2014, Gdynia 2014, *Procedia Computer Science* 35, 2014, pp. 842-850.

[11] E. Medina, R. Fruland, S. Weghorst: Virtusphere: Walking in a Human Size VR Hamster Ball, *Proc. of the Human Factors and Ergonomics Society Annual Meeting* 52, No. 27, 2008, pp. 2102-2106.

[12] A. Mitchell: Lost in l-Space. *INAVATE IML*, 10 2012, pp. 3-4. http://www.inAVateonthenet.net .

[13] VirtuSphere Inc.: Virtusphere. *The Virtual World Immence. Let's Immerse Together*. http://www.virtusphere.com/index.html .

[14] Ł. Wiszniewski and T. Ziółkowski: Real-Time Connection between Immerse 3D Visualization Laboratory and KASKADA platform. *TASK Quarterly* 19, No 4, 2015, pp. 471-480.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1647–1655
DOI: 10.15439/2016F302
ACSIS, Vol. 8. ISSN 2300-5963

Anticipated, Momentary, Episodic, Remembered: the many facets of User eXperience

Patrizia Marti
University of Siena, Department of Social, Political and Cognitive Science and Eindhoven University of Technology, Department of Industrial Design,
Via Roma 56, 53100 Siena Italy
Email: patrizia.marti@unisi.it

Iolanda Iacono
University of Siena, Department of Social, Political and Cognitive Science
Via Roma 56, 53100 Siena Italy
Email: iolanda.iacono@unisi.it

Abstract— **User experience (UX) has been defined in several ways. In general terms, it refers to everything that is individually encountered, perceived, or lived through. The literature on UX reports studies mostly focused on specific interaction events, which may have an impact on the user's emotions and feelings. This paper provides a reflection on how UX evolves over time. We performed a medium term study comparing four types of UX: Anticipated, Momentary, Episodic and Remembered (or Cumulative) experience [1]. Anticipated UX refers to the period of time before first use, and focuses on the expectations a person has on the product, service or system. Momentary UX refers to any perceived change during the interaction in the very moment it occurs. Episodic UX is an appraisal of a specific usage episode extrapolated from a wider interaction event. Remembered UX is the memory the user has after having used the system for a while. The different facets of UX have been analysed in a medium term research spanning over four weeks. The study compared the experience of ten users of a pedometer/fitness app that counts steps and burned calories all day long. The results show that the experience of use changed over time decreasing significantly before, during and after the interaction. The evaluative judgment related to the overall satisfaction with the product, was largely formed on the basis of an initial high expectation on pragmatic aspects (i.e. utility and usability) before and during the first encounters. After four weeks of use, the problems related to usability, reliability of data, and battery drain became a dominant aspect of how good the product was perceived. Hedonic qualities and Attractiveness were negatively impacted as well. The continuous reflection on the use, documented in online diaries, made the problematic aspects prevailing on the overall UX in particular on the evaluation of Episodic and Remembered UX. This prevented any change in behaviour in the participants.**

I. Introduction

Roto et al. [1] edited the "User Experience White Paper", a document reporting the results from Dagstuhl Seminar on Demarcating User Experience, held in September 2010, where 30 experts from academia and industry worked together to define the concept of UX. In this document the editors highlight the multidisciplinary nature of UX, which has led to several definitions of and perspectives on UX. Interestingly they underline the importance of analysing time spans of user experience, stating that the actual experience of usage does not cover all relevant aspect of UX. Time spans matter in determining the UX. People have expectations on a certain product or system before the first encounter. Expectations are often generated by advertisements or others' opinions and have impact on the way people approach the system and prepare to use it.

At the first encounter and during the actual use people may change their appraisal of the system. Pragmatic and hedonic qualities of the product play a fundamental role in determining visceral responses related to momentary feeling perceived during usage [2]. Different episodes of momentary experiences lead to a reflection on the experience itself. Reflection often determines a person's overall impression of a product and many factors come into play when thinking back and reflecting upon the total appeal and experience of use. The outcome of episodic experience is not necessarily equal in value to the sum of momentary experiences. Over time the perception of usage might change again. In remembering the overall experience, people select only few elements, positive or negative, which will determine the general opinion of the product and the chance that it will be recommended to others for later use Fig. 1.

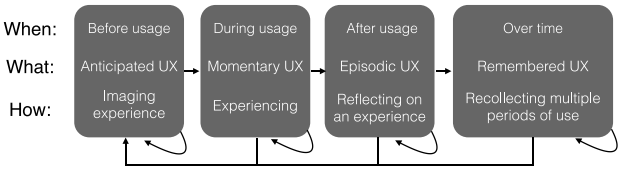


Fig. 1 Time spans of user experience adapted from Roto et al. 2011

Barbara Fredrickson and Daniel Kahneman [3] proposed the model of remembered utility, which dictates that an event is not judged by the entireness of an experience, but by prototypical moments or "snapshots" that are considered representative of an event under uncertainty. The remembered value of snapshots determines the actual value of the entire experience. Fredrickson and Kahneman [3] explained such phenomenon saying that the selected snapshots correspond to the average of the most affectively intense moments of an experience and are related to the resulting feeling experienced in the end. So the duration of the experience does not affect the final

judgment (“duration neglect” effect) while the most intense moments experienced in the end do.

Kahneman [4] studied the differences in perception between the actual and remembered experience through a series of experiments. In 1996, Redelmeier and Kahneman [5] assessed patients' appraisals of a painful colonoscopy procedure. They found that patients evaluated the discomfort of the experience in relation to the intensity of pain occurring in the end of the procedure (peak-end rule). So a peak painful short event occurring in the end is remembered as more negative than a prolonged painful episode occurring in the beginning or the middle of the experience. Length or variation in intensity of pain does not matter.

II. State of the Art

In the field of Experience Design the evaluation of the experience of prolonged use of interactive products is becoming a critical issue. Until few years ago, UX studies have mostly focused on short-term evaluations and the aspects relating to the initial adoption of new product design.

Only recently, an increasing number of studies have started focusing on assessing the changes in a person’s experience in interaction with a product over time [6], [7], [8].

Consequently new methods and models have been defined to understand how the relationship between the user and the product evolves over long periods of time.

Karapanos et al. [9] developed “UX Curve”, a method which aims at assisting users in retrospectively reporting how and why their experience with a product has changed over time.

Mahlke and Thüring [10] developed a model which defines three components of user experience: perception of instrumental qualities (usability and usefulness), emotional reactions and perception of non-instrumental qualities (appeal and attractiveness). Applying this model, they provided evidence that instrumental and non-instrumental qualities influence emotional reactions in the use of interactive systems.

However, the majority of current UX evaluation methods still concentrate on single behavioural episodes and momentary evaluations. Vermeeren et al. [11] report that only 36% of methods focus on long-term period of experience.

Whilst measuring first encounters and momentary experiences is important for collecting feedback from users in particular in the early prototyping phases of the development process [11], recent researches demonstrated that different user experience aspects changes over time [12], [9].

Marti and Iacono [13] compared the experience of use of two tablet applications for zooming in and out while taking photos. They confronted two interaction modalities in the short and medium term: the classic “Slide to zoom” and the novel “Squeeze to zoom”, a squeezable interface. Results obtained in the short-term evaluation revealed that “Squeeze to zoom” was awarded higher values than the “Slide to zoom” in the hedonic quality-stimulation and attractiveness dimensions, whilst it obtained lower values in the pragmatic quality and hedonic quality-identity. However, in the longitudinal study, the usability of “Squeeze to zoom” improved whilst the attractiveness of “Slide to zoom” decreases significantly. Furthermore “Squeeze to zoom” was significantly more appreciated for its hedonic qualities and the effect was maintained over time.

Karapanos et al. [9] evaluated the experience of use of six participants for 1 month after the purchase of an Apple iPhone. They found that the relevance of novelty quickly faded away, while over time different the hedonic quality of the iPhone emerged.

Fenko et al. [12] found also that the perception of importance of sensory modalities changes over time. At the moment of purchasing a mobile phone, they found that vision was the most important perceived modality. After 1 month touch and audition became more important than vision.

In the following we report the result of a medium term study assessing the experience of use of a fitness application for mobile phone, conducted with ten participants.

The study compares four types of experiences as defined by Roto et al. [1]: anticipated experience, momentary experience, episodic experience and remembered experience. Anticipated UX refers to the expectations a person has before the first encounter with the product. Momentary UX refers to individual interaction episodes and the perceived change in use. Episodic UX refers to a usage episode extrapolated from a wider interaction event. Remembered UX refers to the memory of the user after having used the system for a while.

III. Experimental Protocol

The study was conducted in Siena, Italy. Participants were asked to try out over four weeks, Pacer, a fitness application running on smartphone.

Pacer is a free app developed by Pacer Health, Inc. [14] running on Android and iOS platform Fig. 2.

Fig. 2 Pacer interface

It allows to track the steps, whether the phone is in the hand, pocket, in a belt or bag. Pacer records steps, distance, active time and calories burned all day, every day, and gives reminders to keep the person going. It allows to set health goals (e.g. to set the ideal weight) and to stay on target. Pre-defined programs like "from walking to slow ride" are also available. Through a GPS the app allows to track the walking, running or bicycling routes on a map. The ultimate goal is to bring together people based on common health goals and interests with the objective to improve health behaviour change outcomes. Users can create groups, connect with friends via Facebook, motivate each other in physical activities, achieve and compare performances, and ultimately create competitions.

B) Methodology

Ten subjects (M = 5 and F = 5) with an average age of 25.90 were involved in the study for a period of four weeks on a voluntary basis. Five participants were students of the MA course in Experience Design (University of Siena). Five participants were invited to join the study among their groups of friends.

As said above the study aimed to analyse any change among the anticipated, momentary, episodic and remembered experience of use over a month.

The study was conducted using different methods of data collection: an ad-hoc questionnaire to appraise the anticipated and remembered experience, an online-shared diary to assess the momentary experience, and AttrakDiff [15] to assess the episodic and remembered experience Fig. 3.

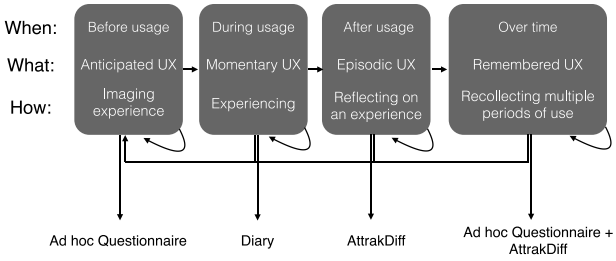


Fig. 3 Overview of the methods used to assess the four types of UX

More in detail the anticipated experience was assessed using an ad-hoc questionnaire focused on the main functionality of Pacer: Step counter, Burned calories, Community, Reminders/Notifications, Settings and use of GPS.

A 5-point Likert scale with values from -2 to 2 was associated to each of the abovementioned functionality. The values were represented in the form of emoticons (-2 = very negative, -1 = negative, 0 = neutral, 1 = positive, 2 = very positive). To evaluate the anticipated experience, the questionnaire was administered at baseline before the app was installed on the participants' smartphones. None of the participants had ever used Pacer. At the beginning of the study all of them received a brief description of the six assessed functionality.

The momentary experience was evaluated using self-reporting. A closed group on Facebook was created to keep a shared diary. All participants joint the group.

The subjects were asked to take note of their experience on the everyday use of the application. The aim of keeping a diary was to express in a narrative form the impressions resulting from the use of the app in the very moment they were experienced by the subject (Momentary UX). The diary entries could be expressed in a free format, using text or images (e. g. screenshots of the app). However, participants were asked to associate an emoticon to each diary entry, the same used to assess the anticipated and remembered experience (-2 = very negative, -1 = negative, 0 = neutral, 1 = positive, 2 = very positive).

The episodic experience was evaluated using AttrakDiff, a questionnaire administered 5 times over a period of four weeks: T_0 = first encounter, T_1 = after 1 week, T_2 = after 2 weeks, T_3 = after 3 weeks, T_4 = after 4 weeks at the end of the study.

AttrakDiff, is a method developed by [15] to assess the user's experience and feelings in relation to interactive products and therefore a product's overall attractiveness. The questionnaire uses the technique of the semantic differential on pairs of opposite adjectives to evaluate the user experience. Users are asked to assess their experience and their perception of the product, responding to

pairs of opposite adjectives. The adjectives are assessed on a seven-point Likert scale, from -3 to 3, in which 0 indicates neutrality. The questionnaire was developed in German and then translated into many languages including English. It consisted of 28 items, broken down into four dimensions:

- *Pragmatic quality or PQ*: describes a product's usability. Indicates how the user can successfully achieve his or her goals using the product. A product need not be particularly beautiful or well-designed to satisfy this quality.
- *Hedonic quality – Identity or HQ-I*: indicates to what extent the product allows the user to identify with it in a certain social context. It relates to what we communicate socially when we use a product. Identification with a brand, for example a certain type of mobile phone, defines our inclinations and preferences of use of that product. Some products are preferred by certain categories of users because they are seen as cool, and not necessarily for the features they offer.
- *Hedonic quality – Stimulation or HQ-S*: indicates to what extent the product can support users' needs in terms of novelty, content, stimulating interaction, presentation of style. It is defined by attributes that encourage users to improve their skills of use of the product. Examples of hedonic stimulation are those features of software applications that are usually little used, and the shortcuts for some commands. Some products offer the user flexibility of use, and the person feels gratified to learn or to find alternative or more effective and efficient modes of use of the product.
- *Attractiveness or ATT:* describes the product's overall value on the basis of perceived quality.

Hedonic and pragmatic qualities are independent of one another, but together contribute to determining attractiveness.

For the present study we used an Italian version of the questionnaire translated by the authors. The same version was used in a previous study [16].

The questionnaire contained 28 items broken down as follows:

Pragmatic quality: *Technical- Human; Complicated-Simple; Impractical- Practical; Cumbersome- Straightforward; Unpredictable- Predictable; Confusing - Clearly structured; Unruly- Manageable*.

Hedonic-identity quality: *Isolating- Connective; Unprofessional- Professional; Tacky-Stylish; Cheap-Premium; Alienating-Integrating; Separates me- Bring me closer; Unpresentable-Presentable.*

Hedonic-stimulation quality: *Conventional-Inventive; Unimaginative-Creative; Cautious-Bold; Conservative-Innovative; Dull-Captivating; Undemanding-Challenging; Ordinary-Novel*.

Attractiveness: *Unpleasant–Pleasant; Ugly-Attractive; Disagreeable-Likeable; Rejecting-Inviting; Bad-Good; Repelling-Appealing; Discouraging –Motivating.*

The remembered experience was assessed in two different ways: 1) using the same ad hoc questionnaire used to evaluated the anticipated experience, in order to compare what was expected with what was remembered; 2) conducting a paired-samples t-test to compare the four UX dimensions of AttrackDiff at time T_0 (first encounter) and T_4 (end of the study).

IV. Results

A) Anticipated and Remembered UX (ad hoc questionnaires)

The data collected on the anticipated UX are reported in Fig. 4.

The 10 participants had a high expectation on the use of the functionality Step counter, Community, Settings and GPS. They did not expect to have a similar positive experience associated to the functionality Calories and Notifications.

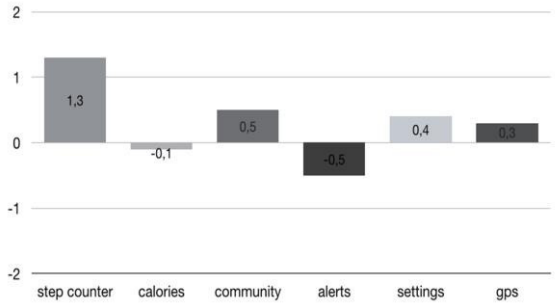


Fig. 4 Mean value of the Questionnaire on Anticipated UX

The data related to the remembered UX after 4 weeks are reported in Fig. 5.

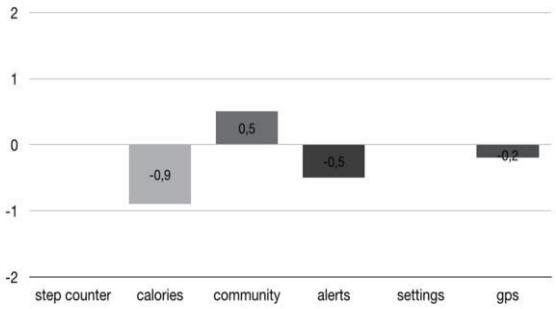


Fig. 5 Mean value of the Questionnaire on Remembered UX

Apparently participants had negative memories on the use of the app. After four weeks all functionality were rated between 0 and -1, except for Community, which maintained the same evaluation of the Anticipated UX.

B) Momentary UX

The 10 participants kept the diaries regularly recording events as soon as they occurred. In total, the corpus of 10 diaries contained 59 entries of which 29 comments were negative, 6 were neutral and 24 were positive. The negative comments were mostly related to poor usability of the app and to an improper way of functioning, which did not comply with the user expectations. These outcomes are consistent with the data collected through the questionnaire on the Anticipated UX.

After the first week of use, a 22-year-old boy wrote the following comment "*The app drains the battery. Therefore if you value more saving battery rather than being sporty and burning calories, you have to start and stop it continuously...*".

A similar comment was entered by a 25-year-old girl "*I receive continuous alerts on the battery draining. This is annoying. I'll try to find a way to stop this. ☹*". (Fig. 6).

Fig. 6 This app could drain the battery.

Another negative comment on the usability was reported by a 33-year-old girl who wrote "*not very positive ... I would say daunting: I tried a challenge but I failed and the app did not tell me why*". A 26-year-old boy wrote: "*After one hour walking the app marks only 223 steps. It is unreliable. I would like to uninstall it*".

After two weeks of use, a 25-year-old boy wrote "*I received a notification asking me if I would recommend the app to a friend, based on my experience of use. I discovered also the possibility to send feedback to developers. Honestly I would not recommend this version to a friend, and I'm tempted to send a full list of negative remarks to the developer"* (Fig. 7).

Fig. 7 "Based on your experience with this version of Pacer, would you recommend it to a friend? Certainly yes, probably, maybe or maybe not, probably no, definitely no."

A 23-year-old girl commented the following: "*I have received this notification for the third time today ... I find it annoying especially if you have been walking the whole day. It pops up when having meal or when sitting for more than an hour ☹. I wish to turn the notifications off*" (Fig. 8)

Fig. 8 "You've been sitting for more than an hour, start moving!"

A 25-year-old girl manifested her disappointment: "*Yesterday afternoon I did jogging and tried the program "from walking to jog". I started the application and discovered that the program required a Premium subscription. ☹ I downloaded another free app that offers the same service for free*".

Some comments explicitly referred to the hedonic-stimulation quality. A 33-year-old girl wrote "*After the initial excitement, after four weeks I find it not really useful. I have not been very active in the past days and Pacer does not motivate me in doing more ☹"*

The most positive comments relate to the Community, that is the possibility to connect to others and share the performance and the achievement of common goals. A 22-year-old boy wrote "*I've just beaten my record of 15.000 steps today!" ☺. This is the notification from Pacer!!!* (Fig. 9).

Fig. 9 Goal achievement

A 33-year-old girl wrote: "*My friend invited me to join her network ... it is nice, I can send her messages and see how many steps she does during the day ☺... I didn't think this functionality would be so engaging for me*".

A 30-year-old boy wrote "*I can see the percentage of people who walk less than me ☺... It is an interesting information since it relates to all people using the app and not only my friends (the numbers wouldn't have been meaningful)*". A 23-year-old girl reported: "*As soon as I woke up this morning, I received a notification of yesterday activity. I discovered I walked more than 32% of users…it is a small but meaningful achievement for me… very positive ☺*" (Fig. 10).

Fig. 10 Goal achievement

After the first week of use, a 25-year-old boy wrote "*I discovered a fantastic functionality!!! I set the program "Sleep eight hours a day and exercise the abdominal muscles". Just after, a weekly calendar appeared on the screen associated to a chat where it was possible to share in real time comments of all users who set the same goal. ☺ An entire world of opportunity disclosed to me*".

To summarise, the negative diary entries were mostly related to low usability of the app, to an untimely use of notifications, and a scarce accuracy of data (e.g. the burned calories or steps). Some participants confronted the data obtained with Pacer with the ones provided by other step counters, realising that Pacer was not reliable. The app was not really motivating for participants and the majority of them uninstalled it at the end of the study. Furthermore, Pacer does not seem to meet the requirements of runners. It displays the pace stat as the overall pace for the whole run, rather than a lap-by-lap breakdown of the pace, which is what the typical runner's app shows. Runners want to know whether the first mile was as fast as the third, for example, and Pacer doesn't tell this.

The positive comments were mainly associated to the Community, Social sharing and Security aspects. In fact, to use the product, it is not necessary to create an account and provide the email address and personal data. For those who are concerned about security, that is a plus.

Overall Pacer did not offer much to explore beyond the basics and this caused a drop of interest after four weeks. The diary entries followed a negative trend over four weeks. Negative comments increased at T_3 and T_4.

C) Episodic UX

As said before, the episodic experience was evaluated using AttrakDiff. The graph presented in Fig. 11 shows the mean values obtained for the 4 dimensions of analysis (PQ; HQ-I; HQ-S and ATT) at time T_0, T_1, T_2, T_3, and T_4.

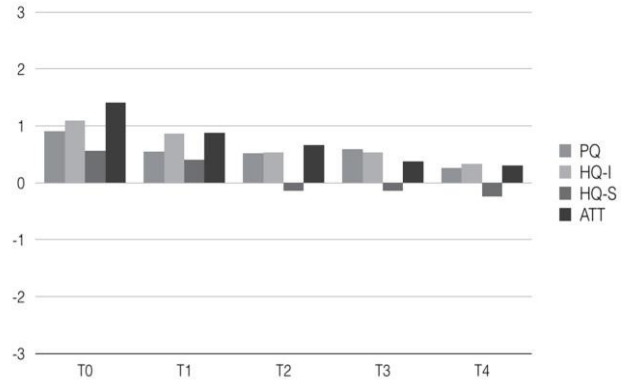


Fig. 11 Mean values for the four AttrakDiff dimensions over the time

Data show a decreasing trend for all dimensions of the analysis, from an initial positive attribution to all dimensions, to a progressive decreasing assessment over the four weeks. The HQ-S at T_4 scored below zero. The dimensions that obtained the highest values during the

evaluation are ATT, HQ-I, and PQ although the decreasing trend is the same for all of them.

Table I provides the average values obtained for the four dimensions and the relative standard deviation (Table I).

A closer look at the evaluation of specific items contained in the PQ dimension *Impractical- Practical; Cumbersome- Straightforward* show clearly how the judgement on pragmatic qualities decreased over time (Fig. 12). At the end of the study, the product was considered non-practical to use and cumbersome. The assessment of the item *Impractical- Practical* changed significant since over four weeks the pragmatic aspects were evaluated in real contexts of use (e.g. battery drain affected the entire use of the smartphone, the step counter stops when the person receives a call), and compared with other products considered more effective.

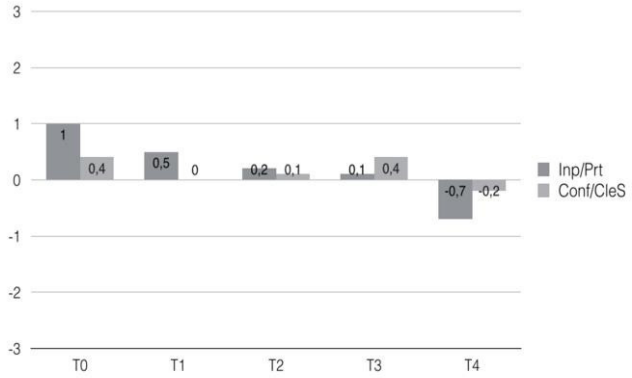


Fig. 12 PQ items *Impractical- Practical; Cumbersome- Straightforward* over four weeks

Also the items related to HQ-I: *Isolating- Connective; Alienating-Integrating; Separates me- Bring me close,* decreased over time (Fig. 13), even if the social features like the possibility to form or join groups and to create personal goals were generally appreciated in the diaries, and judged positively in the questionnaire on Anticipated and Remembered UX.

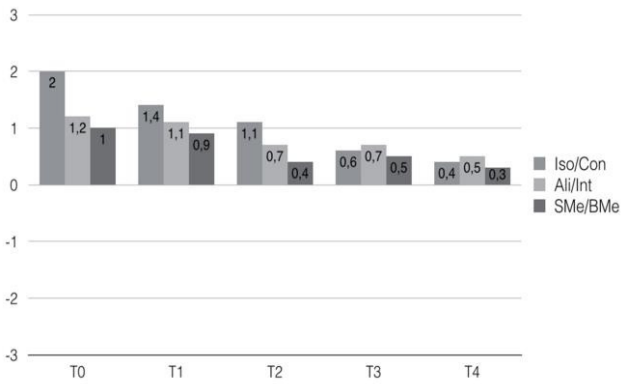


Fig. 13 HQ-I: *Isolating- Connective; Alienating-Integrating; Separates me- Bring me close* over four weeks

The app was unsuccessful in motivating participants. The items relating to the ability of the app to create enjoyable and captivating experience (HQ-S, ATT) decreased significantly over time (Fig. 14, Fig. 15). The following Fig. 14 reports the assessment of the HQ-S items: *Unimaginative-Creative; Dull-Captivating; Undemanding-Challenging; Ordinary-Novel;* and Fig. 15 the ATT items: *Unpleasant–Pleasant; Ugly-Attractive; Disagreeable-Likeable; Rejecting-Inviting; Bad-Good; Repelling-Appealing; Discouraging –Motivating.*

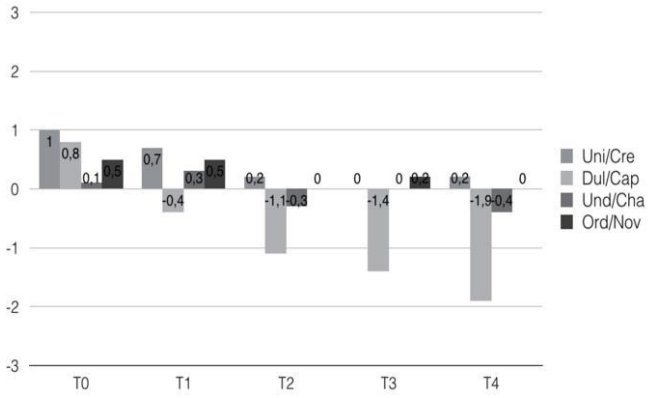


Fig. 14 HQ-S: *Unimaginative-Creative, Dull-Captivating, Undemanding-Challenging; Ordinary-Novel* over four weeks

TABLE I.
MEAN AND STANDARD DEVIATION OVER THE TIME

	PQ		HQ-I		HQ-S		ATT	
	Mean	***SD***	***Mean***	***SD***	***Mean***	***SD***	***Mean***	***SD***
T_0	0,90	1,31	1,09	1,59	0,56	1,39	1,40	0,89
T_1	0,54	1,03	0,86	1,54	0,40	1,31	0,87	0,96
T_2	0,51	1,33	0,53	1,35	-0,14	1,32	0,66	1,05
T_3	0,59	1,37	0,53	0,65	-0,14	1,60	0,37	1,17
T_4	0,26	1,57	0,33	1,40	-0,24	1,66	0,30	1,01

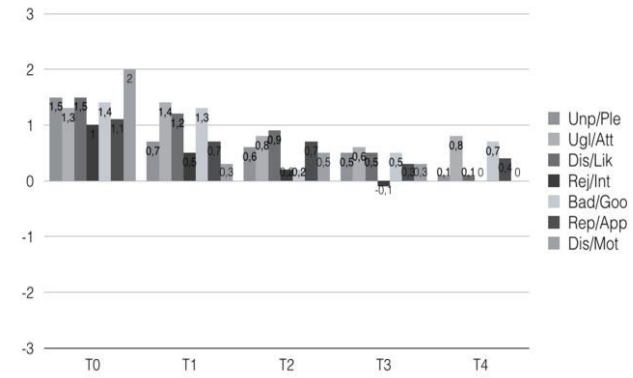


Fig. 15 ATT: *Unpleasant–Pleasant; Ugly-Attractive; Disagreeable-Likeable; Rejecting-Inviting; Bad-Good; Repelling-Appealing; Discouraging –Motivating* over four weeks

There were no differences between male and female participants (Fig. 16 and Fig. 17).

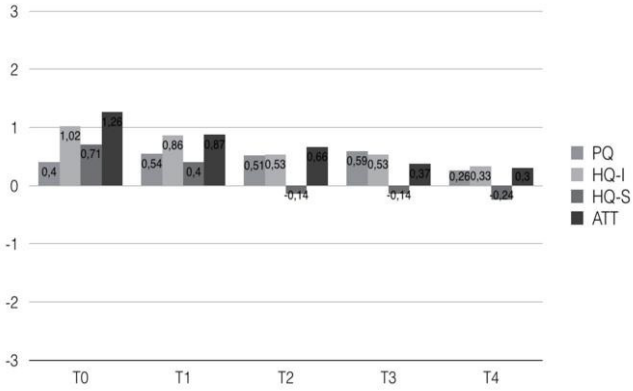


Fig. 16 Mean values for the four AttrakDiff dimensions over the time for male

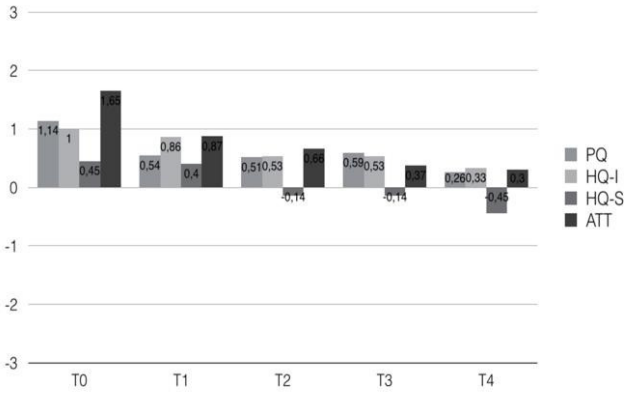


Fig.17 Mean values for the four AttrakDiff dimensions over the time for female

D) Remembered UX (AttrakDiff)

As said above, the remembered experience was assessed using an ad hoc questionnaire and AttrackDiff. The results of the ad hoc questionnaire are reported in section A above. The results of AttrakDiff related to the remembered experience were analysed conducting a paired-samples t-test to compare the four UX dimensions (PQ; HQ-I; HQ-S; ATT) at the first encounter (T_0) and after four weeks (T_4).

The test revealed that there were statistically significant differences between T_0 and T_4 on three dimensions HQ-I; HQ-S and ATT. The scores for HQ-I at T_0 (M=1,08, SD=0,62) was higher than the HQ-I at T_4 (M=0,32; SD=0,65) , t(9)=3,03; p=0,014. The score for the HQ-S at T_0 (M=0,55, SD=0,34) was higher than the HQ-S at T_4 (M=-0,24; SD=0,30) , t(9)=3,00; p=0,015. The score for the ATT at T_0 (M=1,40 , SD=0,19) was higher than the HQ-S at T_4 (M=-0,30; SD=0,179), t(9)=4,40; p=0,02.

On the contrary, there is no statistically significant difference for PQ.

Apparently after four weeks the participants remembered far better their dissatisfaction related to the hedonic qualities and the overall attractiveness of Pacer. These qualities were predominant with respect to the pragmatic qualities of the app. In fact other fitness app offer similar functionality, therefore the overall attractiveness makes the difference for a memorable UX. The participants clearly reported this in their qualitative comments.

The paired-samples t-test confirms that the user experience of using Pacer decreased over the time and the difference is statistically significant.

V. Discussion and Conclusions

The data obtained from the questionnaire on the Anticipated UX were consistent with those collected with AttrakDiff at T_1 (Episodic UX) and the diaries after the first encounter (Momentary UX). The same consistency can be noted between the data related to the Remembered UX and those collected with AttrakDiff at T_4.

Time seems to have an impact on the importance people attribute to different qualities of the experience with interactive products, as confirmed by previous studies [9]. Despite the crucial importance of usability in the product's initial acceptance, aspects of reliability, motivation, comparison with other products, change in behaviour and touch points (how the product communicates with the user, for example by notifications and alerts) are even more crucial for a user to resonate with a product and value it in the long term. That is why the UX evaluation in the long term is crucial.

Furthermore, even if it is not possible to proceed to general conclusions after a study involving a limited number (10) of subjects, the present study offers an original contribution that we hope could stimulate additional studies taking time systematically into account using different methods for the evaluation. Longitudinal studies on UX evaluation reported in literature assessed users' per-

ceptions focusing on specific times (Episodic UX) rather than assessing how their perceptions changed over time (Momentary and Episodic UX) and what memories people form in the long term that are crucial in stimulating later use. The diaries combined to questionnaires allowed us to reduce concerns about the reliability of the absolute measures collected with AttrakDiff where judgments were taken at predefine times and without reference points to single functionality. From one side diaries allowed us to assess the qualities of UX in context, that is in specific moments of interaction that were meaningful for participants. On the other side, the questionnaires on Anticipated and Remembered UX allowed us to associate the evaluation on expectations and memories to specific functionality of the product. The importance of such judgement was recognised also by Jordan and Persson [17] who suggested a hierarchical structure of qualities that contribute to positive experience, having the functionality of the product as a baseline. In addition to Jordan and Persson [17], we assumed the importance of UX qualities to vary with several personal and contextual factors including time as a fundamental source of diversity in UX, considered in its many facets Anticipated, Momentary, Episodic and Remembered.

I. References

[1] V. Roto, E. Law, A. Vermeeren, and J. Hoonhout, (eds), "User Experience White Paper. Outcome of the Dagstuhl Seminar on Demarcating User Experience", Germany, 2011. Available to the link http://www.allaboutux.org/uxwhitepaper

[2] D. Norman, "Emotional design: Why we love (or hate) everyday things", New York: Basic Books, 2004.

[3] B. L. Fredrickson, D. Kahneman, "Duration neglect in retrospective evaluations of affective episodes". Journal of Personality and Social Psychology 65 (1), 1993, pp. 45–55.

[4] D. Kahneman, "Evaluation by moments, past and future". In D. Kahneman, A. Tversky, "Choices, Values and Frames". Cambridge University Press. 2000, p. 693.

[5] D. A. Redelmeier, D. Kahneman, "Patients' memories of painful medical treatments: real-time and retrospective evaluations of two minimally invasive procedures". pain 66 (1), 1996, pp. 3–8.

[6] M. von Wilamowitz-Moellendorff, M. Hassenzahl, A. Platz, "Dynamics of user experience: How the perceived quality of mobile phones changes over time". In User Experience – Towards a unified view, Workshop at the 4th Nordic Conference on Human-Computer Interaction, 2006, pp 74-78.

[7] V. Mendoza, D. Novick, "Usability over time". In Proc. of the Special Interest Group on Design of Communication (SIGDOC), 2005, pp. 151-158.

[8] E. Karapanos, J. Zimmerman, J. Forlizzi, J. B. Martens, "Measuring the dynamics of remembered experience over time". Interacting with Computers, 22(5), 2010, pp. 328-335.

[9] E. Karapanos, J. Zimmerman, J. Forlizzi, J.B. Martens, "User experience over time: an initial framework". In Proc. of the 27th International Conference on Human Factors in Computing Systems. ACM, 2009, pp. 729–738.

[10] S. Mahlke, M. Thüring, M. "Studying Antecedents of Emotional Experiences in Interactive Contexts". In Proc. of the SIGCHI Conference on Human Factors in Computing Systems, New York: ACM Press, 2007, pp. 915-918.

[11] A. Vermeeren, E. Lai-Chong Law, V. Roto, M. Obrist, J. Hoonhaut, K. Väänänen-Vainio-Mattila, "User experience evaluation methods: current state and development needs". In Proc. of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries, 2010, pp. 521-530

[12] A. Fenko, H. N. J. Schifferstein, P. Hekkert, P., "Shifts in sensory dominance between various stages of user-product interactions". Applied Ergonomics 41, 2010, pp. 34–40.

[13] P. Marti, I. Iacono, "Experience over time: evaluating the experience of use of an interactive device on the short and medium term". International Journal on Multimedia Tools and Applications, in press, ISSN: 1380-7501.

[14] http://www.pacer.cc/ (last checked on May, 7th 2016)

[15] M. Hassenzahl, "The interplay of beauty, goodness, and usability in interactive products. Human-Computer Interaction, 19, 2004, pp 319–349.

[16] P. Marti, I. Iacono, "Evaluating the Experience of Use of a Squeezable Interface". In Proc. of s of the 11th Biannual ACM Conference on Italian SIGCHI Chapter, Rome, Italy 28-30 September 2015. ISBN: 978-1-4503-3684-0, 2015, pp 42-49.

[17] P. W. Jordan, S. Persson, "Exploring users' product constructs: how people think about different types of product". International Journal of CoCreation in Design and the Arts, 3. 2007, pp. 97-106.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1657–1661
DOI: 10.15439/2016F434
ACSIS, Vol. 8. ISSN 2300-5963

Designing effective educational games - a case study of a project management game

Jakub Miler, Agnieszka Landowska
Gdansk University of Technology,
Faculty of Electronics, Telecommunications and Informatics,
Narutowicza Street 11/12, 80-233, Gdansk, Poland
Email: {jakubm, nailie}@eti.pg.gda.pl

***Abstract*—This paper addresses the issues of designing effective educational games. We aim at investigating how the cognitive, behavioral and emotional aspects of the games influence their educational effectiveness. The results were obtained with an observational user experience study extended with affect analysis carried out for a project management game GraPM. We analyzed the players' understanding of the game mechanics and logic, their engagement and emotional state. Then we confronted it with the educational effects achieved. In this case study the key identified factors of educational effectiveness were: understanding game mechanics, player's engagement, and feeling of control. Invoking other desired emotions was not required for effective education, which was also generally unrelated to the player's initial knowledge.**

I. Introduction

EDUCATION has always been an important part of human's activity. Effective education requires satisfaction of multiple goals according to Bloom's taxonomy and its newer revisions [1]. One of the recognized tools to provide multi-layer educational experience are games, with computer games in particular. The use of games in education is additionally promoted by the "serious games" approach [2].

Playability is a commonly used term for computer games, however it can be evaluated the same way as usability [3]. Usability of software as defined by ISO 9241 standard includes effectiveness, efficiency, and satisfaction with which specified users achieve specified goals in particular environments [4]. User experience (UX) extends traditional usability with affective aspects forming more holistic picture of human-system interaction.

The evaluation of user experience is particularly important for educational games as the cognitive, behavioral and emotional aspects of a game can largely influence its educational effectiveness. Our research goal is to understand the factors that influence the educational effectiveness of computer games and utilize it to design better educational games. This leads to the following research questions of this paper: (RQ1) how usability of a game affects its educational effects? (RQ2) how user experience of a game affects its educational effects?

This paper presents an extended user experience study of an educational game GraPM and is organized as follows. Section II discusses the related work. Section III describes briefly the game under study and its usability goals. Sections IV and V provide the study plan and results, respectively, followed by the concluding remarks in section VI.

II. Related Work

The related work for our study can be divided into two main topics: (1) educational games; (2) usability and user experience research.

We are mostly interested in the management games in general and project management games in particular. The example project management training games are Symulator Projektu and Symulator Zachowań Menadżerskich from Grupa ODiTK [5], That Project Management Game [6], or Scrum Game [7]. No usability, user experience or educational effectiveness studies for most of these games were identified. Scrum Game has been evaluated preliminarily, but only with a simple questionnaire.

Broad discussions of the issues of game usability are presented in [8], but they do not address specifically the educational games. P. Mirza-babaei introduced biometrics to evaluate the gameplay experience, but applied it to a non-educational FPS-type game [9]. Parodi *et al.* measured the player's education with their Competence Performance Analyser tool with limited use of UX and biometrics [10]. Raabe, Santos, Paludo, and Benitti carried out an extensive evaluation of serious games teaching project management, however the evaluation did not include the usability [11].

Research on both user experience and affective computing is broad and have already been summarized several times, e.g. in the work of Vermeen et al. [12] or the book by Albert and Tullis [13]. There are a few studies on fusing affect recognition and usability evaluation [14]–[16]. The most important work is the one by Ahn and Picard, that proposed the Affective-Behavioral-Cognitive (ABC) framework for

This work was supported in part by Polish-Norwegian Financial Mechanism Small Grant Scheme under the contract no. Pol-Nor/209260/108/2015 as well as by DS Funds of ETI Faculty, Gdansk University of Technology.

the user experience evaluation. The framework was validated with an experiment on beverages [14]. Lew, Olsina, Becker, and Zhang applied the affect evaluation in the quality assurance procedures for web applications [15]. Kołakowska, Landowska, Szwoch, Szwoch, and Wróbel proposed the application of affect recognition in usability evaluation in four different scenarios: first impression test, task-based usability test, free interaction test and comparative test [16].

III. The Game under Study

We carried out this research with a project management educational game GraPM, which was conceived and designed by J. Miler based on his earlier research on project management [17]–[18]. GraPM stands for "game on project management", where "gra" is game in Polish. This game puts the player in the role of a project manager. The target group of this game includes: (1) people who want to increase their project management knowledge and skills; (2) players who like management and strategy games.

The GraPM game was implemented within an engineering diploma project in 2014 [19]. It is a JavaScript-based rich internet application run in a web browser, see Fig. 1.

Fig. 1 The GraPM game user interface mid-game

The user experience goals for the GraPM game assume that the player understands the game mechanics after the 2nd gameplay and the game logic after the 5th gameplay. The player's engagement should be maintained high across multiple gameplays to ensure game replayability.

The game is also expected to generate the following emotions desired for educational effectiveness: (a) interest – eagerness to learn; (b) slight confusion – feeling there is something to learn; (c) joy – satisfaction from learning and playing better; (d) feeling of control – the confidence of fully controlling the in-game project and winning the game. On the other hand, the following emotions are undesired for the educational effectiveness and should be avoided in the gameplay: (a) fear – being afraid of controlling the game and learning; (b) strong confusion (frustration) – feeling lost due to not knowing how to play the game; (c) anger – irritation due to not understanding the game and inability to win; (d) boredom – loss of motivation to play, learn and win; (e) disregard – considering the game poor, unchallenging and non-educating.

IV. User Experience Study Design and Execution

A. UX Study Design

In the study of the user experience of GraPM, we have used typical scenario-based procedure for usability testing. The procedure was performed and analyzed using the Affective-Behavioral-Cognitive (ABC) approach, that combines participant perception (cognitive component retrieved by self-report) with observational scores (behavioral component) and emotion analysis techniques (affective component). The study was a part of an experiment that aimed at evaluation of applicability of automatic emotion recognition in the context of usability testing [20]. It was performed at the Emotion Monitor stand, which is a multi-modal setting designed for human-computer interaction observation [21].

An initial questionnaire was filled before playing the game. It covered: sex, age, year and field of study, knowledge of the GraPM game or other project management games, participation in project management courses and evaluation of initial project management knowledge (IPMK).

The gameplay (which was performed 5 times) was intertwined with questionnaires that included: subjective competence progress evaluation and self-report on emotions. The questionnaires covered cognitive component of the ABC approach. In order to perform behavioral analysis, the Morae Recoder software was used, that allowed to capture screen and user mouse/keyboard activities.

Affective component used in this study was based on three input channels: video capturing for facial expressions analysis (with Logitech Internet camera and compatible software solution), self-report based on PAD (Pleasure-Arousal-Dominance) emotion representation model [22] and physiological signals recording as a reference (skin conductance, blood-volume pulse and respiration rate recorded with ProComp Infiniti hardware coder together with Biograph software).

B. UX Evaluation Criteria

We have applied three general UX evaluation criteria: understanding, engagement and enjoyment, as well as purposely added educational effectiveness. All evaluation criteria were further operationalized into metrics.

Understanding covers the cognitive aspect of user experience and describes the comprehension of the game by the player on the level of its mechanics and its internal logic. The understanding of game mechanics means that the player is able to manipulate the game and affect the gameplay. The understanding of game logic means that the player learned the rules of the game and is able to win. According to the user experience goals of GraPM, the understanding of the

game mechanics should be evaluated after the 2nd gameplay, while the understanding of the game logic should be evaluated after the 5th gameplay.

Engagement is related to the behavioral aspect of the player-game interaction and characterizes the ability of the game to attract the player, maintain his interest and motivate him to play, learn and win. It should be evaluated over several gameplays.

Enjoyment covers the emotional aspect of the user experience and describes to what extend the game is able to stimulate desired emotions and avoid undesired emotions, as defined by the game designer. It should also be evaluated over multiple gameplays.

The educational effectiveness corresponds to the main business goal of the GraPM game. The other factors of user experience form, in our opinion, an indispensable foundation for the educational effects to emerge. The educational effects should be evaluated after the 5th gameplay.

C. Operationalization of UX Evaluation Criteria

The understanding factor was operationalized with two metrics: understanding of game mechanics (UM) and understanding of game logic (UL). UM measured the number of mechanics understood by the player out of 15 mechanics. UL measured the number of game logic rules understood out of 10 rules in total. The values of these metrics were measured by the game designer based on the recorded gameplay.

The engagement factor was operationalized with the following metrics: (EN1) number of pauses to adjust the project; (EN2) percentage of risks to which player reacted in any way; (EN3) number of threats eliminated (reduced to zero); (EN4) number of opportunities enhanced or exploited (materialized); (EN5) number of tooltips read; (EN6) number of mouse clicks (left button only); (EN7) distance of mouse movement (in thousands of pixels); (EN8) gameplay duration (in minutes); (EN9) summary evaluation of engagement based on EN1-EN8 in a 5-point Likert-type scale (very low VL, low L, medium M, high H, very high VH). The EN1-EN8 metrics were measured in an open scale of intensity. Metrics EN1-EN5 were measured by the game designer based on the recorded gameplay. Metrics EN6-EN8 were measured automatically by Morae Manager tool. EN9 was evaluated by the game designer.

Enjoyment factor was operationalized with the following metrics: (EJ1) average level of valence per user, representing positive vs. negative dimension; (EJ2) average level of arousal per user, representing calm vs. energetic dimension; (EJ3) compatibility of the EJ1 and EJ2 with desired or undesired emotional states in a closed scale (D – desired emotional state; UD – undesired emotional state, O – other emotional state than specified); (EJ4) dominance after 5th gameplay, which is the proxy for the feeling of control.

Educational effectiveness was operationalized with the following metrics: (ED1) self-reported degree of improvement in the project management competencies in 5-point Likert-type scale, from 1 - “very low” to 5 - “very high”; (ED2) number of additional unique project management aspects listed by the player after 5 gameplays as compared to the initial questionnaire.

D. UX Study Execution

The study was carried out in April and May 2016. The entire experiment involved 10 participants aged 23 to 43 (8 of them belonged to the game target group), 5 male and 5 female. 5 participants were selected for the user experience study based on the following criteria: (1) belonging to the target group; (2) no prior playing GraPM; (3) diverse fields of study; (4) inclusion of both male and female; (5) diverse levels of initial project management knowledge. In the entire experiment we recorded 50 gameplays, 5 for each participant. The time of a single participant recording (5 gameplays and the questionnaires) varied from 34 up to over 90 minutes. The players in the study are coded P1 to P5.

V. User Experience Evaluation Results

A. Understanding

The understanding of the game mechanics and logic by the sample players are presented in Table I. It can be observed that, apart from player P1, the players understood at least 80% of the game mechanics only after the 2nd gameplay. The results also show that 2 out of 5 players (P1 and P4) understood at most 50% of the game logic, while the other 3 at least 80%. The players P1 and P4 were not able to win the game as they had considerable usability problems only at the game mechanics level. Only 2 players (P3 and P5) could achieve full project success and win the game.

TABLE I.
EVALUATION OF UNDERSTANDING

Metric	P1	P2	P3	P4	P5	Avg.
UM	8	13	15	12	13	12,2 (2,6)
UL	5	8	9	4	9	7 (2,3)

B. Engagement

The results of the engagement evaluation are presented in Table II. The values of metrics EN1 to EN5 were averaged over the 2nd and the 5th gameplay, while the values of metrics EN6 to EN8 were averaged over all 5 gameplays of a particular player.

Interpreting these results, we can see that the least engaged player was P4. P4 influenced the project course to a very limited extent (EN1), generally ignored the risks (EN2–EN4), clicked and moved the mouse little (EN6, EN7) as well as played for the shortest time (EN8). The most engaged player was P3. P3 adjusted the project many times (EN1), managed successfully the risks (EN2–EN4), read many tooltips (EN5), intensely clicked and moved the mouse (EN6, EN7), and played for the longest time (EN8). The

engagement of P2 was also very high, similar to P3. The P4's engagement was high, close to the average. The anomalous case is the player P1. P1 did not understand how to pause the game (EN1) and react to risks (EN3 and EN4) on the game mechanics level (compare to Table I), thus the 0 values of these metrics. However, P1's engagement was at least medium, which can be evaluated from the metrics EN2, and EN6 to EN8. These metrics show that P1 wanted to control the project and manage the risks, but did not know how to manipulate the game to do this. These interpretations are reflected in the values assigned to the metric EN9.

TABLE II.
EVALUATION OF ENGAGEMENT

Metric	P1	P2	P3	P4	P5	Avg.
EN1	0	12,5	25,5	1	14	10,6 (10,5)
EN2	80%	100%	100%	42%	72%	78,7% (24%)
EN3	0	3,5	3,5	0,5	2	1,9 (1,6)
EN4	0	1,5	1,5	0	1	0,8 (0,8)
EN5	0	0	7	4	0	2,2 (3,2)
EN6	50 (14,9)	87 (6,2)	108 (31,3)	17 (5,0)	73 (11,8)	67,2 (34,9)
EN7	59,0 (16,9)	97,5 (21,4)	90,8 (18,0)	30,7 (14,1)	75,8 (8,6)	70,8 (26,9)
EN8	5:44 (1:26)	6:32 (1:26)	8:32 (2:40)	4:29 (0:38)	6:24 (0:35)	6:20 (1:28)
EN9	M	VH	VH	L	H	H

C. Enjoyment

The evaluation of enjoyment in 5 gameplays for each player is presented in Table III and Fig. 2.

TABLE III.
EVALUATION OF ENJOYMENT

Metric	P1	P2	P3	P4	P5
EJ1	3,6 (0,9)	6,2 (0,4)	5 (0)	5,4 (1,3)	6,4 (0,5)
EJ2	3,8 (0,4)	5,6 (1,1)	1 (0)	5,2 (0,8)	2,4 (1,7)
EJ3	D	D	O	D	O
EJ4	5	5	5	3	7

Desired emotional states region contains majority of the 1st quarter together with the states of neutral arousal from the negative region. Out of the four specified desired emotional states (section III), two are not expressible in the valence-arousal model: interest as a cognitive state and a feeling of control, which is expressible with the third dimension of PAD model – dominance (EJ4). Undesired emotional states were clustered as two regions: negative states of very high arousal and negative states of very low arousal. Please note, that fear, anger and frustration are very close to each other in this model.

The players P1, P2 and P4 fall into the desired emotional states region. The players P3 and P5 do not fall into the desired nor un-desired region; it seems, that for them the game was a relaxing experience.

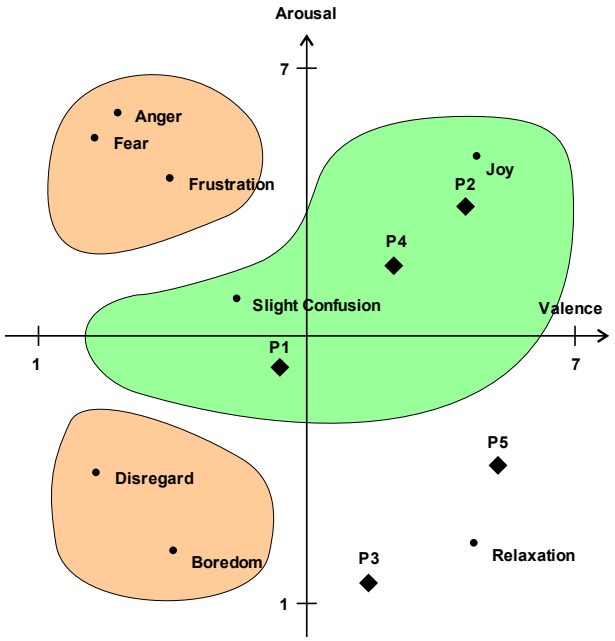


Fig. 2 Players' emotional states compared to the desired and undesired emotional state regions in the valence-arousal model

D. Educational Effectiveness

The evaluation of the educational effectiveness after the 5th gameplay is presented in Table IV. The values of metrics ED1 and ED2 are confronted with the initial player's knowledge of project management before playing the GraPM game (IPMK). It can be observed that the players P2 and P5 reported the highest educational effectiveness, however in case of P5 the high self-reported estimate (ED1) was not confirmed with the knowledge verification (ED2). P3 reported low estimate of education (ED1), but still exhibited considerable increase in knowledge (ED2). The lowest educational effectiveness was observed for P1 and P4.

TABLE IV.
EVALUATION OF EDUCATIONAL EFFECTIVENESS

Metric	P1	P2	P3	P4	P5	Avg.
ED1	2	4	2	1	4	2,6 (1,3)
ED2	1	3	3	2	1	2 (1,0)
IPMK	3	4	5	2	1	3 (1,6)

E. Summary and Discussion of Results

This study leads to the following observations:

- the understanding of the game mechanics forms the basis for understanding of the game logic;
- the engagement seems unrelated to understanding of the game mechanics, which can be seen by cases of player P1 (at least medium engagement with low understanding of the mechanics) and player P4 (low engagement with high understanding of the mechanics);
- the engagement is related to understanding of the game logic, however the nature of this relation is still to be investigated;

- the educational effectiveness is enhanced by the player's understanding of the game mechanics and logic, as well as his engagement (P2, P3, and P5);
- the educational effectiveness seems unrelated to the initial knowledge of the player: players P2 and P3 reported the highest initial knowledge but still exhibited high educational effects; player P5 reported the lowest initial knowledge and high educational effectiveness; players P1 and P4 reported medium initial knowledge and medium educational effects;
- the educational effectiveness was not disturbed with undesired emotions for any player, however 2 out of 3 players (P3 and P5) exhibited rather ambivalent emotions regarding the game goals (relaxation);
- the educational effectiveness is related to the player's feeling of control: the player P4 with the lowest educational effects also had the lowest feeling of control, while the player P5 with high educational effects exhibited the highest feeling of control.

We are aware of the fact that the validity of this study has some limitations. We identified and addressed the following threats to its validity: (1) sample size – we engaged 5 users as the usability tests show that 5 users reveal 75-90% of the usability issues; (2) sample as a group of convenience – we introduced the initial questionnaire and selected the sample for UX evaluation to ensure its diversity; (3) confounding variables – we performed the study in a strictly controlled environment, where we limited the possible influences of external factors; (4) subjective measurements – we operationalized most of the variables to objective metrics, the number of subjective self-reports is minimal; (5) small number of gameplays – due to resource limitations in manual measurements we focused on 2nd and 5th gameplay, which were specified in the UX design goals.

VI. Conclusion

The study provides preliminary evidence that usability and user experience affect the educational effectiveness of a computer game. The results revealed the flaws in the GraPM game UX design at the cognitive and behavioral levels. Together with the analysis of the emotional impact it may be used to improve the game and increase its effectiveness.

Based on the data collected we plan to evaluate the GraPM learning curve based on the understanding of logic across gameplays, analyze players' emotional state in consecutive gameplays as well as identify affects related to particular game events.

Acknowledgment

J. Miler thanks the engineering diploma project team that implemented the GraPM game: Rafał Piechowski, Damian Płatek, Anna Wasik, and Sylwia Grabowska. The authors thank Dominika Makowiecka, who helped in the experiment setting and execution.

References

[1] L. W. Anderson, D. R. Krathwohl, P. W. Airasian, K. A. Cruikshank, R. E. Mayer, P. R. Pintrich, J. Raths, and M. C. Wittrock, *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives, Abridged Edition*, Pearson, 2000

[2] C. Abt, *Serious Games*, University Press of America, 2002

[3] D. Novick, J. Vicario, B. Santaella, I. Gris, "Empirical analysis of playability vs. Usability in a computer game," LNCS vol. 8518, PART 2, Springer Verlag, 2014

[4] ISO 9241-210:2010 Ergonomics of human-system interaction -- Part 210: Human-centred design for interactive systems, ISO, 2010.

[5] ODiTK Symulator Biznesu Sp. z o.o., http://www.symulator.oditk.pl/ [retrieved May 2016]

[6] R. Charney, That Project Management Game, http://thatpmgame.com/ [retrieved May 2016]

[7] A. Gkritsi, "Scrum Game: An Agile Software Management Game," M.S. thesis, School of Electronics and Computer Science, Univ. of Southampton, Southampton, UK, 2011.

[8] K. Isbister, and N. Schaffer, *Game Usability: Advancing the Player Experience*, CRC Press, 2008

[9] P. Mirza-babaei, "Biometrics to improve methodologies on understanding player's gameplay experience," in Proc. *25th BCS Conference on Human-Computer Interaction*, Swinton, UK, 2011, pp. 546–549.

[10] E. Parodi, M. A. Bedek, P. Seitlinger, M. Vannucci, C. Jennett, M. Ruskov, and J. M. Celdran, "Analysing players' performance in serious games", *International Journal of Technology Enhanced Learning (IJTEL)*, vol. 6, no. 3, 2014, pp. 237–248.

[11] A. Raabe, E. Santos, L. Paludo, and F. Benitti, "Serious Games Applied to Project Management Teaching,", in Handbook of Research on Serious Games as Educational, Business and Research Tools, IGI Global, Hershey, PA, 2012, pp. 668–692.

[12] A. P. Vermeeren, E. L. C. Law, V. Roto, M. Obrist, J. Hoonhout, and K. Väänänen-Vainio-Mattila, "User experience evaluation methods: current state and development needs," in Proc. *6th Nordic Conference on Human-Computer Interaction: Extending Boundaries*, ACM, 2010, pp. 521-530.

[13] W. Albert and T. Tullis, *Measuring the user experience: collecting, analyzing, and presenting usability metrics*, Newnes, 2013

[14] H. I. Ahn and R. Picard, "Measuring Affective-Cognitive Experience and Predicting Market Success", *IEEE Transactions on Affective Computing*, vol. 5, no. 2, 2014

[15] P. Lew, L. Olsina, P. Becker, and L. Zhang, "An integrated strategy to systematically understand and manage quality in use for web applications," *Requirements Engineering*, vol. 17, no. 4, 2012, pp. 299-330.

[16] A. Kołakowska, A. Landowska, M. Szwoch, W. Szwoch, M. Wróbel, "Emotion recognition and its applications", *Human-Computer Systems Interaction: Backgrounds and Applications 3*, Springer, 2014, pp. 51-62.

[17] J. Miler and H. Wesołowska, "Improvement of Task Management with Process Models in Small and Medium Software Companies," in Proc. 19th EuroSPI Conference, CCIS Systems, Software and Services Process Improvement, vol. 301, Springer, 2012, pp. 145-156.

[18] J. Miler, "A Method of Software Project Risk Identification and Analysis," Ph. D. thesis, Faculty of Electronics, Telecommunications and Informatics, Gdansk Univ. of Techn., Gdansk, Poland, 2005.

[19] R. Piechowski, D. Płatek, A. Wasik, S. Grabowska, "Educational game on project management", B.Sc. thesis, Faculty of Electronics, Telecommunications and Informatics, Gdansk Univ. of Techn., 2014.

[20] A. Landowska, J. Miler, "Limitations of Emotion Recognition in Software User Experience Evaluation Context," in Proc. of FedCSIS, 2016, accepted for publication

[21] A. Landowska, "Emotion monitor-concept, construction and lessons learned," in Proc. of Federated Conference on Computer Science and Information Systems (FedCSIS), 2015, pp.75-80.

[22] A. Kołakowska, A. Landowska, M. Szwoch, W. Szwoch, M. R. Wróbel, "Modeling emotions for affect-aware applications," in Wrycza S. Information Systems Development and Applications, 2015, pp. 55-64.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1663–1670
DOI: 10.15439/2016F504
ACSIS, Vol. 8. ISSN 2300-5963

Peepdeck: a dashboard for the distributed design studio

Jesús Muñoz-Alcántara
Eindhoven University of Technology
De Rondom 70, 5612AP,
Eindhoven, The Netherlands
Email: j.munoz.alcantara@tue.nl

Petr Kosnar,
Eindhoven University of Technology
De Rondom 70, 5612AP,
Eindhoven, The Netherlands
Email: hello@iampetr.com

Mathias Funk, Panos Markopoulos,
Eindhoven University of Technology
De Rondom 70, 5612AP,
Eindhoven, The Netherlands
Email: {m.funk, p.markopoulos @tue.nl}

***Abstract*—Designers adopt a large amount of general-purpose tools for supporting their remote collaborative tasks. Each tool provides very diverse functionalities: from file sharing to instant communication and video collaboration. The designer struggles when filtering and combining the right information spread across the multitude of tools. This research extends McGrath's framework of task circumflex to map the collaborative demands of the design practitioner and proposes Peepdeck, a design exploration to support them. Peepdeck is a dashboard that assembles information scattered across multiple tools in a personalized and organized way. Through two design iterations followed by evaluations of the user interface, several requirements were identified for supporting collaboration awareness in design teams. Insights confirmed the relevance of combining information from different but already familiar tools, rather than attempting to replace them. It was identified the importance of optimizing for visual scanning, supporting search of content and allowing users to customize the tool.**

I. Introduction

Design studios operate in a much more distributed fashion nowadays than in the past. Designers may be distributed geographically, potentially even in different time zones. This paper discusses the design of an application for supporting distributed collaborative design work. There is a myriad of new tools and online services available as shown by *Fig 1*. Tools like Dropbox, GoogleApps, Basecamp, Atlassian, Slack, Skype and Trello are just some examples of the diversity of commercial applications available on the Internet that support teamwork in the design studio. These technologies support very diverse functionalities and services, from file sharing, online edition of documents to instant communication and video collaboration.

Previous research [13] revealed a collection of patterns of behavior that the designers conduct in the context of collaboration tools. A surprising finding was that designers adopt a large amount of generic tools that they appropriate for supporting collaborative design tasks, choosing different tools for different parts of the design process. Tools like Facebook (social media), Dropbox (file sharing) and Skype (instant communication) were some of the most popular.

While some of the tools are used throughout the whole design process, each of the categories of behavior imposes its own challenges to each of the tools. Individuals switch fluently from one tool to another depending on the activity, the personal needs, the project needs and how the tool covers those needs. The designer must be able to filter and combine the right information spread across the given multitude of tools, each with different information (file, contact, folder, application) and user interfaces. This is the starting point that motivates the present work in order to facilitate the current collaborative practices of designers.

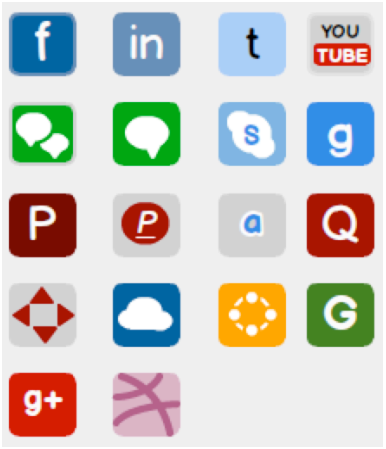


Fig 1. Set of icons representing the diversity of commercial applications available on the Internet. Original icons by Designyantra used under Creative Commons License.

The current research presents a design exploration into how collaboration practices for distributed design teams can be supported. The paper starts by taking as reference McGrath's framework of task circumflex [12] to analyze how the current collaboration tools are incapable of covering the demands of the design practitioner. Then, the paper proposes as a solution an application called Peepdeck, which assemble information scattered across multiple tools in a personalized and organized way. Afterwards, the two iterations of the design and evaluation of the user interface of the application are presented. Finally, the paper discusses a reflection to the proposed design, the design requirements

This work was funded by the COnCEPT project as part of the European Commission 7th Framework ICT Research Programme: Project Number 610725. For further details visit: http://www.concept-fp7.eu

and further issues pertaining the design of such groupware applications.

II. Fitting Technology in Practice

The vast amount of existing research in the field of CSCW [21] has illustrated how groups adopted IT tools and integrated them into their social dynamics to support teamwork. Researchers have identified specific patterns of behavior supported by the appropriation of collaboration tools [10; 17; 18; 19] and how those patterns changed through their appropriation [24].

McGrath [12] defined a framework, called task circumflex, to classify technology according to the type of task that they were supporting within the team. McGrath classified collaboration tools in 4 quadrants based on whether they supported (1) generation activities: for planning and idea creation, (2) choice activities: for problem solving and decision-making, (3) execution activities: for the task execution and performance, or (4) negotiation activities: for tasks that focus in resolving conflicts between the individuals (*Fig 2*).

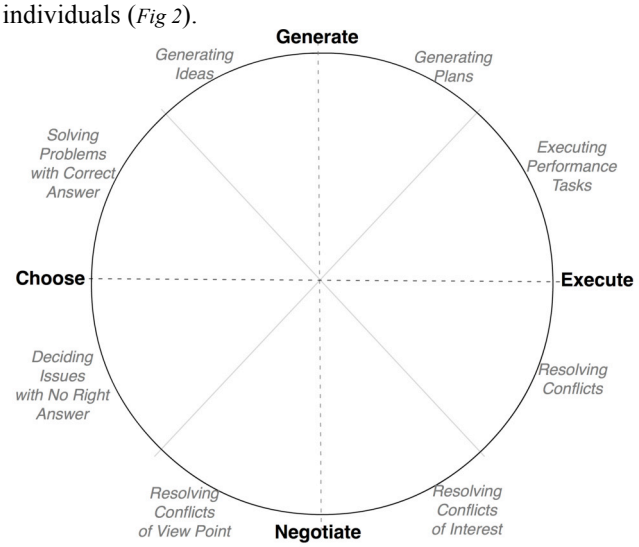


Fig 2. The task circumflex defined by McGrath (1984)

In the context of remote design teams, Muñoz-Alcántara and colleagues [13] performed a series of interviews and surveys to professional designers in order to understand the activities that designers engage to support their collaborative practices. Their results described 5 categories of collaborative activities: (1) creating ideas and concepts (e.g., ideation, brainstorming and inspiration), (2) developing ideas and concepts (e.g., sketching and prototyping), (3) making sense of the material, resources, and experiential knowledge (e.g., giving and receiving feedback), (4) keeping the team on track (e.g., notifying the team and solving issues in the team), and (5) managing the development of the project (e.g., defining and managing tasks and deadlines). As shown in *Fig 3*, the first 3 categories described the core activities of the design practice, while the last 2, described the social dynamics that enable the completion of the first set of processes.

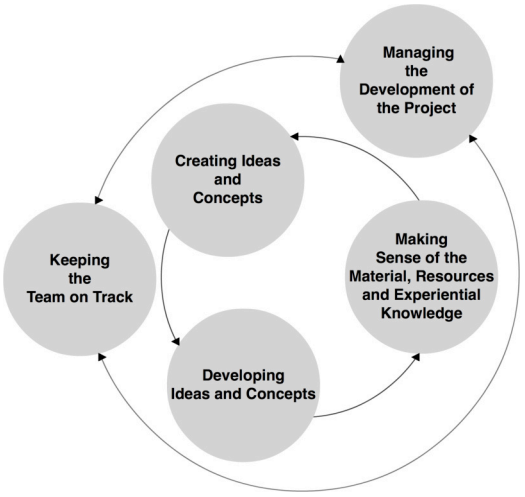


Fig 3. This diagram shows the flow of the five categories of collaborative activities described by Muñoz-Alcántara et al. (2015)

Additionally, Muñoz-Alcántara et al. [13] compiled a list of specific tools that designers use during each of the activities, as shown in Table I. The list reveals that professional designers use few specialized tools compared to the total amount of tools involved during the creative stages of their work. Furthermore, Table I illustrates how various tools appear repeatedly in more than one of the activities. This clearly shows that a large collection of different tools supports the designers on more than one of their daily tasks.

Table I. List of tools used on each design activity

Activity	Tools involved
Creating ideas and concepts	Google Docs, Google Spreadsheets, Mindmaps, Skype, Evernote, Dropbox, Pinterest, Facebook
Developing ideas and concepts	Paper, iPad (Adobe Ideas, Paper 53), Axure, Excel, Illustrator, SVN, Github, Bitbucket, Dropbox, Google Docs
Making sense of the material, etc.	Skype, Evernote, Facebook Groups, Facebook Chat, Lynk, email and Google Hangouts, Email, WeTransfer, Dropbox, Axure, Interactive documents
Keeping the team on track	Facebook Group, Facebook Chat, Skype, email, Whatsapp
Managing the development of the project	Redmine, Gantt charts, Teambox, spreadsheets, Trello, Outlook, Google Docs, Dropbox, Google Drive, Facebook Groups

Each of these collaborative processes can be mapped into the classification given by McGrath in order to reflect how each tool is used on every stage of the creative design process. The core process of creating ideas and concepts belongs to the quadrants of generation support tools and choice support tools. Tools aimed for generation and for execution activities mainly support the process of developing ideas and concepts. The process of making sense of the material is mainly covered by the quadrant of tools for choice but it also includes tools focused on negotiation. Keeping the team on track is primarily achieved by the support of negotiation tools while managing the development of the project depends on generation and

execution means. Fig 4 displays the task classification defined by McGrath combined with the group activities described by Muñoz-Alcántara and colleagues. These sets of behaviors provide the starting point for understanding how each different activity provides a specific value to the effort of the design team.

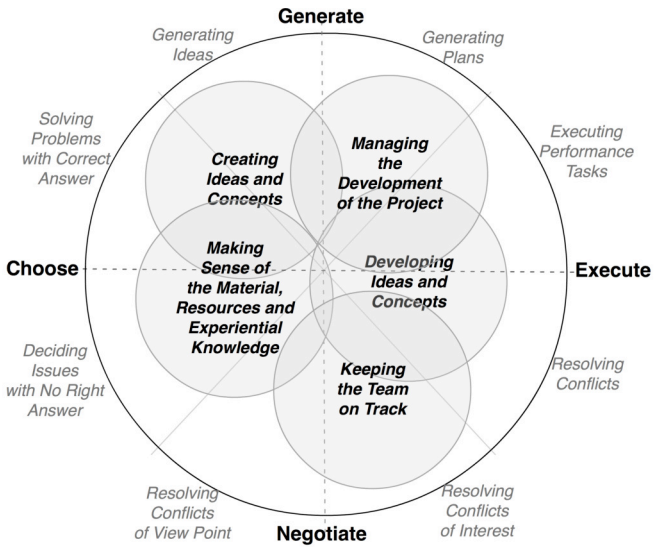


Fig 4. This diagram maps the task circumflex of McGrath (1984) with the specific patterns of behavior of a designer described by Muñoz-Alcántara et al. (2015).

Traditionally, design studios have a high material character. For example, office walls and desks are typically full of notes, post-it, sketches, prototypes and physical models. The material and physical aspects embedded in the design studio such as designer's practices, the use of artifacts and the space during collaboration, have a key role on the coordination of the creative activities [23]. Design artifacts (e.g. sketches, physical prototypes) located in the studio inform activity and progress of the group and trigger awareness in the team [2, 6]. The use of space is also productivity-focused, displaying information that supports time management, scheduling and division of workload. The designer takes advantage of the space and the artifacts to enable the coordination with the team. The design process often demands fast switching between activities, and thus, between artifacts and tools. Most activities require the selection and combination of information spread across several sources. Moreover, every person requires a unique set of information based on the roles, responsibilities, and also personal preferences and interest in the projects they are working on.

However, on a distributed setting, the digital information is usually managed and visualized through very different interfaces and organized in different information architecture. Since general-purpose tools are generally adopted to support certain design activities, extra work is needed for annotations, documentation, and organizing files and folders on each of the tools. As a consequence designers fail to integrate and coordinate the information, effort and outcomes of other team member's activities. Furthermore, when using a large collection of tools it is difficult to have an overview of the design process and the performance of the individual is affected by the huge amount of information provided by the collection of tools.

The question raised is how a groupware application could support design collaboration by focusing on the integration of the current tools while addressing the core requirements of the design tasks.

The design of the user interface of the dashboard followed a user-centered approach with two iterative cycles. Each design iteration consisted of the following steps: specification finding, conceptualization, development of the concept through prototyping and the evaluation of the prototype. The first iteration focused on creating and evaluating the concept of the dashboard. Drawing from the problems we identified during the literature research, this iteration explored what is the meaningful content of such a tool, what is the way how designers can use it, its information architecture and how they can incorporate it into their typical work routine. This was first evaluated on a paper prototype. In the second iteration an interactive digital prototype was developed and its UI was evaluated in terms of interactions, ease of use, and clarity of the application. The final outcome of this work is to bridge the gap between the existing qualitative fieldwork studies and the design recommendations based on the actual design and evaluation of a tool that enables remote collaboration in the design practice.

III. Initial Concept

The insights gathered from the literature review suggested to focus on solving the problem of scattered information and losing oversight rather than creating yet another specialized tool that solves narrowly scoped problems for designers. The concept of a dashboard emerged. The dashboard aggregates information from (multiple) existing tools and services, and displays them in a personalized manner at one place. This approach enables keeping a single UI for the user regardless of the actual tools and data sources integrated on a lower level. As one of the consequences, exchanging of the backend infrastructure or the connected tools does not need to be noticeable by the users.

IV. First Design Iteration

I. Design specifications

The success of the dashboard mainly depends on addressing the challenge of accessibility of the information and implementation complexity. A number of design challenges were identified on which the success of the dashboard depends:

- Making a glanceable display [11] that collects salient information for design collaboration in one place?
- What is the right amount of information? [4] (Lack of awareness vs. Information overload)

- How to design simple and intuitive interface? (Minimizing learning curve [15; 16])
- Information visualisation: overview / information display (Optimized for visual searching [8; 9])

The dashboard should incorporate the existing tools and provide an extra layer where designers can access the selected content and information streams, and filter them as (individually) preferred. Finally, it should also enable users to stay updated by checking only a single place (tool), instead of multiple different tools and their different notification areas, status bars, streams, feeds, and other elements meant for updating the user.

II. Peepdeck concept

The Peepdeck concept was conceived to address the challenges identified above. Peepdeck is envisioned as an online service that connects existing tools and services that users already use for collaborative design; it aggregates information from the connected services, and classifies it into four main categories resolved from the common groups of items described in the enabling activities discussed above: tasks management, shared calendar items, communication streams, and cloud storage with file management facilities. These categories contain information merged from various sources in a way that the information does not define the attributes or features of the item (i.e., a file from Dropbox is treated equally as a file from OneDrive, or Google Drive).

The central goal of Peepdeck is to display the information that is currently relevant to the user – reflecting their role, context, projects they are involved in, co-workers, personal preferences, and other aspects. The user can adjust the content of the dashboard so it matches his requirements. The dashboard contains shortcut to the connected services – so the user does not replace them with the Peepdeck, but Peepdeck just enables easy and quick access to the important parts (folders, files, and other items) in each connected tool.

III. Concept development

Concept development was the goal of the first iteration. After defining the concept of the Peepdeck as described above we created minimal version of the UI in a form of a wireframe, and then created a paper prototype that we used in the following user research.

IV. Wireframe

The concept of the UI consisted of the vertical columns placed one next to each other (see *Fig 5*). Each of these columns consists of a title and number of items of different types. These items could be in a stream of information from selected source – such as stream of the updates of files from the connected cloud storage, or stream of the activities of the teammates.

The content of each item in each stream is dependent on the nature of the item, which can be a file, an activity, a calendar item, or a task. The items that are not in a stream represent a list of tasks, list of persons, or a block of multiple items of a different character. Variants of the prototype represented different levels of complexity for different items.

A wireframe was designed that was aimed at minimizing the content, abstracting away from the source of the content, and focusing on the type of information. That means, that the items of the same character (e.g., file, task) are represented on the screen similarly, regardless of the source of the item (e.g., whether the file is from Dropbox, OneDrive, or Google Drive). This way, the items are displayed with emphasis on their meaning, instead of structural features such as location where they are stored.

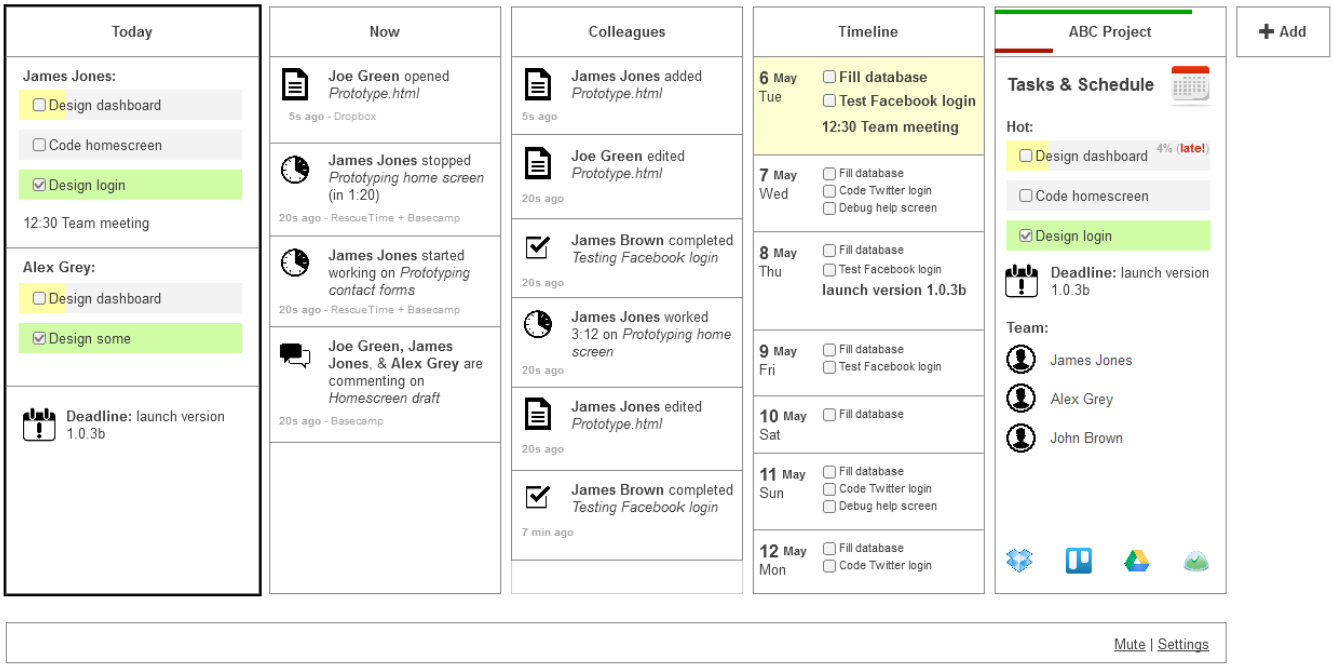


Fig 5. Wireframe of the first version of the UI of Peepdeck

V. Evaluation

The purpose of the evaluation was to validate the design concept and to investigate how to display relevant content at a single place. The test sessions also explored the amount and type of information and how to combine the information in a simple and intuitive way.

A formative user test was conducted with seven participants (six males and one female, with an average age of 30 years old). All participants were full-time employed designers working in the Netherlands on various design positions such as interaction designer, industrial designer, usability engineer, game designer, or graphic designer. All of them worked at middle- to large-size Dutch or international companies, in a team of six to ten people, some of them in multiple teams at the same time. User sessions were conducted in English; the interviewer and the five participants were not native English speakers, however all were proficient English speakers and no communication difficulties were noticed. Each session took about 60 minutes.

The user test involved a combination of a semi-structured interview and thinking-aloud method followed by co-design activities, where participants proposed alternative designs. . In the beginning of the session the interviewer asked about tools currently used, the size of the team, the roles of their current teammates, and typical information required in their work. Additionally, they were asked about the way they collaborate, and what tools or methods they use in their team.

For the evaluation, a paper prototype of the interface was created from the wireframe models so that interactions could be simulated by manipulating paper cut outs (*Fig 6*). We experimented with excluding different types of items to test whether users actually miss them. Participants were given a set of tasks pertaining to understanding one's current state in the context of the status of tasks, projects, deadlines, files and the activities of other team member activities. During the tasks participants were thinking aloud, and after the (successful or unsuccessful) completion of each task, the interviewer asked about the relevance of the elements of the dashboard for completing that task.

Fig 6. User testing of the version 1 of Peepdeck on a paper prototype

During the whole session both participant and interviewer had pens and markers available (each of different color) and participant was asked to add information into the dashboard that was missing for completing any task. Participants were also asked to cross out information that they find useless, or draw extra elements of the dashboard they are missing (*Fig 6*).

VI. Results

The results of the user test revealed that the concept of the dashboard was evaluated very positively. Participants often related the UI elements, interaction or their expectations about the interface to the existing tools they were familiar with; and not only tools they use for work. Examples of the UI elements they mentioned were "badge counters" known from iOS applications, showing in a small field at the corner of the app's icon number of notifications this app has; stacking information as known from Facebook (e.g., when an item is shared by multiple persons, the item is shown once with the list of all persons who shared it, instead of repeating the item for each person); or iconography known from OS X and Facebook. In general, participants often related to the tools they were using at work already, which confirmed previous findings about the list tools they actually use. Participants usually asked for better visual clarity, showing context of the items and relationships between related items. All participants executed the tasks in different ways suggesting the need to allow flexibility in the user interface.

In terms of general problems and requirements, participants stated that they often work on multiple projects at the same time, or switch between projects every few days during a month. Therefore, the interface should support multiple project views and easy handling of them (adding, hiding, showing, and removing). The presented categories of items (such as events, tasks, files and messages) were understood, however the relationships between items on the dashboard and the persons or projects were not clear. Especially for project-related items such as tasks or deadlines, they missed clear hints about the project it relates to. The opinions and preferences on how to present extra information were divergent. A common remark was that the dashboard must be easy to scan visually supported by colors, spacing and the typography.

When referring to missing information or features, the most common concern was to include filtering and search functions. Participants referred to the possibility to stack similar items together, manipulate content (list of tasks, current files), edit privacy settings (shared or not shared), and to show and hide details of each item (tasks' status, deadlines, last time of synchronization). Besides, each of the items should have a link to the original source and other related items. The dashboard should also include the availability of colleagues (chat, IM, call, personal meeting). An integrated social media stream was also suggested with the possibility to disable specific applications when they do not want to be disturbed.

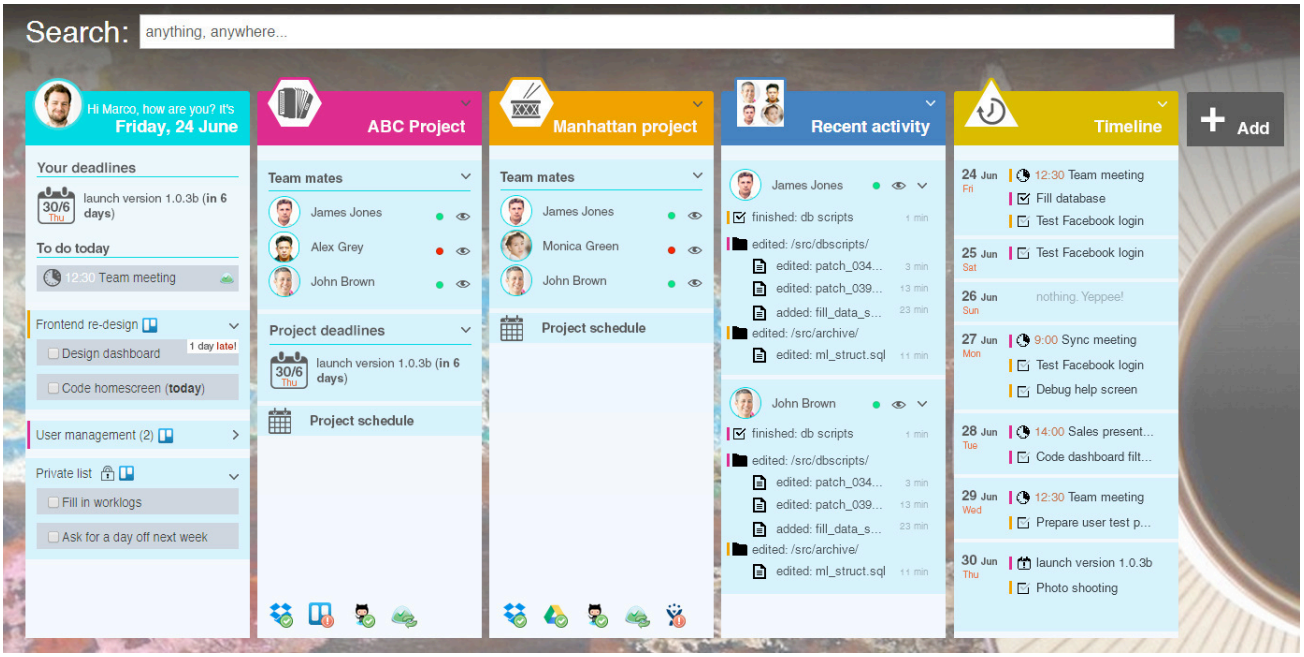


Fig 7. Final version of the UI of Peepdeck on a high-fidelity prototype

V. Second Design Iteration

The second iteration focused on creating and testing an interactive digital prototype implementing the guidelines resulting from the first iteration and user test. For instance, visual clarity, spaciousness and segmentation of the interface should dominate in order to support visual searching and effective scanning of the updates on the dashboard. Other components such as integrated search and availability of team members were included in the UI.

I. UI design

The wireframe of the first version of the dashboard was redesigned and then with the help of a mood board a visual design was created using colors and shapes with a meaning to support visual searching and peripheral scanning of the dashboard [1]. The interface, shown in *Fig 7*, consists of several columns, where the basic set of columns included: a "Today" column, a "Project" column, a "Team activity" column and a "Timeline" column. A full text search is placed visibly above all, to indicate that it searches all the content of the dashboard.

The "Today" column contains today's date and day of the week, photo and greeting of the user to attract attention and make user start reading the screen here. Then it lists the upcoming deadlines, lists of tasks, and a private list of tasks that are stored locally, and not shared. The "Project" column contains the header with the project name and an icon. This column contains a list of teammates, which are aggregated from tools that contain the defined team (e.g., Basecamp), or the users from the shared project folder (e.g., Dropbox). Each person has an indicator of availability on any connected communication service, and a visibility icon that can toggle the content related to this person in the whole dashboard. Underneath there is a list of upcoming deadlines of this project, a link to the project schedule (in selected tool where the project shares the calendar), and icons of all connected tools/services with the indication of synchronization status; these icons are shortcuts to the relevant project pages (e.g., Dropbox folder, or Basecamp project). The "Team activity" column contains the blocks of users and their recent activities such as edited files, or completed tasks. The "Timeline" column lists the simple linear calendar view with the upcoming events such as deadlines, meetings, and tasks. Users can add more columns into the dashboard, which will scroll horizontally.

II. Prototype development

An interactive prototype with high visual refinement based was created in the software Axure RP Pro 7.0. The result was an interactive mockup on HTML, CSS and JavaScript. The prototype implemented the designs created in the previous stage. It also added interactions allowing to manipulate the interface, such as collapsing and expanding the elements in the columns, hiding and adding new columns, do full text search with autocomplete function showing the search results during typing of the search phrase, adding new items on the private task list, hiding and showing content of each person on the dashboard, and showing extra information about elements on mouse hover. The prototype displayed in *Fig 7* is the improved version after the insights from the second evaluation. The prototype is available on the URL: http://ojwz1v.axshare.com/.

III. Evaluation

User tests were conducted with seven designers (four males and three females, all in their early 30s). All participants were employed full-time at the time of the tests by middle- to large-size Dutch or international companies, and they work in a team of six to ten people, often being part of multiple teams at the same time. The individuals held various design positions such as interaction designer,

industrial designer, usability engineer, game designer, or graphic designer. User sessions were conducted in English. Each session took about 60 minutes.

The goal of the evaluation was to test the UI, regarding its visual clarity, the information architecture and the amount of information presented. The connections and data on the dashboard were simulated. User interactions with the interface were specially observed: how they explore and understand the elements on the screen. The terminology and wording used in the interface were also evaluated.

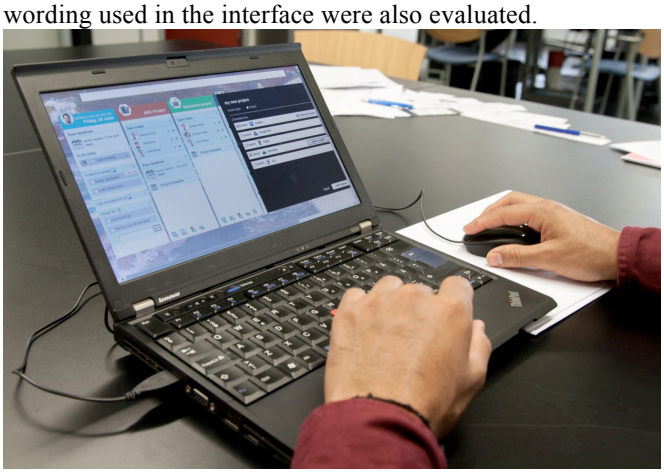


Fig 8. User evaluation of the version 2 of Peepdeck's UI using an the interactive prototype

The interactive prototype was tested on a computer with a mouse, in a full screen mode (*Fig 8*). A similar protocol to the first evaluation was executed during this evaluation. The user test was set up as a combination of a semi-structured interview, thinking-aloud while completing given tasks, and observation. The prototype was presented to participants, and then they were asked to complete a set of tasks while thinking aloud. These tasks were very similar to set of task performed during the first evaluation. The tasks pertained to the current status of one's activities, other member's activities, deadlines and projects. Some tasks focused on the added features such searching for a particular file, directly contacting a team member or unsubscribing from certain notifications. After each attempt the interviewer was asked about which elements of the dashboard were helpful for completing this task, and which parts of the dashboard were not. At the end, participants were asked several open-ended questions about situations in which they find this tool useful, and in which they think it would not be useful.

IV. Results

In summary, the feedback on the tool was very positive as some participants noted that they wanted or needed such a tool already. The visual segmentation and organization of the dashboard was positively evaluated, and people reported that searching was easy, and clear. The results showed that users want to customize the order of the sections – natural action is drag and drop the column to another place (almost all participants attempted to do this). Columns should have an option to manage its content (add, remove, edit, rename and reorder panels) and must allow the management of the tools connected to each column. Let people develop their own widgets for the dashboard (either elements in the columns or whole columns).

The observations also revealed that people were very actively exploring the interface, and learnt how to use it immediately or within one or two tries. This implies that the interface also supports fast learning, and motivates users to explore its functionality. Participants related many actions and interface elements to the tools or environments they are familiar with – Facebook, iOS, OS X, Gmail. It should be possible to collapse columns into tabs or icons, so that they can be opened easily and immediately when needed, but are not taking place on the screen when not needed (analogy to browser tabs, or minimizing applications in OS).

The participants highly valued the aggregation of the tools they usually use. As one designer explains: "I think I would really love it. Because from here, from this interface, I can easily go straight to certain file, certain person, (…) instead of searching: Where is this? Where is that?" [ppn1]. Another designer commented: "It's handy because you don't need to learn all the new tools. If someone prefers this and someone prefers others" [ppn3].

Since the dashboard integrates many activities, services and files, participants noted that there should be a visible confirmation after each action with the possibility to undo the action. Some suggested that, ideally, the dashboard should have a full integration with the connected services and with local files and that the interface should point directly to the aggregated content. Finally the interface should provide means for immediate interaction (instant message or email) with the people in the list of contacts.

VI. Discussion

Designers use many general-purpose tools for different collaborative tasks. Mastering the diversity and combination of the tools and functionalities, and keeping an overview of the process is a recurrent problem in design practice. The present study has explored means to link different tools, to filter relevant information across them and to switch easily between tasks during collaborative design projects. Through two design iterations followed by user tests, it was confirmed the relevance of a dashboard application presenting status information compiled from several general purpose tools for enabling designers to maintain an overview of their work and of the activities of the design team.

So far, feedback from users confirms the general design direction chosen: the dashboard should combine information from different but already familiar tools [7], rather than attempting to replace them. The first iteration identified the importance of optimizing the dashboard for visual scanning and support the search of content. The second iteration identified one more important requirement: to allow users to customize the dashboard as needed (by individual preference, role, workload, number of projects and team size).

Awareness is critical for collaboration. It provides an understanding of others and their activities and it helps guiding ones actions [2]. Awareness is dynamically built through practices [20]. In the design studio, the use of space and artifacts support the creation of awareness [23].

However, on remote environments, different awareness mechanisms should be developed to support seamless collaboration [4]. Through the integration and visualization of the different collaboration tools, Peepdeck aims to facilitate some the most relevant components of the design practice: social interactions supported by the exchange of artifacts [22] and information. Additionally, Peepdeck uses a user-centered approach [5] and adopts several awareness mechanisms such as personalization [4], workspace awareness [6], team availability [3] and work progress [2]. Further iterations involving different methods of evaluation on real scenarios must be done to ensure the generalization, precision and realism of these findings [14]. One can imagine that tools like Peepdeck can support other professionals for other types of collaborative work e.g., to quickly interact with the team (through sending emails or instant messages directly to team members), keeping and self-updated about the project development (following the activity of the colleagues, following the deadlines, tasks and their statuses), and accessing the shared resources easily.

VII. Conclusion

Two iterations were presented of the design, prototyping and testing of PeepDeck an application that aggregates information from tools and online services that are popular amongst design teams, allowing them to be aware of project work. This process identifies several requirements for the design of tools to support collaboration awareness for design teams, these are: a) support the use of collections of widespread tools rather than replace them with a special purpose one b) design the system as a glanceable display to support awareness and peripheral interaction c) allow customization to individual needs and practices.

While these requirements have been identified in related literature regarding collaborative work, the emphasis on the combining general-purpose tools for supporting design activities is new. Further, the notion of personalization and glanceability refer to specific needs of design teams, which differ from the interpersonal awareness applications that have occupied CSCW literature in the past. Future work will explore how functional prototypes of such awareness functionality can be implemented and the extent to which they can be accepted by design teams and to which they succeed in fostering awareness.

References

[1] S. Coradeschi and A. Saffiotti, "Perceptual anchoring of symbols for action", in: *Proceedings of the 17th International Joint Conference on Artificial Intelligence (IJCAI)*, Seattle, WA, pp. 407–412, 2001

[2] P. Dourish and V. Bellotti, "Awareness and coordination in shared workspaces". In Proceedings of the ACM CSCW '92 Conference on Computer Supported Cooperative Work, pp. 107-113, 1992

[3] W. W. Gaver, "The Affordances of Media Spaces for Collaboration", In Proceedings of the Conference on Computer-Supported Cooperative Work - CSCW'92 (Oct. 31-Nov. 4, Toronto, Canada). N.Y.: ACM, pp. 17–24, 1992

[4] T. Gross, "Supporting Effortless Coordination: 25 Years of Awareness Research", Computer Supported Cooperative Work, vol. 22, pp. 425–474, 2013

[5] T. Gross, C. Stary and A. Totter, "User-Centered Awareness in Computer-Supported Cooperative Work-Systems: Structured Embedding of Findings from Social Sciences", International Journal of Human-Computer Interaction (IJHCI), Vol. 18, no. 3, pp. 323–360, 2005

[6] C. Gutwin and S. Greenberg, "A Descriptive Framework of Workspace Awareness for Real- Time Groupware", Computer Supported Cooperative Work: The Journal of Collaborative Computing, Vol. 11, no. 3–4, pp.411–446, 2002

[7] P. Hekkert, D. Snelders, and P. C. W. Van Wieringen, " 'Most advanced, yet acceptable': Typicality and novelty as joint predictors of aesthetic preference in industrial design". British Journal of Psychology, 94, 1, pp. 111-124, 2003

[8] J. Holmgren, J. Juola and R. Atkinson, "Response latency in visual search with redundancy in the visual display", *Percept. Psychophys.* 16, pp. 123-128, 1974

[9] J. Juola, *Cognitive Psychology,* Cengage Learning, 2009

[10] H. Karsten and M. Jones, "The long and winding road: Collaborative IT and organisational change" in *Int. Conference on Computer Supported Work* (CSCW'98), 1998

[11] T. Matthews, "Designing and evaluating glanceable peripheral displays" in *Proceedings of the 6th conference on Designing Interactive systems,* pp. 343-345, 2006

[12] J. E. McGrath, "Groups: Interaction and performance", Vol. 14. Englewood Cliffs, NJ: Prentice-Hall, 1984.

[13] J. Muñoz-Alcántara, P. Markopoulos and M. Funk, "Social Media as ad hoc Design Collaboration Tools" in *Proceedings of the European Conference on Cognitive Ergonomics 2015* (ECCE '15), 2015

[14] D. C. Neale, J. M. Carroll and M. B. Rosson, "Evaluating computer-supported cooperative work: models and frameworks" in *Proceedings of the 2004 ACM conference on Computer supported cooperative work* (CSCW '04), pp. 112-121, 2004

[15] D. Norman, *The Design of Everyday Things*, The MIT Press, London, 1998

[16] R. Oppermann, *User Interface Design*, Institute for Applied Information Technology, GMD Forschungszentrum Informationstechnik, Germany, 2002

[17] W. Orlikowski, "Sociomaterial Practices: Exploring Technology at Work," *Organization Studies*, 28, pp. 1435-1448, 2007

[18] V. Pipek and V. Wulf, "Infrastructuring: Towards an Integrated Perspective on the Design and Use of Information Technology", in *Journal of the Association of Information System* (JAIS), Vol. 10, Issue 5, pp. 306-332, 2015

[19] C. Rossitto, C. Bogdan and K.S. Eklundh, "Understanding Constellations of Technologies in Use in a Collaborative Nomadic Setting" in *CSCW Journal,* vol. 23, no. 2, 2014

[20] K. Schmidt, "The Problem With Awareness: Introductory Remarks on Awareness in CSCW"., *Computer Supported Cooperative Work: The Journal of Collaborative Computing*, Vol. 11, no.3–4, pp. 285–298, 2002

[21] K. Schmidt and L. Bannon, "Constructing CSCW: The First Quarter Century" in *Computer Supported Cooperative Work*, 22: pp. 345–372, 2013

[22] K. Schmidt and I. Wagner, "Coordinative artefacts in architectural practice" in *Proc. of COOP 2002*, pp. 257–274, 2002

[23] D. Vyas, G. van der Veer and A. Nijholt, "Creative practices in the design studio culture: collaboration and communication" in *Cognition, Technology & Work*, 15, 4, pp. 415–443, 2013

[24] V. Wulf, "Evolving Cooperation when Introducing Groupware – A Self-Organization Perspective" in *Cybernetics and Human Knowing*, 6(2), pp. 55 – 75, 1999

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1671–1674
DOI: 10.15439/2016F306
ACSIS, Vol. 8. ISSN 2300-5963

Simulation of Universal Design by a Functional Design Method and by Gamification of Building Information Modeling

Jukka Selin
Department of Electrical Engineering and Information Technology, Mikkeli University of Applied Sciences
Ltd Patteristonkatu 2, 50100 Mikkeli FINLAND
jukka.selin@mamk.fi

Markku Rossi
Department of Electrical Engineering and Information Technology, Mikkeli University of Applied Sciences
Ltd Patteristonkatu 2, 50100 Mikkeli FINLAND
markku.rossi@mamk.fi

***Abstract*—We have developed a method and a process with which we can simulate the functionality and the space requirements of different actions in buildings. We utilize Gamification of Building Information Modeling (BIM). The essential goal of Building Design is to produce designs corresponding to customers' needs. In our method the electronic CAD model is inserted into a Game Engine and is gamified. Features and limitations that correspond to the reality as exactly as possible are generated for the virtual objects steered in the game by the designers. In addition, the space requirements for the actions performed by a simulated user can be dimensioned by our Functional Design Method. We generate 3D space objects of maximum space requirements for actions from videos of real human actions. The 3D space objects created this way can be Collider Objects used by the designers attending to the Gamification. We get simulated User Experiences from the rooms under design. This method helps to understand better than before different end users and their needs. The result is buildings that fit the requirements of the users.**

I. Introduction

The Pilot Experiments described in this article use our method to generate computer game like environments out of a Building Information Models (BIM) by Gamification. The gamified Building Model enables us to simulate actions in buildings.

With the aid of the gamified model we can simulate different User Experiences in a building already when the building is under design.

The ”Players” following different roles can move like the First Person Controller (FPC) or the Third Person Controller (TPC) in the virtual space. Features and limitations according to a role (e.g. a Handicapped using a wheelchair) can be realised.

The Game Object can include the space requirements belonging to the actions of the role. The space requirements are from our Functional Design Method, FDM [1],[2]. From the space requirements we can generate variable Collider Functions around the Game Object. The method makes it possible to simulate different functionalities and space requirements related to actions. The Player can e.g. take the role of a person using a rollator. With the aid of the gamified Building Model he can get the User Experience of moving around with a rollator. We detect already in the beginning the possible problems and shortcomings, and corrections are then low coast and fast compared to during the construction or even after the building project.

The developed method also helps designers of the Construction Industry to better understand the needs and limitations of buildings. It is also possible to involve the user groups to the design via the method of gamified Building Model. The end users can get a view that is understandable and real enough. We can then achieve buildings that match more closely their purpose.

With the method it is also possible to automate the testing against the Universal Design principles, by moving 3D-models with their space requirements (like a wheelchair patient with two assistants) automatically through the model. After the tests the possible problems and their locations can be found in the report automatically generated.

If the designer uses VR glasses, like Oculus Rift, the User Experience of the gamification of Building Information Models can be quite realistic.

The gamification of the models serves all phases of the construction project from sales and marketing until the completion of the Architectural Design. By the aid of the gamified model it is possible to communitise and crowdsource by distributing the gamified model in the Internet. The end users can then be given an opportunity to present ideas, test functionalities and so attend to the design.

This article first describes the methodologies and software related to the research. Then we present the new method and the process for gamification of Building Information Models as an aid in design. Finally, we observe the meaning of the new results and scenarios for the next steps.

II. Methodologies and software used in the Research

The developed method utilises a work based invention report at Mikkeli University of Applied Sciences Ltd (Mamk) filed by Jukka-Pekka Selin. The maximum space requirement in dimensions x,y and z can be derived from video clips of real human actions. In this method the real actions are videoed with at least two video cameras that are situated at right angles. The need of space from the actions in three

dimensions is measured and a corresponding geometrical object is created. This 3D object in IFC file format is compatible with CAD software. The goal is to ensure that the required actions can fit the space under design. [1],[2],[3].

During the research program the application that is BIM compatible, under development and is meant for Lifecycle Management of Building Data, and the application extension Value Add Data of Mamk R&D, were used for creating 3D objects. The application extension realizes the functionality of the FDM. The Colliders describing the space requirements for different actions to the gamified model were created via the execution of the Value Add Data software.

The 3D-model that contains a Building Model was realised with the ArchiCAD design software and it was gamified with the Unity Game Engine.

III. The FDM together with the Gamification of the Building Information Models enhance the quality, visibility and understandability in Design

When the practices of BIM develop, new ways to design, visualise and analyse buildings are developed in parallel. The earlier the issues are detected and corrected, the lower are the costs and related work efforts. The FDM we developed earlier is a method and process born out of need. It can be applied very well in parallel with BIM, and could even be standardised to be an integral part of it.

The FDM can be the basis for developing new analysis functions, techniques and methods to support the design process and the designers.

We combined the FDM to the gamification of the Building Models and did research on whether it could be possible to create User Experiences about how the usage of e.g. a wheelchair or a rollator succeeds in the spaces under design.

With the aid of the 3D space objects from the FDM we can generate e.g. Colliders around the functional unit (e.g. a wheelchair user with two assistants). Collider is a term used in Game Design. The Collider then corresponds to the measured space requirements of the human action. In the First Person Controller concept the Colliders created this way can be attached to the virtual model of a wheelchair Use Case. After this we could move around with the wheelchair in the gamified Building Model and swap the Collider according to the actions the handicapped and his assistants were doing.

When VR glasses like Oculus Rift are used, we got a very realistic User Experience about how it would feel to move and perform different kinds of functions with a wheelchair in a real building.

Our partners agreed this kind of visualisations and generation of User Experiences are one of the directions for forcefully developing BIM and its applications. This kind of approach is quite new in the relatively conservative Contruction Industry, but we think it will become more common.

The virtualisation of buildings and gamification open new perspectives to the whole construction business and the design processes used there. It e.g. enables to involve the future users to the idea gathering phase and to the design work in ways not possible earlier.

A. The space required by Human Actions is converted into a 3D object by the FDM

The design needs described earlier get means of enhancement from the FDM from Phil.Lic. Jukka-Pekka Selin. The method helps especially in a situation where the building operations are not yet active but there exists a digital model of the building.

In addition to a mere idea we offer a ready process that enables to transfer the space requirements of real Human Actions to the BIM based designs via our 3D IFC-objects. It is possible to test and simulate with the Building Model how the functional demands fit the rooms under development. The diagram below shows the process of the FDM:

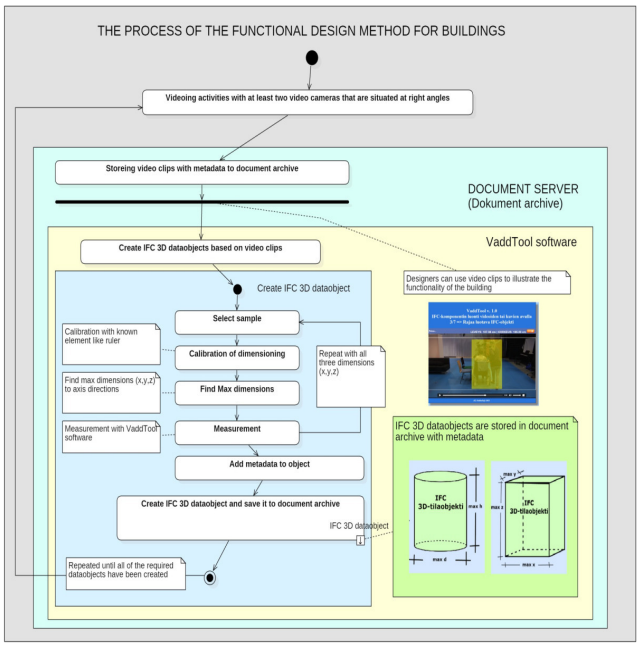


Fig 1. The process of the FDM for Buildings.

An example of the meant human activities is “an assistant who helps a wheelchair user to dress on outdoors clothes“. The invention to use real human activities and turn them into 3D CAD objects is from Licentiate of Philosophy Jukka-Pekka Selin who is the principal lecturer of data processing at Mamk.

Fig. 2 shows how the dimensioning (x, y, z) of the forthcoming IFC data object is performed.

Fig 2. A screen shot of the Space Required by Activity dimensioning function of the Value Add Data software prototype.

B. The Gamification of Building Models with the Unity Game Engine

We realised a pilot project at Mamk where we gamified the building model of the headquarters of the construction company U.Lipsanen Oy. The foundation was the 3D-model from ArchiCAD software. It was a part of the building model [4]. The Unity Game Engine supports importing data in various 3D file formats. After iterative testing and piloting we ended up to recommend that the import format should be the FBX format developed by Autodesk, Inc. [5]. Unity has native FBX support.

C. Simulating the actions of a handicapped building user

We piloted the methods by simulating a wheelchaired, moving person. The space requirements for different actions where first generated by the FDM. The original material was a set of video clips about real actions that were input to our Value Add Data software for creating IFC files.

Fig 3. A 3D building model converted first to FBX has been imported to the Unity Game Engine.

In the pilot we created a Third Person Controller-type Player who moved with a wheelchair. As much realism as possible was programmed for the dimensions, the movements and the building automation. We also created swappable Colliders around the virtual player. This way the space requirements can be swiftly selected according to the action under test. We can see and experience in a concrete way how well the action fits the spaces under design.

The dimension data from the FDM, representing the space need of a real action, can be utilised as such when creating Colliders in the Game Engine. One of the future research topics is to develop an automated creation procedure of Colliders based on the 3D IFC files. The next table below shows as an example the space requirements of two actions studied in piloting.

TABLE I.
MAXIMUM SPACE REQUIREMENTS FOR THE DIMENSIONING OF THE COLLIDERS CREATED BY THE FDM

The dimensioned Action (VaddTool)	Space Requirement (m)
The wheelchair user is helped to dress on or off his shoes by his assistant	1.10 x 1.90 x 1.51 (x,y,z)
The assistant walks with the wheelchair user on the side or behind the wheelchair	1.07 x 1.90 x 1.30 (x,y,z)

Respectively it is possible to dimension and simulate all kinds of actions, also not related to Building Design, by using the Functional Design Method and Gamification of 3D Models. There is a lot of potential in generalisation.

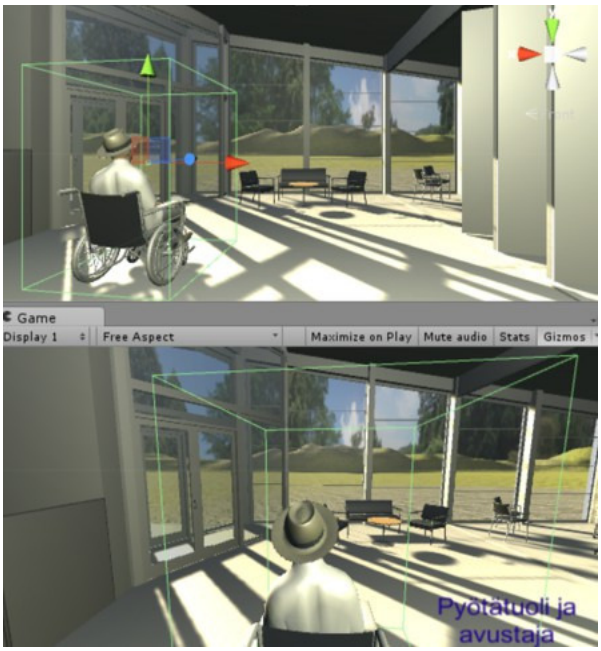


Fig 4. A handicapped virtual player with an added Collider Box (in green) derived with the FDM. Respectively it is possible to dimension and simulate all kinds of actions, also not related to Building Design, by using the FDM and Gamification of 3D Models.

D. Work process oriented, location aware document management and viewing services for the whole Lifecycle

After testing the creation of 3D IFC objects and short term document storage with the Value Add Data software prototype we began to negotiate with the public financiers about developing a service that would manage the documents during the whole lifecycle of buildings. In December 2015 the construction company U.Lipsanen Oy won a partial public financing for their development project RedHal.

The new targeted functionality differs from the presently available services in several features. After the documents reach the status " As built – Ready ", we will move the documents to a long term data repository with a targeted service life of 100 years. The University is an expert in Data Migration through storage technologies and in semantic digital archiving [7].

We are targeting to use indoor positioning in the 0.5 metre accuracy range together with tablets and Augmented Reality services. The data structures both in the Pre As built service and in the long term repository are according to the semantic structures of buildingSMART.

The document viewing will be of an active type. It takes into account the role of the viewer, the location, the orientation and the current phase of the design or construction process. The co-operation between designers will be based on Building Models according to the BIM standards. This project will be finalised by June 2018.

IV. Results and Conclusions

The following picture describes the process aspect in piloting our method.

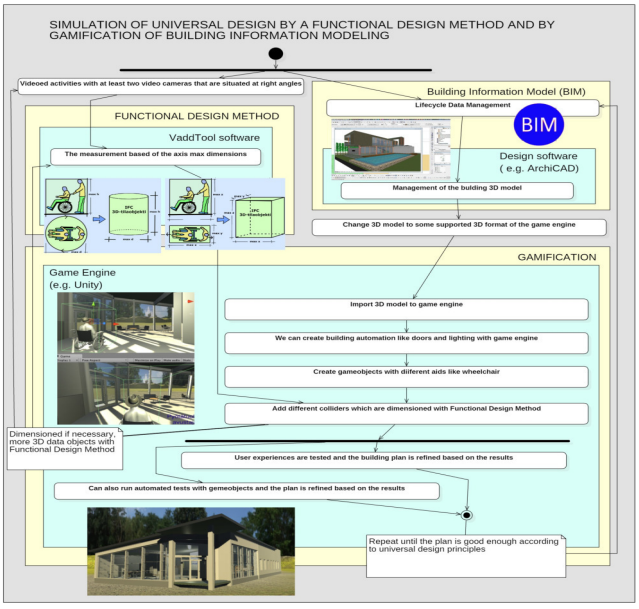


Fig 5. The process developed at Mamk in simulation of Universal Design with the aid of a gamified Building Model.

The results from the piloting are promising. The User Experience in the gamified model can be highly natural. The attendants of the Design Game thought that by bringing a handicapped virtual player to the game, a demonstrative and mind-expanding understanding was created about how well the building corresponds to users' needs.

Their opinion was that the method helps any designer to take the position of a handicapped user and this way to understand their needs and take them into account. The method also helps to test how well the building automation (e.g. active doors and adaptive lighting) serve different users. The FDM was found during the piloting also to be practical in ensuring that the spaces are dimensioned according to real needs. The idea behind the developed method about FDM based Colliders for the Players in a simulation based on gamified Models was found to be good and practical.

We also studied the automating of dimensioning testing by moving Players with Space Requirement Colliders automatically through pre-selected rooms in a way that the collisions and reroutings were documented to a log. Now we saw at once the locations which were too small or narrow for the required actions. It was also possible to study the steepness of ramps. The partners said the gamification of Building Models and the virtualisation of the usage of buildings create new and necessary dimensions to the whole building design process and we have the right future direction in our research. The development of Design Processes can now move towards crowdsourcing and communitisation by bringing the gamified Building Model e.g. via the WebGL technology to the Net and by enabling the future users of buildings access to the preliminary designs, for commenting. We think these processes can help in creating buildings that really fit the users' needs.

References

[1] Toimintatilan määrittävä objekti tietokoneavusteista suunnittelua varten. Finnish patent application 20135286FI. Applicant Mikkeli University of Applied Sciences Ltd (Jukka-Pekka Selin).

[2] An action space defining object for Computer Aided Design. PCT patent application WO 2014/154942 A1. Applicant Mikkeli University of Applied Sciences Ltd (Jukka-Pekka Selin).

[3] Rossi, Markku J., Dave, Bhargav., *Digitalization and quality enhancement initiatives in sw assisted design processes in building and construction industries.* 9th International Conference on Computer Engineering and Applications, Dubai, 2015. ISBN 978-1-61804-276-7.

[4] Rakennusliike U.Lipsanen, Rakennusliike U.Lipsanen Oy Headquarters building (Lipatie 1, 76850 Naarajärvi). A 3D Model designed with ArchiCAD saved in the 3DS format. Contruction Company U.Lipsanen Oy, Pieksämäki, Finland, 2015.

[5] Autodesk Inc. The official website of the Autodesk FBX format. Accessed on March 30, 2016. http://www.autodesk.com/products/fbx/overview

[6] Autodesk Inc. The official website of the FBX Converter software. Accessed on March 30, 2016. http://usa.autodesk.com.

[7] Uosukainen, Liisa (ed.), Open Source Archive. Towards open and sustainable digital archives. Mikkeli University of Applied Sciences, Series A, No. 94, 2014. 100 pp. http://urn.fi/URN:ISBN:978-951-588-456-5 . Accessed on May 9, 2016.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1675–1679
DOI: 10.15439/2016F529
ACSIS, Vol. 8. ISSN 2300-5963

Evaluation of Affective Intervention Process in Development of Affect-aware Educational Video Games

Mariusz Szwoch
Gdańsk University of Technology
Faculty of Electronics, Telecommunications and Informatics
G. Narutowicza St. 11/12, 80-233 Gdańsk, Poland
Email: szwoch@eti.pg.gda.pl

***Abstract*—In this paper initial experiences are presented on implementing specific methodology of affective intervention design (AFFINT) for development of affect-aware educational video games. In the described experiment, 10 student teams are to develop affect-aware educational video games using AFFINT to formalize the whole process. Although all projects are still in progress, first observations and conclusions may already be presented.**

I. Introduction

AFFECTIVE computing is an emerging field of computer science that deals with human affects. As affective applications try to automatically recognize, interpret and react to human emotions, their development demands interdisciplinary research incorporating not only computer vision or pattern recognition but also human behavior studies such as psychology and cognitive science. If such affective awareness is built into an application to extend its functionality, it is called *affect-aware*; in contrary, the primary goal of *affective applications* is focused on human emotions. The concept and potential of affective and affect-aware applications may be effectively exploited in the nearest future in many fields, such as healthcare, education, entertainment etc.

Video games seem to be among the most natural application area of affect-aware concept. Practically all entertainment provided by video games to a player is somehow based on his or her emotions. It is usually informally introduced into the game at its development stage based on the assumed model of so-called representative player. Unfortunately, such a static approach does not take into account that each player differs to a certain extent from that averaged model and, more importantly, a player's affective state can dynamically change, even radically, from session to session making it almost impossible to predict the current emotions at the development stage. That is why, it is so important to create at the development stage emotional model of the player and implement some methods of emotion recognition or estimation.

Unfortunately, there are no standards of development affective or affect-aware software. Existing methodologies of software engineering have no intrinsic rules or templates that would enable development of affect-aware or affective applications. Recently, a new methodology of affective intervention design (AFFINT) has been proposed [1] for development of affect-aware intelligent systems. The proposed process consists of 10 development steps with a predefined order of their implementation. The practical implementation of AFFINT process have been explained using three case studies of Gerda tutoring system and two prototype affect-aware video games.

Despite the exhaustive description of AFFINT process and given case studies the question arise whether it is already ripe enough to be directly used by software engineers to develop affect-aware and affective applications. In this paper, the initial experiences are described of using AFFINT process in development of 14 prototype affect-aware educational video games. Although, games are developed for different platforms and using different technologies their common denominator is their educational aspect and mandatory use of AFFINT. Gathered programmers experiences allow to supplement AFFINT description with some practical comments and examples.

II. Background

In general, the main goal of adding affect-awareness into any software is to enhance its primary purpose, or functionality, by adjusting some of its elements to the current emotional state of the user. Such modification of a system behavior according to the user's affective state is called *affective intervention*. There are possible different aims of such affective intervention, depending on such aspects as the application area, the system goal etc. For example, an educational program may try to keep the user in the so called *flow state* that provides the best learning

This work was supported in part by Polish-Norwegian Financial Mechanism Small Grant Scheme under the contract no Pol-Nor/209260/108/2015 as well as by DS Funds of ETI Faculty, Gdansk University of Technology.

effects [2]. In contrary, a video game do not always has to keep the player in a specific emotional state; instead, it can adapt the gameplay, especially its difficulty, to give the player the best playing experience at the moment [3].

In order to effectively implement affect-aware functionality in any application software engineers have to formally define *affect model* of its user and *affective intervention model*. The first model defines what emotional states of the user are taken into consideration, while the second one defines conditions and realization way of the performed affective intervention [1]. What is the most important these models have to be defined in the very early development stage of software development, regardless the methodology used. Such approach guarantees that affect-awareness is taken into account already in the project phase preventing from adding affect-awareness as additional feature to already developed application.

Unfortunately, there are no specific software development methodologies defined for development of affect-aware of affective applications and AFFINT process is the only formalized proposition enabling the design and evaluation of affective intervention models [1]. Affect-aware and affective software is obviously dependent on various methods of recognition or estimation of user's emotion. Such methods may use, in general, different input channels according to different ways of emotions' expressing [4]. The most frequently used emotion recognition methods include:

- facial expression recognition (FER) based on video input channel [5], thermovision or depth sensors[6];
- voice analysis based on audio channel [7];
- analysis of physiological signals such as heart rate or skin conductance [8][9];
- textual input analysis in using a system interface [10];
- analysis of different behavioral patterns in using standard input devices such as mouse, keyboard, and pad [11][12];
- analysis of the current user's progress within the application, e.g. in a quiz or gameplay, and additional application events, such time lapse, new challenge [3].

Although these methods can be used alone, better results are usually obtained by fusing information from diverse input channels (*early fusion*) or different methods and algorithms (*late fusion*) [13].

The second element of affective and affect-aware software is *affective intervention*, which is a program response to the recognized emotional state of a user. There are many studies concerning affective phenomena in human-computer interaction (HCI). They focus on many different aspects, such as defining users' emotional states appearing during satisfying and unsatisfying experiences with applications [14], using affective interventions to reduce users' frustration [15] and increase their efficiency in the performed tasks in particular application domain, e.g. e-learning.

Another group of publications focuses on design and evaluation of affective applications and their affective interfaces and interactions [16]. Many studies emphasize the fact that abandoning the concept of a 'standard user' (or player for video games) in favor of adaptive affective approach often leads to greater users' satisfaction and more efficient and effective performance of their tasks [3][16]. On the other side, affective interventions must be subject to certain rules, restrictions and limitations concerning their frequency or influence upon the user [17]. Such rules allow combining emotion recognition methods and affective interventions into *affective feedback loop*.

III. AFFINT PROCESS

AFFINT approach proposes a ten-step process that formalizes incorporation of affect-awareness into the software development methodology [1]. These ten activities are numbered in the desired application order and mapped into the four stages of system development (Fig.1):

I. System definition consisted of three activities:
1) *Application goals and tasks* that should be supported by affective subsystem;
3) *Available input channels in application environment* that can be used by emotion recognition or estimation algorithms;
6) *Available output channels in application interface metaphor* that can give the user a feedback about the recognized user's emotional state or the performed affective intervention;

II. Affective intervention solution set that includes:
2) *Effective emotional activations* that defines a subset of user's emotional states that are optimal to reach the application's goals;
4) *Available emotion recognition solutions and representation models* that can be used in the application;
7) *Possible affective interventions of an application* that define list of possible scenarios;

III. Affective intervention model layer that comprises:
5) *Emotion recognition granularity and methods* that define the specific emotion representation model and characteristics to be used in emotion recognition;
8) *Affective intervention triggering rules* that binds possible emotional states of the user with affective interventions;
9) *Affective intervention constraint rules* that limits the frequency and scale of affective interventions to create less artificial human-computer interaction;

IV. Evaluation of intervention model layer containing
10) *Validation with end users*, which is a natural assessment of application's quality.

As the proposed order of performing particular activities is not absolute within each layer, additional precedence, or dependence, relations are proposed to ensure their proper and logical sequence (Fig.1). The detailed description of AFFINT is available in [1].

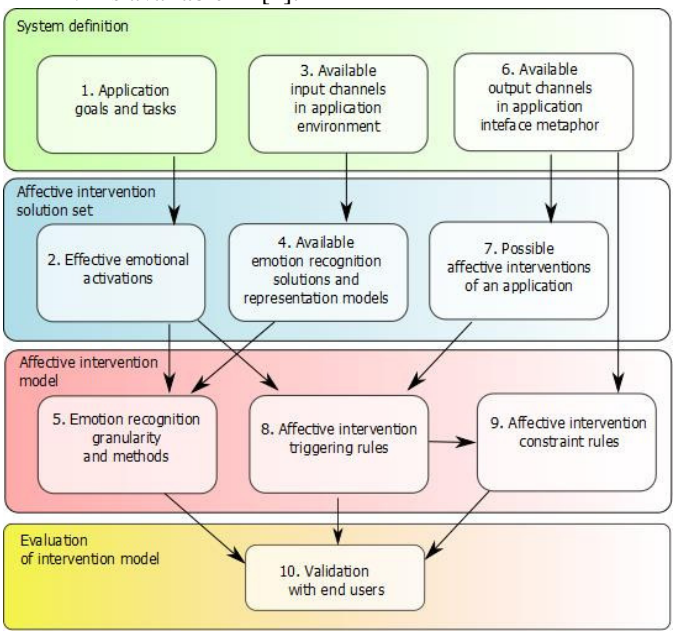


Fig. 1 AFFINT process of affective intervention design (arrows indicate precedence) [1]

IV. Experiment Assumptions

Although AFFINT process and its three case studies have been described in details, its applicability in practical software projects remains unknown. Three important questions have to be answered. Firstly, whether the AFFINT process is general enough that it can be used in wide spectrum of affective and affect-aware applications? Secondly, whether its description is sufficiently comprehensive, detailed and coherent to be easily followed be software engineers for such applications? Finally, is the proposed order of defined activities correct and how it should be mapped into development stages of different software development technologies?

An exhaustive answer to all these questions demands in-depth analysis of many different projects developed in different fields of applications using different methodologies of software engineering. This paper describes the first of series of planned experiments. In this experiment, 10 groups of experienced IT students at Gdańsk University of Technology (GUT) were given a task of designing and developing an affect-aware educational video game during a one-semester project within Interactive Multimedia Systems course. Narrowing the topic allows to focus on the second and the third of the asked questions, e.g. easiness of AFFIT's implementing in his specific domain by software engineers with no previous experience in affective computing.

All participating groups were allowed to choose the main target platform of the game, their preferred development environment and, of course, to design their own concept of the game. One additional requirement of using specific Emotion Recognition Framework (ERF) [3] was bound to PC target platform. Finally, four teams decided to develop a game for PC platform, three teams target at mobile devices (with Android system), and two groups decided to develop web applications (Table I). Almost all teams preferred to use Unity 3D environment, except one team whose members decided to use Phaser framework. These choices where made mainly based on their knowledge of particular environments. Additional advantages of Unity taken into account were its popularity and the fact that it allows deploying the same project to different target platform with relative easiness.

TABLE I.
PROJECT DECISIONS ON TARGET PLATFORM AND DEVELOPMENT ENVIRONMENT

Target platform	Development Environment	Number of teams
PC (Windows)	Unity	5
Mobile (Android)	Unity	3
Web (HTML5)	Unity	1
	Phaser framework	1

V. Validation of AFFINT Process

All teams' members participated in a special lecture dedicated to affective computing, video games and educational software. This allowed to specify the general characteristics, goals, limitations, and minimum requirements. With this background, students were given the article [1] describing the AFFINT process to verify whether the given description and case studies provide a sufficient basis for its direct implementation. Due to the short time of the project, the development process has been arbitrarily divided into four reported stages, namely Requirements Specification (RS), Game Concept and Design (GCD), Implementation and Tests (IT), Verification of Requirements and Product Validation (VRPV). Within the experiment, all teams were to map and to define particular AFFINT activities into these stages on their own.

Unfortunately, most teams reported that despite the detailed description, they still had some doubts and questions. The most important problem was that the suggested order of implementing particular activities is not always obvious and possible to follow. For example, it was quite easy to define some elements of the activity 5 (e.g. emotion recognition granularity) at the GCD or even RS stage, while proper identification of available emotion

recognition solutions demands sometimes in-depth studies that postpone reaching another development phase.

All these problems indicated that design and development of affect-aware applications, even using formal and detailed AFFINT model, requires some knowledge and experience in the field of affective computing. In this experiment, additional training lectures were offered to the students as well as individual tutoring for each project. This allowed to overcome most initial difficulties and problems with proper definition of AFFINT activities. In order to deal with some knowledge gaps at the early stages of the project, an incremental approach was accepted, that will allow to extend or even modify previously defined AFFINT activities when new information becomes available.

After two first stages of video games' development, some interesting observations and conclusions may be drawn from using AFFINT process in the described experiment. Definition of activities 1-3 & 6 was quite easy and natural. Application goals and tasks (act.1) are strictly bound to the specific concept of the application. For the definition of effective emotional activations (act.2), all teams, except one, assumed usage of one axis of the Pleasure-Arousal-Dominance (PAD) emotional space. It significantly simplified the description of users' emotions by using single variable with negative and positive values. All these teams defined exactly three recognized emotions: positive (e.g. joy), zero (neutral), and negative one (e.g. sadness). One team assumed usage of two axes (PA) and recognition of five emotions, but subject to possible reducing after initial tests of recognition methods.

All teams targeting at PC platform planned to use video as a base input channel (act. 3) additionally supported by analysis of usage of standard input devices, e.g. keyboard (three teams) and mouse (two teams). This was an optimal choice taking into account existence of few off-the-shelf libraries for face detection and facial expression recognition like Noldus FaceReader, and other. Unfortunately, there are no such trusted and freely available solutions for the mobile devices, and web applications has serious limitations in the access to system resources. That is why other teams relied mostly on emotions estimation by analysis of the players' behavior during the gameplay. Additionally, one team planned to use fitness bend with Android driven smartphones and tablets, while developers of web application planned to analyze mouse movements and clicks during the play. Finally, available output channels in all designed games were defined as gameplay difficulty and additional visual effects.

Although all teams were on the same development stage, the progress of their concepts, design and AFFINT description varied considerably. While some teams assumed additional tests and research of emotion recognition possibilities, other teams presented consistent and complete vision of the game. For example, Fish Quiz game for young players assumed development of motor skills of the player as well as broadening his or her knowledge in ichthyology. The goals of the player are to click on different fish species (Fig.2a), avoiding crabs, and correctly answer quiz questions to advance in experience levels and receive medals (Fig.2b). The affective model of the player consists of two states, namely joy and frustration, which are controlled by only one parameter influenced by successful and failed clicks. Thus, the input channel contains only mouse clicks and players advances within the gameplay. In turn, affective intervention controls the fish speed, their attraction to the mouse cursor, and the frequency of crabs appearance.

a)

b)

Fig. 2 Sample screens from Fish Quiz game by J.Atroszko, A.Cholewczyńska, K.Gersten at the developer stage: a) action mode, b) quiz mode

Conclusion and Future Work

The described experiment has brought several interesting conclusion about development of affective and affect-aware software in general, and usage of AFFINT process, specifically. First of all, development of such applications demands software engineers with some experience in this field or at least trained in the appropriate theory. Contrary to initial expectations, it seems impossible to correctly project and develop such applications using AFFINT

without basic understanding of advantages and limitations of affective feedback loop, possible emotion recognition methods and their input channels, and also about the possible spectrum of affective interventions.

Secondly, not all AFFINT activities can be precisely defined at the early stage of requirements specification or game concept and design, as they may need some additional tests or at least tuning. Using an incremental model for AFFINT definition seems to be a good approach. Despite described problems, AFFINT process proved to be very useful formalism that enforces taking affective issues into account during the whole design and development of affect-aware and affective software, especially in e-learning and video games.

Our future work will focus on introducing some modifications to AFFINT process in order to make it more flexible and thus better tailored to deal with possible uncertainty in various aspects in development affect-aware applications as well as with the specificity of agile methodologies of software development. Additionally, analysis of AFFINT documentation from all ten projects along with some feedback from developers will allow us to enrich the process with a set of predefined solutions for different stages of development of affect-aware applications.

REFERENCES

[1] A. Landowska, M. Szwoch, W. Szwoch, “Methodology of Affective Intervention Design for Intelligent Systems,” Interact. Comput. 2016, doi:10.1093/iwc/iwv047.

[2] R. Baker, “Modeling and understanding students’ off-task behavior in intelligent tutoring systems,” Proc. of the SIGCHI conference on Human factors in computing systems, ACM 2007, pp. 1059-1068.

[3] M. Szwoch, “Design Elements of Affect Aware Video Games,” roceedings of the Mulitimedia, Interaction, Design and Innnovation Article No. 18, 2015.

[4] H. Gunes, B. Schuller, “Categorical and dimensional affect analysis in continuous input: Current trends and future directions,” Image and Vision Computing, vol. 31, 2013, pp. 120-136

[5] J. N. Bailenson, E. D. Pontikakis, I. B. Mauss, J. J. Gross, M. E. Jabon, C. A. C. Hutcherson, C. Nass, O. John, “Real-time classification of evoked emotions using facial feature tracking and physiological responses,” International Journal of Human-Computer Studies, 66(5), 2008, 303-317.

[6] M. Szwoch, P. Pieniążek, “Facial Emotion Recognition Using Depth Data,” The 8th Int. Conf. on Human System Interaction, pp. 271-277, IEEE, 2015.

[7] Z. Zeng, M. Pantic, G. Roisman, T. S. Huang, “A survey of affect recognition methods: Audio, visual, and spontaneous expressions,” Pattern Analysis and Machine Intelligence, IEEE Transactions on, 31(1), 2009, pp.39-58.

[8] S. H. Fairclough, Fundamentals of physiological computing, Interact. Comput. 21 (1-2), 2009, pp.133-145.

[9] W. Szwoch, “Using Physiological Signals for Emotion Recognition”, Proc 6th International Conference on Human Systems Interaction, 2013, pp. 556-561.

[10] H. Binali, C. Wu, V. Potdar, “A new significant area: Emotion detection in e-learning using opinion mining techniques,” proc. of 3rd IEEE International Conference on Digital Ecosystems and Technologies, 2009, pp. 259-264.

[11] A. Kołakowska, “Recognizing emotions on the basis of keystroke dynamics,” Proc. of the 8th International Conference on Human System Interaction, 2015, pp.291-297.

[12] A. Kołakowska, “A review of emotion recognition methods based on keystroke dynamics and mouse movements,” Proc. of 6th International Conference on Human System Interaction, 2013, pp.548-555.

[13] H. Gunes, M. Piccardi, "Affect Recognition from Face and Body: Early Fusion versus Late Fusion," Proc. IEEE International Conference on Systems, Man and Cybernetics (SMC ',05), pp. 3437-3443, 2005.

[14] T. Partala, A. Kallinen, “Understanding the Most Satisfying and Unsatisfying User Experiences: Emotions, Psychological Needs, and Context”. Interacting with Computers, 24, 1, 2012, pp. 25–34.

[15] K. Hone, “Empathic Agents to Reduce User Frustration: The Effects of Varying Agent Characteristics,” Interacting with Computers, 18, 2, 2006, pp. 227–245.

[16] K. Höök, “User-Centred Design and Evaluation of Affective Interfaces,” From Brows to Trust, Springer, 2005, pp. 127–160.

[17] L. Chittaro L., R. Sioni, “Affective Computing vs. Affective Placebo: Study of a Biofeedback-Controlled Game for Relaxation Training,” International Journal of Human-Computer Studies, 72, 8–9, 2014, pp. 663–673.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1681–1684
DOI: 10.15439/2016F127
ACSIS, Vol. 8. ISSN 2300-5963

Eye-tracking Web Usability Research

Paweł Weichbroth, Krzysztof Redlarski, Igor Garnik
Gdańsk University of Technology
Faculty of Management and Economics
Department of Applied Informatics in Management
ul. Narutowicza 11/12 80-233 Gdańsk, Poland
Email: pawel.weichbroth, krzysztof.redlarski, igor.garnik@zie.pg.gda.pl

***Abstract*—In this paper we present the results of a study that aims to evaluate the usability of three selected web services, based on eye–tracking and thinking aloud techniques. The gathered comments and observations, recapitulated and supported by particular measures, allow us to discover and describe typical user behavior pertaining to given tasks to solve.**

I. Introduction

WEB portals are one of the main sources of information and news. These were some of the first information services that nearly a quarter of a century ago appeared on the Internet. Initially, they delivered information in a similar way to newspapers, but they also contained links to other subject-related services. However, the most important role of a portal is still to provide timely and reliable information, relevant to Internet users' needs. Unfortunately, as shown by the results of the research presented in this paper, users often have trouble finding even basic information that, due to the nature of the service, should be available there.

It may be noted that Internet users seek information in two ways: if the type of information is sought occasionally (e.g. only once), then they use a search engine, such as Bing, Google or Yahoo [1]. In contrast, when they frequently seek information of the same type (e.g. exchange rates on the current day), then they use proven and trusted web services, such as Internet portals [2].

Over the years, we can observe how, under the influence of development of technology and design trends, the appearance of web pages and the way information is exhibited is changing. User behavior patterns of searching for information are changing as well. These patterns are affected by many factors (which enforces a different way of seeking information), e.g. the more intensive use of mobile devices, the age or even the gender of users [3, 4]. It was the basis for the research enabling the observation and description of patterns that can be found among current users, and also the determination of the reasons why access to some types of information is difficult (or even unavailable) and how to prevent this.

II. Related Works

Montero *et al.* [5] showed that the essence of design patterns is to capture the design experience in such a form which can be used effectively and repeatedly. However, adapting patterns is not a straightforward task, because the designer must demonstrate expertise and flexibility in the design process and operate on a high level of abstraction. The authors proposed conceptualizing web design pattern knowledge into the form of an ontology. The depicted generic hypermedia model, which describes the structure of a hypermedia application, includes such elements as: *node* (an information holder, e.g. a web page, a frame, a pop-up window), *content* (a piece of information, e.g. a text, a binary file or an executable application), *link* (a connection between two or more nodes or contents) and *anchor* (a source or target reference to an internal or external content or node). Design pattern format, usually expressed in natural language, is typically specified by such attributes as: *name* (an unambiguous identifier), *category* (used to classify the pattern body based on several criteria, like purpose or scope), *problem* (outlines the adaption scenario), *solution* (outlines the desired outcome) and *related-patterns* (depicts equivalents).

Based on Jakobson's communication model, Thorlacius [6] introduced and elucidated a visual communication model, where web-specific aspects in terms of navigation and interaction were taken into account. The model consists of the six following factors: *product*, *context*, *medium*, *code*, *the addresser* (actual or implicit) and *the addressee* (actual or implicit). The product is both the content and the form, along with two communication functions: formal and sublime aesthetic. The first one is the concept of visual symbols, considered in terms of colors, illustrations, typography and design in accordance with modern conventions of website layout; it reflects user experiences that contribute to "good look and feel". The second function is the question of when and where to use visual elements like flash animations and expressive illustrations, and whether innovative design should be applied; it arises from "the space between the known and the unknown" and is harder to describe in detail. The context refers to the featured (core)

content of the product. The medium is the connecting link between the addresser and the addressee in order to establish a communication channel. The code is a system of signs where each unique sequence returns a different meaning, which is presumed known to both sides. The addresser is a single person or persons responsible for the content being published on the website, where the actual is the only person who speaks for himself about the real intentions which lay behind the website, and the implicit is directly personified by visible means on the website. Finally, the last factor in the cited model is the addressee, who is the content receiver, who can actually experience the product (actual) or who can be identified through an appearance analysis (implicit).

III. Experimental Setup

A. Participants

43 participants were involved in the experiment: 21 males and 22 females, with the age average of 22.53, all students of the Gdansk University of Technology, the Faculty of Management and Economics.

B. Apparatus

In our research, eye movements (saccades, fixations, pupil diameter) were recorded with the infrared camera-based Tobii TX300 eye-tracker system, where the light source and camera are permanently affixed to a monitor. It has a 300Hz sampling frequency, and the tracking technique is "dark pupil". We used Tobii Studio software working under MS Windows7 (x64).

During the experiment the voice conversation between the moderator and participant was recorded and we made handwritten short notes of comments and remarks, and evidenced the results of tasks given to solve. Tobii Studio allowed us only to use Internet Explorer as a web browser, from this reason we were not able to examine other web browsers.

C. Objects and stimuli

The set of objects consists of three different web portals that bring information together from diverse topics and sources in a uniform way. Onet and Wiadomosci24 are the most recognizable Polish-language web portals in Poland, along with the English-language BBC. All of them are information-aware, where up-to-date news, gossip and advertisements play a major role. The context surrounding the examined websites is largely the same, where the top three mainstream contexts can be distinguished: political, social and economic. However, they differ significantly if the context scope is taken into consideration. The addressers are a group of people involved in developing, maintaining and promoting particular sections of a website, in cooperation with journalists that comment on the current information stream, carry out interviews, and report events and facts on the site. The addressees are generally users who purposely request access to website resources, using a suitable electronic device (e.g. personal computer, tablet or smart phone).

D. Procedure

The procedure consisted of 18 steps (Fig. 1), including 8 instructions (I), 3 questions (Q) and 6 tasks (T). The first instruction (I1) was a welcome screen and briefly described the purpose of the research. Next, in three questions (Q1-Q3) we asked the participant respectively about their sex, age and English language skills. Each of the second to the seventh instructions (I2-I7), preceded and described the subsequent tasks (T1-T6), and the last instruction (I8) was an acknowledgment of participation.

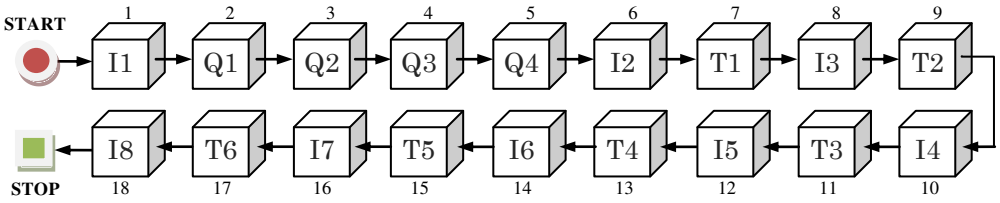


Fig. 1. The research procedure sequence

There were two separate tasks defined: the first was to find the current weather in Gdansk and the second was to find the exchange rate of Polish currency to the Euro. The maximum time allowed per task was 60 seconds. Tasks T1 and T2 were performed on Onet, T3 and T4 on the BBC site and T5 and T6 on Wiadomosci24. We decided to use the thinking aloud technique during the test to help in understanding the manner of each participant's behavior while performing the task.

E. Eye movement analysis

A quite apparent assumption laid the foundation-stone for the concept of areas of interest: *informative regions of a scene receive more of an observer's attention than the less informative regions do*. In this context, the qualifier "informative" reflects the degree of how much that particular area contributes to the understanding (perception) of the scene of sight from which the area is taken. However, a separation of such areas still poses many challenges for the development and research community. In Tobii Studio, AOIs are created by drawing ellipses, rectangles or polygons over any type of stimuli to get statistics for eye-tracking metrics for one or more recordings.

IV. Results

Quantitative and qualitative methods to collect data were concurrently used that respectively correspond to the eye-tracker device and thinking aloud technique. The quantitative analysis (heat map method) helped to identify the main areas of the participants' interest which drew their attention the most.

The analysis of the Onet service (Fig. 2) shows the largest concentration of the participants' sight on the information in the upper right corner of the site. The influencing factor for this might be the fact that the participants declared the greatest knowledge of this web service.

Fig. 2 The Onet heat map

In case of the BBC service (Fig. 3) we observed an increased interest in the upper part of the service. This was the location of the menu bar which was thoroughly followed by the participants.

Fig. 3. The BBC heat map

Interestingly, the participants also declared the lowest knowledge of the BBC service. This means that their behavior was the most natural and intuitive, and that there were no disturbance factors, such as prior acquaintance with the site.

The analysis of the Wiadomosci24 heat map (Fig. 4) allows for the observance of an increased concentration on the largest graphic and a field in the upper right corner of the service. The interest in the field (in the upper right corner) could result from the order of task execution.

Fig. 4. The Wiadomosci24 heat map

In this case the participants sought required information in the similar areas like in the case of Onet service. Such behavior was probably a result of prior connotations and users' experience with previously tested websites.

In addition, the quantitative analysis involved the reconstruction of video records and the verification of handwritten notes from the study. The data analysis helped to identify three dominant patterns of behavior among users in terms of how they try to reach the necessary information:

- *search tool*, a method based on finding the website search tool, which leads the user to the desired information by entering a user keyword;
- *menu bar*, a method based on finding the options bar or toolbar; analogically to the software application, the user tried to use a systematized and categorized tool (e.g. the menu bar), which should lead the user to the desired information;
- *home page*, a method based on searching the website's home page content for information; the user that decided to search the information service expected that they would find information on the home page.

Table I shows how often the particular method of

Table I
The Results Summary Of The Conducted Observations

Method of search information	Onet		BBC		Wiadomosci 24		AVG
	T1	T2	T1	T2	T1	T2	
Search tool	2	5	25	19	27	20	16.3
Menu bar	15	25	30	42	39	40	31.8
Home page	41	42	39	40	39	37	39.7
Number of participants	43	43	43	43	43	43	-

information search were used in the particular web service. The results of the first task showed the differences among the participants in the selection sequence of a particular model of behavior. Some of the participants decided to seek information through the menu bar, some by analyzing the contents of the home page, and some by using the search tool (both available in the information service, as well as a built-in web browser). Such behavior patterns may have many origins: previous experience in use of the particular service or knowledge of the language in which the content is presented (the participants declared different level of English language skills). These are certainly factors requiring further research on their impact on the behavior of users of information services in their search for desired information.

In the second stage of the analysis of the results of these studies, information for previously identified ways to search was verified. For this purpose, we defined areas of interest (AOI) for each of the services. They included a specific part of the site which contained information belonging to one of three identified groups ("search tool", "menu bar", "index/home page").

The quantitative analysis of models of user behavior, in the context of the search for information in each individual information service, was carried out using the following variables: *time to first fixation*, *fixations before the participant appears in the AOI, fixation duration*, *fixation count*, *visit duration*, *total visit duration*, *percentage fixated*, and *percentage clicked*.

An analysis showed significant differences regarding the usability of the home site, as well as allowed to identify changes in the behavior of tested users, depending on the home page services. In the case of the Wiadomosci24 service the study participants focused their sight on a two-column layout. The structure of other services resulted in a single-column data analysis. The reason for this and no other particular behavior was probably the difference in the system frame, which in the case of Wiadomości24 was more diffuse. The participants enjoyed situations when sought information was displayed directly on the screen and it did not require the user to additionally scroll through. However, the lower the content was on the home site, the decreasing level of interest was observed. The Wiadomosci24 service was a kind of exception. In this case, the participants analyzed the content available at the bottom of the website for much longer. The reason for this was probably the form that resembles the structure of a typical “toolbar menu”. For the rest of the websites there was the same trend (as in the case of the Onet and BBC services) – the lower the position of the content, the less the interest of the user.

V. Conclusions

The aim of the study was to analyze the typical behavior of users of web services in the search for information. The research helped to identify not only common patterns and behaviors that occurred among users. The first stage of the analysis of the study results allowed typical patterns to be identified, which consisted in an attempt to find the desired information through the search tool, the menu bar, or an analysis of the home page. Then quantitative analysis confirmed the foundations of the first stage of data analysis and identified probable causes of the different behavior of study participants.

All participants completed the first three tasks, whereas none of them were able to complete the last three. We purposely designed a set of tasks in such a manner where the first part was solvable while the second was unsolvable. Such a scenario aimed to stimulate participants to comment on obstacles and constraints, which eventually led to evaluate the usability of a particular service on the one hand, while on the other hand, random surfing progressively and subconsciously increased their concentration and cognition in order to solve given tasks. In consequence, we observed different kind of reactions: surrender (task abandonment after a few failed attempts), impatience (increasing tension and distraction over time) and self-inventiveness (out-of-the box actions, i.e. taking advantage of web browser functionality, opening new tab windows, simultaneously using well-known search engines).

To the best of our knowledge, eye-tracking analysis combined with the thinking aloud technique can provide valuable guidance to designers on the construction of information services and a starting point for further refinement of usability. In the near future, we plan to evaluate different structure patterns of websites and investigate the scale of interactions between three different groups of artefacts (i.e. standards, interface schema and data flow diagrams) in order to embody modifications included in subsequent prototypes, developed in Python. Moreover, we will verify the usefulness of the evaluation matrix template, introduced in [7].

References

[1] B. Sparrow, J. Liu, and D.M. Wegner, Google effects on memory: cognitive consequences of having information at our fingertips. Science, vol. 333(6043), 2011, pp. 776–778.

[2] E. Go, K. H. You, E. Jung, and H. Shim H, *Why do we use different types of websites and assign them different levels of credibility? Structural relations among users’ motives, types of websites, information credibility, and trust in the press*. Computers in Human Behavior, vol. 54, 2016, pp. 231–239.

[3] S.-Y. Chen, and J.-Y. Tzeng, *College female and male heavy internet users’ profiles of practices and their academic grades and psychosocial adjustment*. Cyberpsychology, Behavior and Social Networking, vol. 13(3), 2010, pp. 257–62.

[4] J. Pokrywczynski, and J. Wolburg, *A psychographic analysis of Generation Y college students*. Journal of Advertising Research, 2001, pp. 33–52.

[5] S. Montero, P. Díaz, and O. Aedo, *Formalization of web design patterns using ontologies*. In: Advances in Web Intelligence. Springer, Berlin 2003, pp. 179–188.

[6] L. Thorlacius, *Visual Communication in Web Design – Analyzing Visual Communication*. In: International Handbook of Internet Research, Springer, 2010, pp. 455–476.

[7] P. Weichbroth, and M. Sikorski, *User Interface Prototyping. Techniques, Methods and Tools*. Studia Ekonomiczne. Zeszyty Naukowe Uniwersytetu Ekonomicznego w Katowicach, 2015, pp. 184–198.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1685–1692
DOI: 10.15439/2016F390
ACSIS, Vol. 8. ISSN 2300-5963

Mouth features extraction for emotion classification

Staniucha Robert* and Wojciechowski Adam†
Institute of Information Technology
Lodz University of Technology
Wólczańska 215, 90-924 Lodz, Poland
Email: *800671@edu.p.lodz.pl, †adam.wojciechowski@p.lodz.pl

***Abstract*—Face emotions analysis is one of the fundamental techniques that might be exploited in a natural human-computer interaction process and thus is one of the most studied topics in current computer vision literature. In consequence face features extraction is an indispensable element of the face emotion analysis as it influences decision making performance. The paper concentrates on classification of human poses based on mouth. Mouth features extraction, which next to eye region features becomes one of the most representative face regions in the context of emotions retrieval. Additionally, in the paper, original mouth features extraction method was presented. It is gradient based. Evaluation of the method was performed for a subset of the Yale images database and classification accuracy for single emotion is over 70%.**

I. Introduction

VISUAL determinants revealing human emotions comprise [1]: emotional voice, body pose, gestures, gaze direction end facial expressions. Thorough and complete human emotion analysis should consider also state of human environment, both current and passed, that may originate emotions. Typical spontaneous, emotions accompanying, muscular activity last usually between 250 ms and 5s – very rarely longer [1]. As a result not only location of human action is important, but intensity and its dynamics as well. In 1971 Ekman and Friesen [2] postulated 6 basic emotions that reveal distinctive set of features with unique facial expressions. These are happiness, sadness, fear, disgust, surprise and anger. Face poses visual recognition or even monitoring may have considerable influence not only on building playful and intelligent social living environment, control employing nonexpert workers (i.e. crowdsourcing [3]) but also on our security as well (i.e. video surveillance system).

Surveying facial expressions, mouth region seems to be the most representative next to the eyes region. This paper concentrates on visual mouth features extraction which have key influence on further facial expression and emotion detection analysis. The new approach bases on local intensity gradients and subsequent dedicated impulse filter responses. Precise extraction of mouth features can increase the efficiency of a classifier by decreasing its complexity. Authors present a new, alternative method of mouth characteristics extraction and evaluate effectiveness of the method on the well known and reliable Yale faces database [4]. The paper shows also classification results based on elaborated method for common used classifiers.

II. Related work

Biologically facial expressions are generated by contractions of facial muscles, which cause face features temporal deformations. The most evident changes concern eye lids, eye brews, nose, lips or skin possible wrinkles. Facial expression intensity can be measured by geometric deformation of facial features or facial texture analysis, i.e. density of face appearing wrinkles.

Among anthropologically justified face core landmarks forming unquestionable framework for face features extraction researches [5] proposed 11 points: pronasale, alare (left & right), subnasale, chelion (left & right), endocanthion (left & right), exocanthion (left & right) and sellion (fig. 1). Subsequently some minor landmarks can be evaluated.

Face features extraction methods, presented in literature, can be classified into two main groups: appearance based and geometric based methods [6]. Though the appearance based approach seems to be currently the most popular, the geometric based methods seem to be recently neglected but still promising. Even though the geometric approach seems to be very well studied [7], [8], according to Pali [9], among the most evident aspects that can be still improved, within face features extraction, there are dimensionality reduction, features extraction techniques and features subset selection. Thus, the geometrical approach can almost automatically reduce space dimension problem and behave more reliably in demanding scenarios where pose (in-plane and out of plane face rotations) and illumination are not controlled [10].

Face features detection is a challenging task as, due to the quantity of local face image structures, classical corner detectors are useless without considering their context. Researches estimating face inherent features and edges date back to well known Kanade work [11] and were further intensively developed (i.e. [12]). Authors attempted to reconstruct horizontal and vertical lines applying Laplacian operator and evaluating horizontal and vertical integral image projection obtaining effectiveness of about 75% on dedicated, self-prepared databases.

Castrillon [10], Castille [13] and Yang [14] suggested Viola-Jones object detection algorithm for coarse face parts (i.e.: eyes, mouth, nose) localization but it did not localize precisely face landmarks and required further, more detailed features detectors. Even Panning [15] and Wang et al. [16] approaches, measuring distances between face regions, detected with originally elaborated features or Lienhart [17] extended Haar-like

978-83-60810-90-3/$25.00

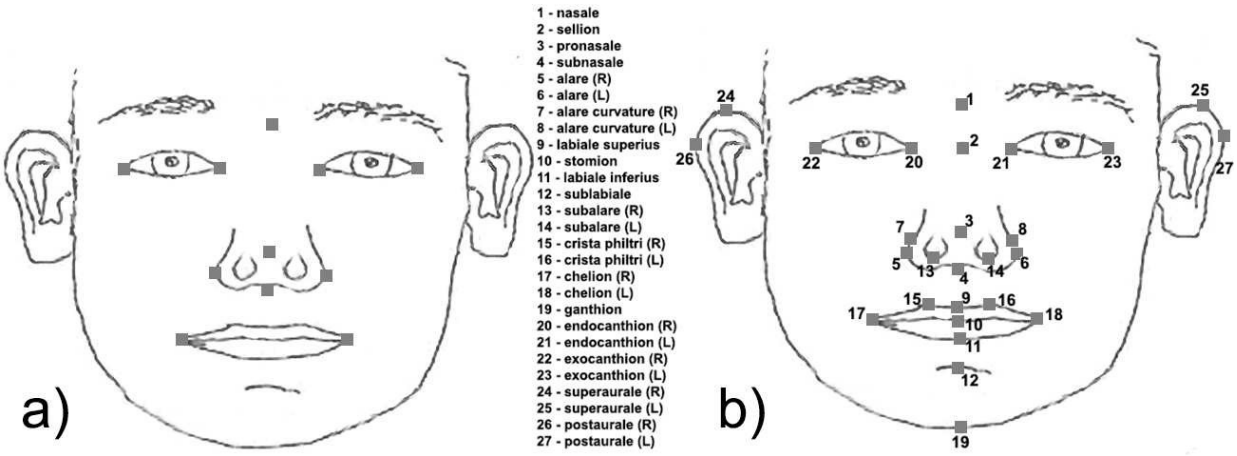


Fig. 1. a) Set of 11 anthropological landmarks (small square) used for face identification and face expression analysis b) Minor landmarks (small square) to improve accuracy

features, required consequent holistic extensive well trained classification.

The process of face landmark thorough detection should be introduced by an appropriate in-image face detection. Subsequently extracted landmarks can be used then for facial expression classification. Face localization and general face features extraction can be performed in tances between face regions, detected with originally elaborated features or Lienhart [17] different manner. Common solutions encompass deploying of a sequence of Haar-like features according to Viola-Jones algorithm [13], [18], eigenspaces [19] also referring to 3D model [20], skin color segmentation [21], statistical methods [22], [23] or active contour or shape models [24], [21].

Some authors [25], [26] exploited local binary patterns (LBP) idea for face image features extraction. Their extension: Local Direction Patterns (LDP) and locally assembled binary (LAB) features [27] appeared to be quite efficient for face edges detection and partly inspired authors of this paper for gradient distribution analysis.

If dedicated mouth features are further required, within face region analysis should be performed. Kim [28] and Chien [29] analyzed grid-based and coordinate-based lips features (width/height of outer/inner lips edges) for Korean language words recognition support. Matthews [30] exploited AAM and ASM for mouth visual shape description for lips reading. Shen [31] and Lewis [32] analyzed color space for lip features retrieving. He et al. [33] proposed modified Biologically Inspired Model of face features extraction improving the SVM classification of face smile. Su [34] suggested geometrical and Gabor filter retrieved face features fusion for better facial expression recognition.

Aforementioned approaches, though well studied and elaborated, lack of generic simplicity which lies in mouth lines detection. Presented method robustly extracts simplified lips edges by means of originally elaborated gradient-based approach which can be subsequently interpreted. Presented solution provides a representative set of features for mouth classification and consequent facial emotions recognition. It was additionally tested on a subset of Yale faces database [4]. It uses many of commonly used classifiers to show that presented method provides sufficient information which enables the detection or identification of face emotions.

III. Mouth features extraction method

To detect mouth shape features we need to find human face and relative localization of the mouth. Next, we can try extract mouth shape from it. Aggregated mouth features extraction process can be completed within subsequent steps:

1) Finding face,
2) Finding mouth on the face,
3) Mouth segmentation,
4) Features extraction.

In image face localization can be performed by means of Haar-like features method [17], but it will not be described here, because it is a part of another problem. Localization of mouth within the face region is a similar problem and can be completed with an analogical set of Haar-like features. That is why further, core mouth features analysis, assumes that face image is cropped to mouth with some border, as shown in fig. III

Mouth segmentation is then performed in several steps. These can be described as:

1) Gradient calculation,
2) Resulting image normalization,
3) Resulting image filtering,
4) Resulting image thresholding,
5) Noise removal.

To retrieve information from mouth images, at the first step **gradient** should be calculated. Gradient of an image is calculated with formula 1.

$$\nabla f = \begin{bmatrix} g_x \\ g_y \end{bmatrix} = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix} \tag{1}$$

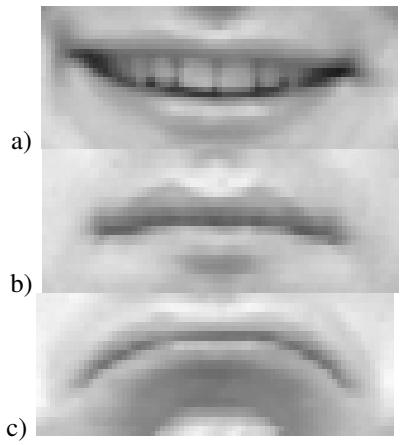


Fig. 2. Example of mouth images: (a) happy, (b) neutral, (c) sad, for subject no. 1

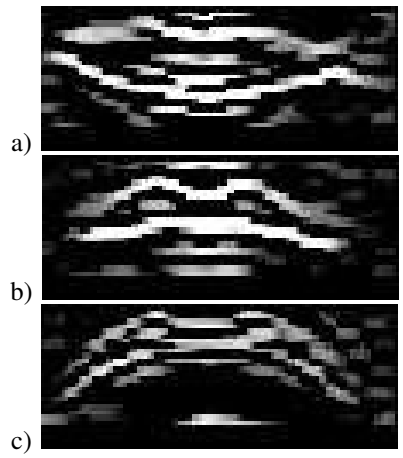


Fig. 3. Example gradient for exemplary mouth images: (a) happy, (b) neutral, (c) sad, for subject no. 1

During the experiments, it was noticed that for descent mouth retrieval there is no need to calculate the whole gradient, but only its vertical part. Using both dimensions of gradient does not give noticeable results improvement, thus, in the segmentation process there was used only vertical component.

To calculate a vertical gradient g_y of image A we can use simple filter, as it is shown in equation 2.

$$g_y = \frac{\partial f}{\partial y} = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} * A \tag{2}$$

Additionally, extended 3x3 matrix (eq. 3) was used to reduce noise, which introduces additional column on the left and on the right side of pixel position.

$$\begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \Rightarrow \begin{bmatrix} -1 & \mathbf{-1} & -1 \\ 0 & \mathbf{0} & 0 \\ 1 & \mathbf{1} & 1 \end{bmatrix} \tag{3}$$

The first step results in images with vertical gradient calculated over the whole their area. Additionally, gradient values were squared in order to level up output and remove useless noise (fig. 3).

$$A = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \quad B = \begin{bmatrix} -3 & -10 & -3 \\ 0 & 0 & 0 \\ 3 & 10 & 3 \end{bmatrix}$$

$$C = \begin{bmatrix} -1 & -1 & -1 \\ 2 & 2 & 2 \\ -1 & -1 & -1 \end{bmatrix}$$

Fig. 4. Filter matrices used for edge extraction

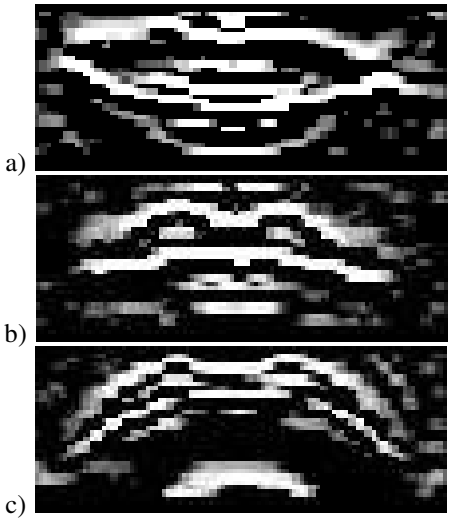


Fig. 5. Exemplary outputs obtained with matrix B from fig. 4: (a) happy, (b) normal, (c) sad, for subject no. 1

The next image processing step was the gradient **normalization**. MIN-MAX type normalization was used to adjust gradient to image full spectrum of brightness.

In the next step gradient image filtering was considered as to extract edges from it. Various filters was tested to extract shapes from images. The best results were obtained by filters represented in fig. 4. Corresponding outputs obtained with matrix B are presented in fig. 5.

After the filtering step a process of **thresholding** was applied to extract shape of lips. It was done by simple cut off image value below certain threshold of pixel brightness. Results were verified for a few threshold values, but overall the best was achieved with 250. Example results obtained with different values of threshold are shown in fig. 6

The subsequent step of the proposed method is the noise reduction. It was achieved by morphological operations performed on image. The best noise reduction was obtained for closing, which is a combination of erosion and dilation. In result shape of mouth was closed as it is shown in fig. 7.

The last stage concerned segmented image features extraction. Mouth corners were selected as initially considered features. They were found by the most extreme edges in all directions: down, left, right and up.

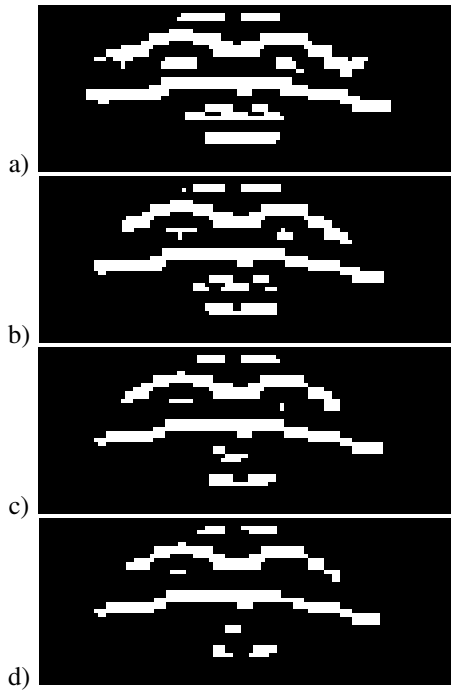


Fig. 6. Example results obtained with different threshold values: (a) 100, (b) 150, (c) 200 and (d) 250, for subject no. 1

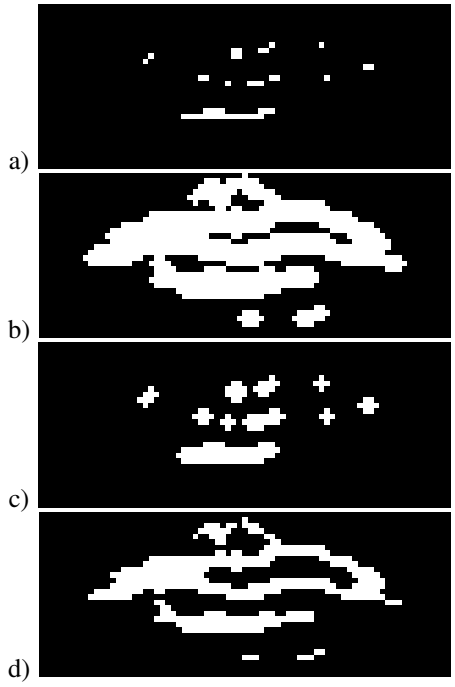


Fig. 7. Example results with different noise reduction: (a) erosion, (b) dilatation, (c) opening, (d) closing, for subject no. 12

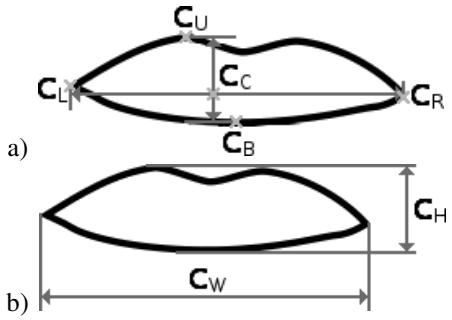


Fig. 8. Feature extraction corners: (a) positions, (b) sizes, for subject no. 1

Edge point c_e is defined as the farthest point in specified direction, as is shown in equation 4.

$$c_e = (x, y) \tag{4}$$

where

$$e = \{ \texttt{B} - \text{bottom}, \texttt{L} - \text{left}, \texttt{R} - \text{right}, \texttt{U} - \text{topmost} \}$$

In case of the left edge $c_{\texttt{L}}$, it is the most distant left point of the largest contour area. $c_{\texttt{R}}$ can be defined analogously, but for the right edge. More difficult is to find the bottom and topmost edge points: $c_{\texttt{B}}$ and $c_{\texttt{U}}$, because often the contour is composed of several separated pieces, so we need to check if the contour size is sufficiently large in relation to the image size. Calculation of the shape size was simply done by following each white pixel connected to initially located edge points ($c_{\texttt{L}}$ and $c_{\texttt{R}}$). In case of contour inconsistence, averaged extreme values (bottom and topmost), retrieved from two separate edges, anchored independently from left and right corners were calculated.

The farthest edges are marked with white and light grey dashed lines. Brighter color is used to mark part of shape with features determining edges (fig. 8).

The collected information was sufficient to easily determine the next two parameters: $c_{\texttt{W}}$ and $c_{\texttt{H}}$. The first one determines the maximum distance stretching horizontally the mouth area, between $c_{\texttt{L}}$ and $c_{\texttt{R}}$, on the average height between them. Similarly value of $c_{\texttt{H}}$ was determined.

With a **cross-section** $c_{\texttt{C}}$ of width $c_{\texttt{W}}$ and height $c_{\texttt{H}}$, it is possible to interpret shape of the the mouth. Cross-section is always defined, because interpreted shape corners c_e are also determined.

Additionally, c_{h_i} as height of some part of mouth shape can be extracted similar to $c_{\texttt{H}}$, but calculated as height between the top most point in shape and the lowest one for some i position horizontally.

IV. Features extraction

To evaluate the method a subset of the Yale's face database was used. It comes from UC San Diego Computer Vision [4]. It contains 165 grayscale images of 15 people, originally sized 320x243 pixels, but had manually cropped mouth region with size of 75x30 pixels. Each subject had several images in different face expressions, quality and light conditions, but for

all people those conditions were the same. It allowed us to test elaborated method in various conditions of image quality.

Each image was marked by expert as shown in fig. 9: shape of mouth was built out of 6 points and it was marked by expert with white line, maximum width and maximum height. Marked elements were measured and noted as m_e, i.e. width as $m_{\mathtt{W}}$, analogically to computed values $c_{\mathtt{W}}$ from authors' method.

Exemplary results are shown in fig. 10. In analogy to the reference images, maximum widths (red) and heights (blue) are marked over the image.

Mouth key features evaluation method results c_e were subsequently compared with reference, expert marked points m_e (labeled data). For i-th image, each of key features c_e estimation accuracy ACC was calculated according to equation 5. For cross-section feature $c_{\mathtt{C}}$ estimation accuracy ACC was calculated according to equation 8, where `distance` is Manhattan distance between points and $\mathtt{distance}_{MAX}$ is the maximum possible distance on the image.

$$ACC_i(c_e) = \frac{\min\{c_e.x; m_e.x\} \times 100\%}{\max\{c_e.x; m_e.x\}} \text{ where } e = \{L, R\} \quad (5)$$

$$ACC_i(c_e) = \frac{\min\{c_e.y; m_e.y\} \times 100\%}{\max\{c_e.y; m_e.y\}} \text{ where } e = \{B, U\} \quad (6)$$

$$ACC_i(c_e) = \frac{\min\{c_e; m_e\} \times 100\%}{\max\{c_e; m_e\}} \text{ where } e = \{W, H\} \quad (7)$$

$$ACC_i(c_c) = 100\% - \frac{(\|c_c.x - m_c.x\| + \|c_c.y - m_c.y\|) \times 100\%}{75 + 30} \quad (8)$$

In table I selected features positioning accuracy results are presented. The results were calculated as averaged value of selected ($n = 66$) Yale database images individually estimated accuracies $ACC_i(c_e)$ (eq. 9).

$$ACC(c_e) = \frac{\sum_i ACC_i(c_e)}{n} \quad i \in \{1, 2, \ldots, n\},\ e = \{L, R, B, U, W, H, C\} \quad (9)$$

As extracted feature vector of single image was tuple of $c_L, c_U, c_R, c_B, c_W, c_H, c_C, c_{h_0} c_{h_n}$, where c_{h_i} means additionally height for a few positions, equally distributed over width of mouth. For tests n was 7.

The highest results of extraction accuracy were reported for left (81.42%), bottom (88.29%) and right (95.53%) part of mouth features detection. More problematic was upper part, what is strongly related to differences received in gradient of mouths. Resulting emotion characteristic determinants: mouth width (92.78%), height (72.77%) and cross-section point (86.18%) revealed also high evaluation accuracy. Weak mouth upper part features estimation influences negatively effectiveness of mouth shape determination, but can be counterbalanced by adding a number of measuring points for the lower edge and the height of shape characteristics measured for several positions, not only centrally located.

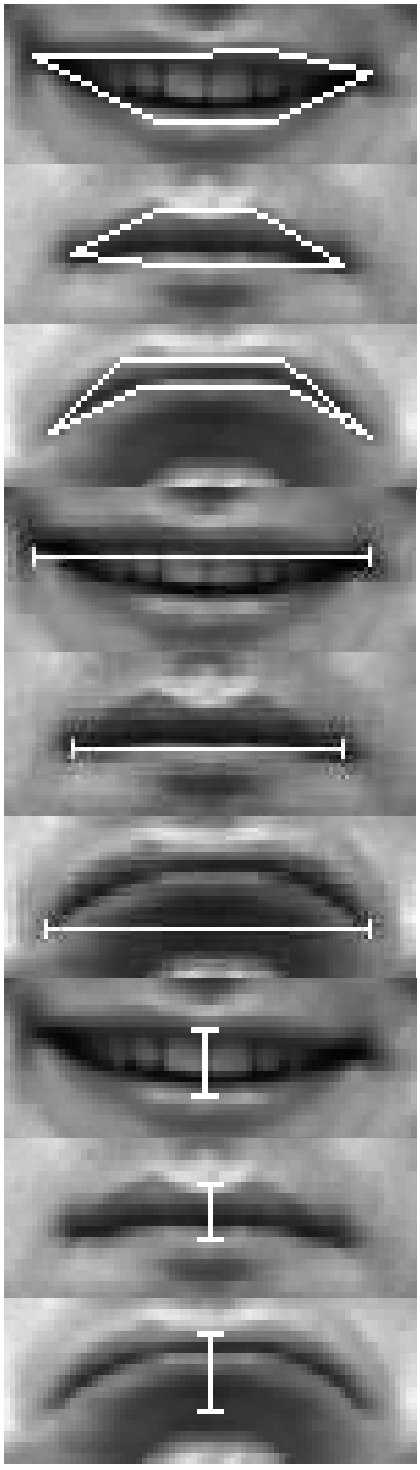


Fig. 9. Example labeled dataset with marked with white lines: shape, width and height

TABLE I
MOUTH FEATURES RECOGNITION RESULTS

Measurement	Size		Position				
	$ACC(e_W)$	$ACC(e_H)$	$ACC(e_L)$	$ACC(e_R)$	$ACC(e_U)$	$ACC(e_B)$	$ACC(e_C)$
Result [%]	92.8	72.8	81.4	95.5	43.6	88.3	86.2
Std dev. [%]	7.1	17.9	18.4	3.8	25.5	13.1	9.4

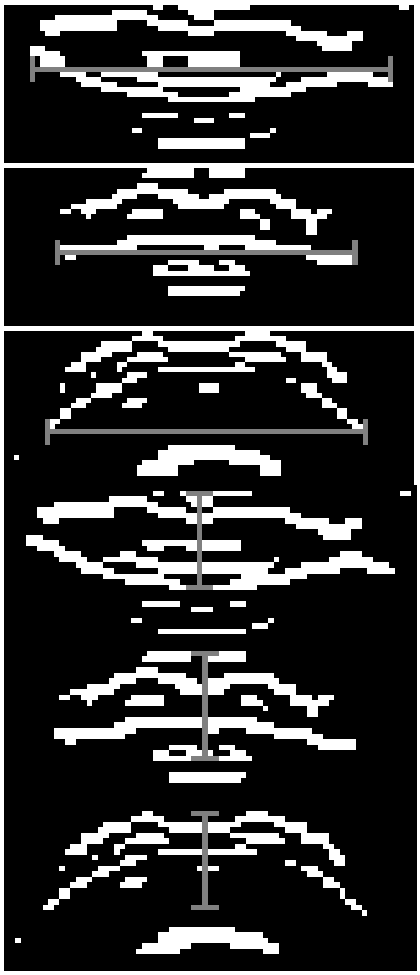


Fig. 10. Example results with marked with white lines: width and height

V. EMOTION CLASSIFICATION

After features' set extraction their classification is performed. Classifier predicts a label – one class from a few available. It is based on mathematical evaluation of input vector, due to selected classifier, and as output it has received a number, which represents assignment of input vector to some class [35], [36].

4 types of classical one-label classifiers were used for classification: Random Trees, *k*–Nearest Neighbors, Multi–Layer Perceptrons and Support Vector Machines. These are well known and widely used classifiers. There also others approaches like statistical methods based on Bayesian modelling combined with a Markov chain Monte Carlo (MCMC) sampling algorithm [37] etc., but was not considered in this paper. Weka was used for implementation, as a good base to train and test various classifiers with different options – explorer module [38].

At first, Random Trees (RT) was used, which was originally described to resolve classification and regression problems [39]. It constructs a tree that considers $log_2($ predicators $)+1$ randomly chosen attributes at each node, it doesn't perform pruning. It has used estimation of class probabilities basing on a hold-out set (backfitting).

K-nearest neighbors (*k*NN) classifier calculates distance between feature vector of input vector and others feature vectors from training set [40]. It has used Euclidean distance metrics, $k = 3$ and distance weighting with equation: $1/$distance.

Multi-Layer Propagation is a classifier that uses backpropagation to classify instances. This network is built by an algorithm: input layer = input features vector, hidden layer = input layer / 2 and output layer = number of classes. The nodes in this network are all sigmoid, except for output when the class is numeric in which case the nodes become unthresholded linear units.

Support Vector Machine (SVM) is a discriminative classifier formally defined by a separating hyperplanes. It has used linear kernel and C–SVC, because it is commonly used for n-class classification. In this research the libsvm library [41] was used.

The effectiveness of all methods was measured by Classification Accuracy (CA) as the basic evaluation measure. This measure provides a very precise evaluation because it reflects relation between set of correct labeled instances and all existing ones in the training set. It is defined as:

$$CA = \frac{1}{N}\sum_{N=0}^{N} y_i = f(x_i) \tag{10}$$

where: x_i are instances, i = $1..N$, N is their total number in the test set, y_i denotes the label of x_i and $f(x_i)$ is predicted label during classification process.

The experiments were carried out on popular database, which is often used to test methods working on face images. From selected images features vectors were extracted, which were used for classification.

Extracted features from previous section were divided into three groups of datasets:

1) Single – which was used for binary classification. All instances of features were labeled to belong to single class or not, for example instance can be classified as

TABLE II
EMOTION BINARY CLASSIFICATION WITH MOUTH FEATURES

	RT [%]	KNN [%]	MLP [%]	SVM [%]
happy	80.3	87.9	88.9	89.4
normal	71.2	75.8	75.8	83.3
sad	77.3	86.4	80.3	83.3
sleepy	75.8	77.3	81.8	83.3
surprised	72.7	86.4	84.8	83.3
wink	74.2	77.3	78.8	83.3

TABLE III
EMOTION CLASSIFICATION WITH MOUTH FEATURES

	RT [%]	KNN [%]	MLP [%]	SVM [%]
Three	53.0	37.9	37.9	45.5
Three2	30.3	31.8	28.8	30.3
Six	24.8	27.3	25.8	18.2

happy or unhappy (opposite emotion). There were six classes: happy, normal, sad, sleepy, surprised, wink.

2) Three – instances were grouped into three simple emotions: happy, normal and sad, where surprised was added to happy and wink to sad, because their vectors were similar.
3) Six – all instances were joined together in one dataset, but original labels were left for classification. This was the hardest dataset, which shows bad separation of instances.

Additionally, second Three dataset (named Three2) was used, which contains features of image space from the last step of features extraction, to better compare results.

Results of binary classification were shown in tab. II and from multi class classification in tab. III. For this context, the best recognized pose was happy, with almost 90% of accuracy detection for single emotion detection, over 70% True Positive (TP) detection in three–poses dataset and over 50% TP in six-poses dataset, with over 50% overall detection. It was high detectable emotion in all experiments.

The next poses which were well diagnosed these are surprised and sad. Detection rate was over 86% and 50% TP in six-poses dataset respectively. Other poses had much worse results, because individual classes were tightly mixed to each other, so they were hard to detect. Almost the same situation exists for binary classification, where some of the poses were hard detectable, but overall results is based on not classified non-poses, e.g. normal emotion was non detected for few of classifiers. SVM almost perfect detects non-poses for all emotions. Generally False Positive ration was very low or 0 for all tests.

VI. CONCLUSION

In this paper, it was proposed method to extract a few facial features from mouth image to expression recognition. The experiments were carried out on popular database, which is often used to test methods working on face images. As was shown in the last section, it gives overall good results, especially in horizontal measurement and they should be enough to right description of many facial expressions. In the future work it will be possible to collect other facial features and merge them together to better understand face and identify poses.

REFERENCES

[1] B. Fasel and J. Luettin, "Automatic facial expression analysis: a survey," *Pattern recognition*, vol. 36, no. 1, pp. 259–275, 2003.

[2] P. Ekman and W. Friesen, "Constants across cultures in the face and emotion," *Journal of Personality and Social Psychology*, vol. 17, no. 2, pp. 124–129, 1971.

[3] C. Chen, P. W. Woźniak, A. Romanowski, M. Obaid, T. Jaworski, J. Kucharski, K. Grudzień, S. Zhao, and M. Fjeld, "Using crowdsourcing for scientific analysis of industrial tomographic images," *ACM Trans. Intell. Syst. Technol.*, vol. 7, no. 4, pp. 52:1–52:25, Jul 2016. doi: 10.1145/2897370. [Online]. Available: http://doi.acm.org/10.1145/2897370

[4] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE TRANS. PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, vol. 19, no. 7, pp. 711–720, 1997.

[5] S. Liang, J. Wu, S. M. Weinberg, and L. G. Shapiro, "Improved detection of landmarks on 3d human face data," in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, July 2013. doi: 10.1109/EMBC.2013.6611039. ISSN 1094-687X pp. 6482–6485. [Online]. Available: http://dx.doi.org/10.1109/EMBC.2013.6611039

[6] S. Mishra and A. Dhole, "A survey on facial expression recognition techniques," *International Journal of Science and Research*, vol. 4, no. 4, pp. 1247–1250, 2015.

[7] E. Hjelmås and B. K. Low, "Face detection: A survey," *Computer vision and image understanding*, vol. 83, no. 3, pp. 236–274, 2001.

[8] C. Zhang and Z. Zhang, "A survey of recent advances in face detection," Tech. rep., Microsoft Research, Tech. Rep., 2010.

[9] V. Pali, S. Goswami, and L. Bhaiya, "An extensive survey on feature extraction techniques for facial image processing," in *Sixth International Conference on Computational Intelligence and Communication Networks*, 2014. doi: 10.1109/CICN.2014.43 pp. 142–148.

[10] M. Castrillón-Santana, D. Hernández-Sosa, and J. Lorenzo-Navarro, "Combining face and facial feature detectors for face detection performance improvement," in *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, ser. Lecture Notes in Computer Science, L. Alvarez, M. Mejail, L. Gomez, and J. Jacobo, Eds., 2012, vol. 7441, pp. 82–89.

[11] T. Kanade, *Computer recognition of human faces*. Birkhäuser, 1977, vol. 47.

[12] R. Brunelli and T. Poggio, "Face recognition: Features versus templates," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, no. 10, pp. 1042–1052, 1993.

[13] M. Castrillón, O. Déniz, D. Hernández, and J. Lorenzo, "A comparison of face and facial feature detectors based on the viola–jones general object detection framework," *Machine Vision and Applications*, vol. 22, no. 3, pp. 481–494, 2011.

[14] M.-T. Yang, Y.-J. Cheng, and Y.-C. Shih, *Facial Expression Recognition for Learning Status Analysis*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 131–138. ISBN 978-3-642-21619-0. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-21619-0-18

[15] A. Panning, A. K. Al-Hamadi, R. Niese, and B. Michaelis, "Facial expression recognition based on haar-like feature detection," *Pattern Recognition and Image Analysis*, vol. 18, no. 3, pp. 447–452, 2008.

[16] Q. Wang, C. Zhao, and J. Yang, "Robust facial feature location on gray intensity face," in *PSIVT 2009*, ser. LNCS, T. Wada, F. Huang, and S. Lin, Eds., vol. 5414. Springer, 2009. doi: 10.1007/978-3-540-92957-4-47 p. 542–549. [Online]. Available: http://dx.doi.org/10.1007/978-3-540-92957-4-47

[17] R. Lienhart and J. Maydt, "An extended set of haar-like features for rapid object detection," in *Image Processing. 2002. Proceedings. 2002 International Conference on*, vol. 1, 2002, pp. I–900.

[18] P. Viola and M. Jones, "Robust real-time face detection," *International Journal of Computer Vision*, vol. 57, pp. 137–154, 2004.

[19] A. Pentland, B. Moghaddam, and T. Starner, "View-based and modular eigenspaces for face recognition," in *IEEE Conference on Computer Vision and Pattern Recognition, Seattle, NA, USA*, 1994, pp. 84–91.

[20] W. Yang, C. Sun, W. Zheng, and K. Ricanek, "Gender classification using 3D statistical models," *Multimedia Tools and Applications*, Mar 2016. doi: 10.1007/s11042-016-3446-7. [Online]. Available: http://dx.doi.org/10.1007/s11042-016-3446-7

[21] Y.-H. Lee, C. G. Kim, Y. Kim, and T. K. Whangbo, "Facial landmarks detection using improved active shape model on android platform," *Multimedia Tools and Applications*, vol. 74, no. 20, p. 8821–8830, Jun 2013. doi: 10.1007/s11042-013-1565-y. [Online]. Available: http://dx.doi.org/10.1007/s11042-013-1565-y

[22] H. Schneiderman and T. Kanade, "A statistical method for 3d object detection applied to faces and cars," in *Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on*, vol. 1. IEEE, 2000, pp. 746–751.

[23] M. A. Berbar, "Three robust features extraction approaches for facial gender classification," *Vis Comput*, vol. 30, no. 1, p. 19–31, Jan 2013. doi: 10.1007/s00371-013-0774-8. [Online]. Available: http://dx.doi.org/10.1007/s00371-013-0774-8

[24] M. Kass, A. Witkin, and D. Terzepulos, "Snakes: Active contour models," in *First IEEE International Conference on Computer Vision*, 1987, pp. 259–268.

[25] A. Hussain, M. S. Khan, M. Nazir, and M. A. Iqbal, "Survey of various feature extraction and classification techniques for facial expression recognition," in *Proceedings of the 11th WSEAS international conference on Electronics, Hardware, Wireless and Optical Communications, and proceedings of the 11th WSEAS international conference on Signal Processing, Robotics and Automation, and proceedings of the 4th WSEAS international conference on Nanotechnology*, 2012, pp. 138–142.

[26] T. Jabid, M. H. Kabir, and O. Chae, "Robust facial expression recognition based on local directional pattern," *ETRI journal*, vol. 32, no. 5, pp. 784–794, 2010.

[27] S. Yan, S. Shan, X. Chen, and W. Gao, "Locally assembled binary (lab) feature with feature-centric cascade for fast and accurate face detection," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, 2008, pp. 1–7.

[28] Y.-K. Kim, J. G. Lim, and M.-H. Kim, "Comparison of lip image feature extraction methods for improvement of isolated word recognition rate," Aug 2015. doi: 10.14257/astl.2015.107.14. [Online]. Available: http://dx.doi.org/10.14257/astl.2015.107.14

[29] S.-I. Chien and I. Choi, "Face and facial landmarks location based on log-polar mapping," *Lecture Notes in Computer Science*, p. 379–386, 2000. doi: 10.1007/3-540-45482-9-38. [Online]. Available: http://dx.doi.org/10.1007/3-540-45482-9-38

[30] I. Matthews, T. Cootes, J. Bangham, S. Cox, and R. Harvey, "Extraction of visual features for lipreading," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 24, no. 2, p. 198–213, 2002. doi: 10.1109/34.982900. [Online]. Available: http://dx.doi.org/10.1109/34.982900

[31] X.-g. Shen and W. Wu, "An algorithm of lips secondary positioning and feature extraction based on ycbcr color space," in *International Conference on Advances in Mechanical Engineering and Industrial Informatics*. Atlantis Press, 2015. doi: 10.2991/ameii-15.2015.271. [Online]. Available: http://dx.doi.org/10.2991/ameii-15.2015.271

[32] T. W. Lewis and D. M. W. Powers, "Lip feature extraction using red exclusion," in *Selected Papers from the Pan-Sydney Workshop on Visualisation - Volume 2*, ser. VIP '00. Darlinghurst, Australia, Australia: Australian Computer Society, Inc., 2001. ISBN 0-909-92580-1 pp. 61–67. [Online]. Available: http://dl.acm.org/citation.cfm?id=563752.563761

[33] C. He, H. Mao, and L. Jin, "Realistic smile expression recognition using biologically inspired features," in *AI 2011*, ser. LNAI, D. Wang and M. Reynolds, Eds., vol. 7106. Springer, 2011, p. 590–599.

[34] C. Su, J. Deng, Y. Yang, and G. Wang, "Expression recognition methods based on feature fusion," in *BI 2010*, ser. LNAI, Y. Yao *et al.*, Eds., vol. 6334. Springer, 2010, p. 346–356.

[35] M. Krzyśko, W. Wołyński, T. Górecki, and M. Skorzybut, "Systemy uczące się," *Rozpoznawanie wzorców, analiza skupień i redukcja wymiarowości. WNT, Warszawa*, 2008.

[36] I. H. Witten, E. Frank, and M. A. Hall, *Data Mining: Practical Machine Learning Tools and Techniques*, 3rd ed. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2011. ISBN 0123748569, 9780123748560

[37] K. Grudzien, A. Romanowski, and R. A. Williams, "Application of a bayesian approach to the tomographic analysis of hopper flow," *Particle & Particle Systems Characterization*, vol. 22, no. 4, pp. 246–253, 2005. doi: 10.1002/ppsc.200500951. [Online]. Available: http://dx.doi.org/10.1002/ppsc.200500951

[38] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The weka data mining software," *ACM SIGKDD Explorations Newsletter*, vol. 11, no. 1, p. 10, Nov 2009. doi: 10.1145/1656274.1656278. [Online]. Available: http://dx.doi.org/10.1145/1656274.1656278

[39] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001. doi: 10.1023/A:1010933404324. [Online]. Available: http://dx.doi.org/10.1023/A:1010933404324

[40] D. W. Aha, D. Kibler, and M. K. Albert, "Instance-based learning algorithms," *Machine Learning*, vol. 6, no. 1, p. 37–66, 1991. doi: 10.1023/a:1022689900470. [Online]. Available: http://dx.doi.org/10.1023/A:1022689900470

[41] C.-C. Chang and C.-J. Lin, "Libsvm - a library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, p. 1–27, Apr 2011. doi: 10.1145/1961189.1961199. [Online]. Available: http://dx.doi.org/10.1145/1961189.1961199

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1693–1697
DOI: 10.15439/2016F507
ACSIS, Vol. 8. ISSN 2300-5963

Applications for investigating therapy progress of autistic children

Agata Kołakowska, Agnieszka Landowska, Michal R. Wrobel
Faculty of Electronics, Telecommunications and Informatics,
Gdansk University of Technology, Poland
Email: {agatakol,nailie,wrobel}@eti.pg.gda.pl

Dominika Zaremba, Dominika Czajak, Anna Anzulewicz
Harimata Sp. z o.o., Poland
Email: {dominika.zaremba,domnika,ania}@harimata.co

***Abstract*—The paper regards supporting behavioral therapy of autistic children with mobile applications, specifically applied for measuring the child's progress. A family of five applications is presented, that was developed as an investigation tool within the project aimed at automation of therapy progress monitoring. The applications were already tested with children with autism spectrum disorder. Hereby we analyse children' experience with the games, as a positive attitude towards the application is the key factor enabling practical application of the solutions in therapy. The study might be interesting for both researchers and practitioners applying e-technologies in autistics therapy.**

I. Introduction

AUTISM is a developmental disorder, that influences the ability to socialize, communicate as well as learning skills. Autistic children have diverse level of deficits in language understanding, speaking and other areas, that make the therapy and education very difficult [1]. Educating children with autism spectrum disorder could be a challenge in the best of circumstances [2]. There are premises for supporting therapy with e-technologies [3], [4], as most of the autistic children are eagerly using computers and tablets once they are taught how to use them [2]. Autistics require repetitive environment for functioning and learning. Technologies are able to perform the same activities in exactly the same way and with indefinite patience. Moreover, systems and devices might be customized in order to adjust to a unique set of deficits of an individual [5].

This study is performed under the AUTMON (Automated therapy monitoring for children with autism spectrum disorder) project, that aims at development of methods and tools to allow for the automatic evaluation of the therapy progress among children with autism spectrum disorder (ASD) [6]. Therapy monitoring is based on automatic detection of behavioral patterns in tablet and application usage. During the project 5 applications were developed or adjusted to fit the deficits of autistic children. The apps have the potential of being applicable in the therapy progress monitoring. However, the crucial question is: whether the autistic children are eager to use those games. This study aims at verification of this issue. In this paper, the applications and their adaptation to autistic users is described, and the data from interaction sessions is analyzed. Moreover, we report our observations on autistics interaction with tablets and applications. The study might be interesting for both producers of solutions for autistic children as well as for researchers investigating effective methods for ASD support and therapy with e-technologies.

The paper is organized as follows: Section II reports work related to this study and especially previous works on monitoring autistic users performance. Section III presents the study context and study design. Section IV and V present the results of games evaluation, and those sections are followed by a summary of results and discussion (Section VI) and the conclusions describing future works (Section VII).

This research is supported by the National Centre for Research and Development, Poland under grant AUTMON no IS-2/6/NCBR/2015, as well as DS Programs of the Faculty of Electronics, Telecommunications and Informatics, Gdansk University of Technology. This research is also inspired by European Cooperation in Science and Technology (COST) Action TD1309 Play for children with disabilities (LUDI).

II. Related work

New technologies are used while working with people suffering from autism both at the stage of diagnosis and therapy. The use of technology in this field is usually based on software implementations for common therapeutic tasks. Supporting diagnosis is limited to the use of computer versions of diagnostic questionnaires. Although there are some advanced technological solutions that can help to diagnose autism, they are still experimental methods, used in laboratories only or tested on small groups of people. In many cases they require specialized hardware. Some of these solutions are based on: eye tracking [7], automatic analysis of children video recordings [8], automatic analysis of individual hand movements [9] and machine learning algorithms applied to predict the state of a person [10].

Much more often than in the diagnosis, new technologies are used in the therapy of autism. In most cases, these solutions are computer implementations of paper-based therapeutic tools. There are numerous programs supporting daily activities, allowing for arranging and following procedures that consist of sequences of activities represented as images or text [5].

Numerous tools are created to help to acquire emotional intelligence skills. These include programs that help in learning face and emotion recognition, learning expressing emotions through facial expressions [11] and recognizing emotions of encountered people [12]. There are also solutions, that incorporate special hardware measuring physiological signals

to interpret an emotional state of a person, which may be especially valuable information in case of autistic children, who often are not able to show their emotions [13], [14].

Apart from solutions designed for people with autism, there are also some applications supporting work of the therapists.

III. Study context and study design

A. Study context and purpose

In 2015 a project named AUTMON was started by a consortium consisting of Gdansk University of Technology, Harimata LTD and Hippotherapy Foundation [6]. Its goal is to develop methods and tools for automated monitoring of progress in the therapy of children with autism.

The goal is addressed by (1) the development of technology identifying behavioral patterns associated with autism, which could be used to test and evaluate the progress of therapy and by (2) implementation of a system that automatically tracks these patterns during therapy. A specially designed tablet applications have been created to record the behavioral patterns during the childs interaction with tablet. These patterns, which may be analyzed for diagnostic or therapeutic purposes, contain information on: touch screen gestures, e.g. tap and swipe device movements identified by accelerometer; the application usage, i.e. navigationpatterns, objects that attract attention (goals/distractors), strategies of decision making etc. The research goal of the AUTMON project is to verify, that the proposed solution might support the therapists by providing them with objective data on the progress of therapy of autistic children in various areas of their development.

The first step in the project research is verification of the children' experience with the games, as attitude towards the application is the key factor enabling practical application of the solutions in therapy. Reporting the study of children experience with the games is the main goal of this paper.

The key research question addressed by the presented study might be formulated as follows: Which applications provide experiences positive enough for the child to perform and continue interaction allowing (as a result) to monitor progress.

B. Applications under investigation

The applications developed during the project were developed in the cooperative and iterative process engaging psychologists, therapists and data scientists. They were designed for the tablet devices. The solution set includes five games, which allows to gather information about children interactions with the device, using touch screen sensors and gyroscope.

Each game consists of training phase and actual game. During the training phase child may get acquainted with the game concept. At the beginning of training session interactive tutorial is presented. Later child may try to use application by herself/himself. During the whole training therapists should instruct the child and may use either verbal and physical guidelines. Depending on game, training session lasts from 2 to 5 minutes. At any time training session may be terminated by tapping three fingers at the top of the screen.

After finishing training session, actual game, which is used to collect data, begins. In order to obtain the undisturbed results, during this phase, the therapist should not interact with the child. Like during the training phase the session can be terminated, when a child looses interest in the game. Fig. 1 presents the games screenshots.

1) Boxes: Boxes is a game designed for warm-up. The goal is to place the balls in boxes by matching corresponding colors.

2) Sharing: The goal of the second game is to share food among four animated children. Child has to tap on the displayed food article (watermelon, apple, cake) and swipe portions to four plates in front of the children.

3) Cat & Dog: Based on the experimental paradigm of Go/NoGo, the game is intended to be used by older children. Typically developing 3-4 years old's are not ready to proceed with this cognitively demanding task.

4) Pinwheel: The pinwheel is slowly turning around while a color ball balances at the base of its stem. A child has to flip the tablet (previously only touch screen data were gathered, now accelerometer and gyroscope provide the data) precisely to move the color ball to the corresponding "pinwheel petal".

5) Creativity: Drawing and coloring pictures seem to be the most rewarding game for most children.

Fig. 1. Sample games screenshots

C. Study design

Two evaluation methods were applied: (1) a behavioral study of video recordings of children interacting with the games and (2) on-line behavioral tagging during measurement sessions.

The first part of the study was performed before measurements sessions were started and aimed at evaluation of user experience with the games. The second part of the study is performed on-line during each measurement session and this paper summarizes the results as a validation of the chosen approach. The target group in those studies are children aged 3 up to 7 (kindergarten education level), diagnosed (or during diagnosis) with autism spectrum disorder. As children measurements require consent from parents, randomization of selection to sample was not possible. Children were recruited among pupils of ten therapeutic centers in Poland. Parents' agreement rate was very high (more than 90%).

1) User experience study: This study was aimed at analyzing the experience of the children play. The first step of this study was to find a definition and unambiguous indicators of good UX. Previous research on the subject mentioned: engagement, playfulness and fun as indicators of good UX in children [15]. Previous studies [16] used some combinations of behavioral observation with survey and interview techniques. This approach was hard to apply in this study, because children with autism are mostly nonverbal or of poor communication skills, so we couldn't ask them straightforward for their opinions. Therefore in this study behavioral observation was chosen as an investigation tool. Moreover, the gameplay was recorded and analysis were made post-hoc by independent observers. This approach allowed for measuring consistency of manual tagging in order to achieve higher reliability. Steps of the study preparation:

1) An exploratory observation of gameplay was performed resulting in a list of observational indicators of interest and having fun with the games. Three independent judges (two psychologist with specialization in autism therapy and a naive observer not familiar with the symptoms of autism) were watching videos and listing observational indicators of good/bad experience of play.
2) Behaviors listed by observers in the previous step were clustered into three categories of: engagement, understanding and enjoyment.
3) The three factors were conceptualized and operationalized with three behavioral indicators each forming 9-item behavioral observation scheme.
4) Each item was assigned a five-point discrete scale (1-5) with descriptive explanations assigned to first and last item. As a result an observation sheet was prepared, that was used in manual tagging of the video recordings.
5) Using the elaborated scale and observation sheet, manual tagging of videos was performed by 5 independent observers. Total number of 21 recordings were analyzed.
6) The manual markers from observers were checked against consistency criteria using commonly used Kappa coefficient. The data was used to evaluate the applications using the understanding, engagement and enjoyment criteria.
7) Final results were formulated regarding inclusion or exclusion of the games from the measurement procedure. Some additional recommendations were formulated for the measurement procedure.

Understanding factor characterizes, that the device functions and the games are understandable for a child. First of all, child realizes that tablet screen can be used to interact with the tablet. There is a range of algorithms of interaction (based on set of possible actions provided) that result in successful interaction with the game, meaning when child achieves the goal of the game. Child wins the game if (s)he acquires a final stage of task completion, e.g. feeds all four children, draws a picture, paints a picture etc.) Children may get to win by themselves or with a standarized verbal or motor prompt from the experimenter.

Engagement factor characterizes, that the function of the tablet and the game is engaging, meaning, that tablet and game attracts childs attention. If in childs sight, in proximity of his/her hands there is a possibility of engaging. It can be observed how often, regular and persistent is childs interaction with the touch screen.

Enjoyment factor characterizes, whether interaction with tablet results in growth of positive affect symptoms, which were operationalized as laughing and vocal expressions, verbal expressions, mimic expressions and motor and postural expressions (clapping hands, jumping, shrugging). Facial expressions were not observable due to to the fact, that video materials were filmed from the chin down (which was explicitly stated in the parent-experimenter agreement for the privacy reasons).

The results of the study are provided in Section IV.

2) Validation during measurement sessions: The second study is on-line observation made during the measurement sessions.

The children are supposed to play all five games during the sessions. The detailed criteria used in the usability study were significantly simplified for the measurement stage. From the 9-item scale only 2 indicators were used in the on-line tagging procedure. For each game the following two factors are estimated by an observer: the level of difficulty, which may be 1-easy, 2-adequate, 3-difficult; the level of interest, which may be 0-none, 1-low, 2-average, 3-high. If a child does not want to play a game in spite of some encouragement, 0 is assigned to the level of interest and no value is given as the level of difficulty. If, for some reasons, a therapist decides not to try to play a game, neither difficulty nor interest are assigned a value. The reasons for childs intentional exclusion from play might fall into following categories: a child performed an extremely negative reaction to the game (eg. fear of sounds); a child is in bad disposition on this specific session day (regular therapist opinion); a child got upset during the study procedure.

The difficulty factor corresponds to understandability criteria from the UX study, the interest factor corresponds to engagement criteria from the UX study.

IV. Usability evaluation of applications for progress monitoring

This section provides results of usability evaluation of the applications. 21 recordings of nine children have been annotated. All children were boys, aged 3-7. Each of them played from one to four of the created games.

To evaluate the consistency of the annotations done by four observers, kappa coefficient was calculated. The values obtained for all nine behavior indicators defined in section III-C were averaged over all games. The highest agreement was obtained for engagement factors (0.79; 0.70 and 0.61 respectively), moderate for understanding factors (0.57; 0.54 and 0.43) and the lowest for enjoyment (0.40; 0.30 and 0.38).

To provide final evaluation scores for the three categories (engagement, understanding, enjoyment) the following procedure has been followed independently for each game:

1) in each category the scores for three behavior indicators were summed getting total category scores;
2) the total category scores were averaged over four judges; as a result three values for engagement, understanding and enjoyment were obtained for each recording;
3) the obtained recording scores were averaged over all films presenting a given game.

The final results are presented in Table I. Sharing received the highest scores for understanding and engagement. Surprisingly, Cat & Dog took the second place in these two categories and the first one in enjoyment. That game seemed to be rather difficult, some of the children did not play it at all. In spite of these obstacles, it turned out to be entertaining. Creativity game received the lowest scores for understanding and enjoyment, Pinwheel for engagement. Although Creativity seemed to be understandable for the children, as most of them intuitively drew lines with their fingers, precise drawing turned out to be quite difficult bringing about the low scores. It can be observed that more understandable games get higher scores in engagement and enjoyment, however the relationship between the variables would require further analysis.

V. On-line tagging - execution and results

To summarize the results the first three sessions of each child have been taken into account. Due to some absence rate not all participants have already gathered three data records. From the total number of 42 children, 29 of them were present during all three sessions. The absence rate is typical for the kindergarten children age range.

The results obtained for each game are averaged over all children. The final scores for the levels of difficulty and interest are shown in Table II.

It can be seen that three of the games (Sharing, Boxes, Creativity) turned out to be much more interesting than other two (Pinwheel, Cat and Dog). Some relation between interest and difficulty may be observed. The more difficult game, the less interesting and vice versa. Boxes and Sharing were the easiest ones. Pinwheel, which requires motor skills while flipping the tablet, and Cat & Dog requiring being focused, were the most difficult, just as it was expected. Some of the kids were not able to play Cat & Dog at all.

Another interesting observation is the fact that the levels of interest and difficulty do not change much over the time. It meant that data collected while playing are not affected by confounding factor such as history effect.

VI. Summary of results and discussion

The aim of this paper was to evaluate, whether the applications developed for monitoring progress of children with autism have a chance to be used in practice. The question was raised, whether the applications are engaging, understandable and enjoying enough to trigger and maintain childs focus and interaction.

The main results might be formulated as follows:

- understanding, engagement and enjoyment are partially dependent, eg. some children are fascinated with pinwheel although they do not understand the purpose of the game, for the others, if they do not understand a game, they are not interested in it.
- both studies confirmed, that three of the games: Sharing, Creativity and Boxes are mostly understandable for the children; out of the three, Sharing seems to be the favorite game;
- game Cats & Dogs and Pinwheel seems less intuitive and half of the children refuse to use them, however, once a child understands the game, it seems to be engaging,
- although it seemed, that some of the children do remember the games after a month (average time distance between recording sessions), the data do not confirm the history effect, which is convenient for measurement.

All recording and measurement sessions revealed also some qualitative observations:

- Most of the children with autism were eager to use tablets, typical behavior was to grab a tablet in the proximity of hand and to follow it, when it was taken away,
- Most of the children know very well, how to turn the tablet on and how to switch to another game, we have used the parent mode in order to prevent children from switching the games off;
- Some lower functioning children got frustrated, when they did not understand the game;
- Some higher functioning children got bored easily and even though they started eagerly, it was hard to keep them play for longer than a minute;

Authors are aware of the fact, that this study is not free of some limitations, such as small number of videos analyzed and low consistency of manual tagging for some of the variables. Despite some limitations of the study, the performed analysis allowed us to determine, which applications provide experiences positive enough for the child to perform and continue interaction allowing (as a result) to monitor progress. The goal of the study was achieved.

TABLE I
USER EXPERIENCE RESULTS OBTAINED FOR THE APPLICATIONS

Game name	Understanding				Engagement				Enjoyment			
	Min	Max	Avg	SD	Min	Max	Avg	SD	Min	Max	Avg	SD
Boxes	10	13	nd	nd	15	15	nd	nd	8	11	nd	nd
Sharing	8	15	12,8	3,3	14	15	14,7	0,6	7	12	8,9	2,3
Pinwheel	4	15	9,5	5,2	8	15	12,9	3,1	4	12	8,8	3,3
Creativity	6	13	8,9	3	11	14	13,4	1,5	2	10	7,8	1,8
Cat & Dog	6	15	11,3	4,3	12	15	14,2	1,5	5	13	9,3	3,3

TABLE II
INTEREST AND DIFFICULTY LEVELS FOR THREE MEASUREMENT SESSIONS

Game name	Interest				Difficulty			
	Total*	1st session**	2nd session**	3rd session**	Total*	1st session**	2nd session**	3rd session**
Boxes n*=100 n**=28	2,5 (0,8)	2,4 (0,8)	2,4 (0,7)	2,5 (0,8)	1,9 (0,5)	1,8 (0,6)	1,9 (0,6)	1,9 (0,5)
Sharing n*=104 n**=28	2,6 (0,6)	2,7 (0,5)	2,7 (0,6)	2,6 (0,5)	1,8 (0,5)	1,9 (0,5)	1,9 (0,5)	1,7 (0,5)
Pinwheel n*=85 n**=22	1.9 (0,9)	2,1 (0,9)	1,8 (1,0)	2,1 (0,8)	2,6 (0,6)	2,6 (0,5)	2,6 (0,5)	2,4 (0,7)
Creativity n*=102 n**=28	2,4 (0,8)	2,6 (0,7)	2,5 (0,7)	2,4 (0,7)	2,3 (0,5)	2,2 (0,5)	2,4 (0,5)	2,2 (0,5)
Cat & Dog n*=49 n**=13	1,8 (1,0)	1,9 (0,9)	1,9 (1,0)	1,9 (1,0)	2,7 (0,6)	2,6 (0,8)	2,8 (0,4)	2,7 (0,5)

n* - total no of gameplays evaluated (all children, up to 3 sessions)
n** - no of gameplays evaluated (only children with 3 sessions)

VII. CONCLUSIONS

Our study shows the potential for use of tablet-based technology in children therapy. Children are generally enthusiastic about tablets and other mobile devices and this rapture can be utilized in the area of research, diagnosis and therapy. Future works in this discipline should focus on to developing tools for monitoring progress and better means for their evaluation. As our study concludes, its not easy to evaluate in terms of child experience, especially in autistic population, where raters agreement is hard to reach.

Objective measures of therapy progress can bring quality data to the therapy providers. As they have better knowledge of childs state and developmental pace and direction, thy can provide adjusted intervention to best fit the individual needs of the child. Better therapy means better life for children with ASD, as enormous amount of evidence shows that personalized and early onset intervention improve their future quality of life.

REFERENCES

[1] F. R. Volkmar, R. Paul, A. Klin, and D. J. Cohen, *Handbook of Autism and Pervasive Developmental Disorders, Diagnosis, Development, Neurobiology, and Behavior*. John Wiley & Sons, 2005, vol. 1.

[2] L. Winerman, "Effective education for autism," *Monitor on Psychology*, vol. 35, no. 11, pp. 46–49, 2004.

[3] A. Landowska, A. Kołakowska, A. Anzulewicz, P. Jarmołkowicz, and J. Rewera, "E-technologie w diagnozie i pomiarach postepow terapii dzieci z autyzmem w Polsce," *e-mentor*, no. 4 (56), pp. 26–30, 2014.

[4] A. Landowska, A. Kolakowska, A. Anzulewicz, P. Jarmołkowicz, and J. Rewera, "E-technologie w edukacji i terapii dzieci z autyzmem w Polsce," *EduAkcja. Magazyn edukacji elektronicznej*, vol. 8, no. 2, pp. 42–48, 2014.

[5] A. Landowska and M. Smiatacz, "Mobile activity plan applications for behavioral therapy of autistic children," in *Man–Machine Interactions 4*. Springer, 2016, pp. 115–125.

[6] "Automated therapy monitoring for children with ASD, project website," http://autmon.eti.pg.gda.pl/, accessed: 2016-05-08.

[7] W. Jones and A. Klin, "Attention to eyes is present but in decline in 2-6-month-old infants later diagnosed with autism," *Nature*, vol. 504, no. 7480, pp. 427–431, 2013. doi: 10.1038/nature12715

[8] J. Hashemi, T. V. Spina, M. Tepper, A. Esler, V. Morellas, N. Papanikolopoulos, and G. Sapiro, "Computer vision tools for the non-invasive assessment of autism-related behavioral markers," *IEEE International Conference on Development and Learning and Epigenetic Robotics*, 2012. doi: 10.1109/devlrn.2012.6400865

[9] E. B. Torres, M. Brincker, R. W. Isenhower, P. Yanovich, K. A. Stigler, J. I. Nurnberger, D. N. Metaxas, and J. V. José, "Autism: the micro-movement perspective." *Frontiers in Integrative Neuroscience*, vol. 7, 2013. doi: 10.3389/fnint.2013.00032

[10] D. Wall, J. Kosmicki, T. Deluca, E. Harstad, and V. Fusaro, "Use of machine learning to shorten observation-based screening and diagnosis of autism," *Translational psychiatry*, vol. 2, no. 4, p. e100, 2012. doi: 10.1038/tp.2012.10

[11] D. Deriso, J. Susskind, L. Krieger, and M. Bartlett, "Emotion mirror: a novel intervention for autism based on real-time expression recognition," in *Computer Vision–ECCV 2012. Workshops and Demonstrations*. Springer, 2012, pp. 671–674. doi: 10.1007/978-3-642-33885-4_79

[12] R. El Kaliouby and P. Robinson, "The emotional hearing aid: an assistive tool for children with asperger syndrome," *Universal Access in the Information Society*, vol. 4, no. 2, pp. 121–134, 2005. doi: 10.1007/s10209-005-0119-0

[13] A. Landowska, K. Karpienko, M. Wróbel, and M. Jedrzejewska-Szczerska, "Selection of physiological parameters for optoelectronic system supporting behavioral therapy of autistic children," in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, vol. 9290, 2014. doi: 10.1117/12.2075020

[14] M. Jedrzejewska-Szczerska, K. Karpienko, and A. Landowska, "System supporting behavioral therapy for children with autism," *Journal of Innovative Optical Health Sciences*, vol. 8, no. 03, p. 1541008, 2015. doi: 10.1142/s1793545815410084

[15] J. C. Read, S. MacFarlane, and C. Casey, "Endurability, engagement and expectations: Measuring children's fun," in *Interaction design and children*, vol. 2. Shaker Publishing Eindhoven, 2002, pp. 1–23.

[16] G. Sim, S. MacFarlane, and J. Read, "All work and no play: Measuring fun, usability, and learning in software for children," *Computers & Education*, vol. 46, no. 3, pp. 235–248, 2006.

The 36th IEEE Software Engineering Workshop

THE IEEE Software Engineering Workshop (SEW) is the oldest Software Engineering event in the world, dating back to 1969, and with the last 35th workshop organized in Heraclion, Crete, Greece, 12-13 October 2012.

The workshop was original run as the NASA Software Engineering Workshop and focused on software engineering issues relevant to NASA and the space industry. After the 25th edition, it became the NASA/IEEE Software Engineering Workshop and expanded its remit to address many more areas of software engineering with emphasis on practical issues, industrial experience and case studies in addition to traditional technical papers. Since its 31st edition, it has been sponsored by IEEE and has continued to broaden its areas of interest.

TOPICS

The workshop aims to bring together all those with an interest in software engineering. Traditionally, the workshop attracts industrial and government practitioners and academics pursuing the advancement of software engineering principles, techniques and practice. The workshop provides a forum for reporting on past experiences, for describing new and emerging results and approaches, and for exchanging ideas on best practice and future directions.

Topics of interest include, but are not limited to:

- Experiments and experience reports
- Software quality assurance and Metrics
- Formal methods and formal approaches to software development
- Software engineering processes and process improvement
- Agile and Lean Methods
- Requirements engineering
- Software architectures
- Real-time Software Engineering
- Software maintenance, reuse, and legacy systems
- Agent-based software systems
- Self-managing systems
- New approaches to software engineering (e.g., search based software engineering)
- Software engineering issues Cyber-physical systems
- Software Engineering for social media

EVENT CHAIRS

- **Bowen, Jonathan,** Museophile Ltd.
- **Hinchey, Mike,** Lero-the Irish Software Engineering Research Centre, Ireland
- **Ryan, Kevin,** Lero-the Irish Software Research Centre, Ireland

PROGRAM COMMITTEE

- **Ait Ameur, Yamine,** LISI/ENSMA, France
- **Banach, Richard,** University of Manchester, United Kingdom
- **Bensalem, Saddek,** VERIMAG, France
- **Bjorner, Nikolaj,** Microsoft Research, United States
- **Broy, Manfred,** Technische Universitaet Muenchen, Germany
- **Carter, John,** University of Guelph, Canada
- **Creissac Campos, José,** Universidade do Minho, Portugal
- **Denney, Ewen,** SGT/NASA Ames, United States
- **Derrick, John,** Unversity of Sheffield, United Kingdom
- **Di Vito, Ben,** NASA Langley Research Center, United States
- **Eleftherakis, George,** The University of Sheffield International Faculty, CITY College, Greece
- **Fantechi, Alessandro,** DSI - Universita' di Firenze, Italy
- **Fidge, Colin,** Queensland University of Technology, Australia
- **Forbrig, Peter,** University of Rostock
- **Fortier, Stephen,** George Washington University, United States
- **Fuhrman, Christopher,** ETS (École de technologie supérieure)
- **Fujita, Masahiro,** University of Tokyo, Japan
- **Gracanin, Denis,** Virginia Tech, United States
- **Groce, Alex,** Oregon State University, United States
- **Grosu, Radu,** Stony Brook University, United States
- **Havelund, Klaus,** Jet Propulsion Laboratory, California Institute of Technology, United States
- **Hsiao, Michael,** Virginia Tech, United States
- **Laplante, Phillip A.,** PennState, United States
- **Malloy, Brian,** Clemson University, United States
- **Nesi, Paolo,** DSI-DISIT, University of Florence, Italy
- **Palanque, Philippe,** ICS-IRIT, University Toulouse 3, France
- **Pastor Lopez, Oscar,** Universitat Politecnica de Valencia, Spain
- **Pullum, Laura,** Oak Ridge National Laboratory, United States
- **Reeves, Steve,** University of Waikato, New Zealand
- **Rouff, Christopher,** Lockheed Martin, United States
- **Rozier, Kristin Yvonne,** NASA Ames Research Center, United States
- **Seceleanu, Cristina,** Mälardalen University, Västerås, Sweden
- **Sekerinski, Emil,** McMaster University, Canada
- **Sun, Jing,** The University of Auckland, New Zealand
- **Taguchi, Kenji,** AIST, Japan

- **Velev, Miroslav,** Aries Design Automation, United States
- **Vilkomir, Sergiy,** East Carolina University, United States
- **Zalewski, Janusz,** Florida Gulf Coast University, United States

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1701–1708
DOI: 10.15439/2016F264
ACSIS, Vol. 8. ISSN 2300-5963

Alvis models of safety critical systems state-base verification with nuXmv

Jerzy Biernacki
AGH University of Science and Technology
Department of Applied Computer Science
Al. Mickiewicza 30, 30-059 Krakow, Poland
E-mail: jbiernac@agh.edu.pl

***Abstract*—For modelling of real-time safety critical systems, when traditional testing techniques cannot be applied, formal system verification is crucial. Alvis is a modelling language that combines possibilities of formal models verification with flexibility and simplicity of practical programming languages. Solutions introduced in Alvis make the development process easier and help engineers to cope with more complex systems. The paper deals with a state-based approach to the verification of Alvis models. Until the research presented in the paper were conducted, the verification process was mostly action-based. The nuXmv tool, as one of the top model checkers, was selected for the task of state-base verification of Alvis models translated into the SMV modelling language. The paper presents a translation algorithm and usability studies performed on existing safety critical systems.**

I. Introduction

Alvis [1], [2] is a formal modelling language developed at AGH-UST in Kraków, Department of Applied Computer Science (*http://alvis.kis.agh.edu.pl*). The motivation behind its creation and development is to provide a formal language which could be used by an average software engineer to model and verify complex systems. To this end, Alvis combines advantages of high level programming languages with a visual modelling language for defining communication channels between subsystems. Its most significant advantage over classical formal methods (Petri nets [3], [4], timed automata [5], [6] and process algebras [7], [8], [9]) is an engineer-friendly syntax. The heavy mathematical foundations are hidden from the user without compromising the capabilities and expressive power of the formalism. Alvis, as a formal language, has an advantage over the industry programming languages – Alvis models can be formally verified using model checking techniques [10]. Using Alvis language, an average software engineer is able to model and verify complex systems which can be then easily implemented. This is particularly important in concurrent and distributed systems where traditional methods of software testing are not applicable. The formal verifications of such systems is a ground for many current scientific projects [11]. The ongoing research on the Alvis language include also building an Alvis simulator, Alvis Virtual Machine [12] and automatic Java code generation [13].

The nuXmv tool [14], [15] is currently one of the top-notch and mainstream model checkers for temporal logics. It features a prominent and state-of-the-art verification engine. The project is still supported and developed, new versions of the tool are released regularly.

The nuXmv can check whether a given finite state model satisfies a given temporal logic formula and if not, it can provide a proper counter-example. System requirements specification, in the form of a set of LTL [16] and CTL [16], [17] temporal logics formulae, can be therefore automatically verified by the tool. In addition, it has a dedicated modelling language called SMV [15] which is relatively easy to use for modelling the system. Furthermore, according to the authors of the project it can verify systems of high complexity, i.e. containing more than 10^{20} states.

These outstanding features initially made nuXmv the best possible choice for the task of state-base Alvis model verification. The main goal of the conducted research was to verify whether nuXmv can be effectively used in the process of verification of complex systems modelled in the Alvis language. In order to prove the concept, a translation algorithm was conceived and then implemented. Extensive experiments of the solution were performed, including modelling and verification of existing real-time safety critical systems.

The paper is organised as follows. Section II contains a short introduction to the Alvis language and basic information about the process of designing and verification of Alvis models. In Section III formal definitions concerning Alvis state space representation are provided. Section IV deals with the algorithm of Alvis model translation into nuXmv. Usability studies conducted on two examples of real-time safety critical systems are presented in Section V. A short summary is given in the final section.

II. Alvis modelling language

An Alvis model is basically a collection of subsystems called *agents* that may run concurrently, communicate with each other, compete for shared resources etc. The concept of *agent* is borrowed from CCS [8], [18]. Agents are divided into *active* and *passive* ones and mimic, to some degree, tasks and protected objects in the Ada programming language [19].

Active agents may perform some activities and are treated as threads of control in a concurrent system. *Passive agents* provide a mechanism for the mutual exclusion and data synchronization.

Interconnections between agents are defined on *communication diagrams*, a visual part of the Alvis language. These

978-83-60810-90-3/$25.00

diagrams present agents as nodes and communication channels as arcs in a directed graph. To model the behaviour of the agents, the *code layer* is used. Alvis source code is similar to the one of high level programming languages. Alvis statements may also incorporate elements of the Haskell functional programming language [20].

Furthermore, the complex systems may be modelled using hierarchical communication diagrams [21]. They introduce a concept of a hierarchical node which represents a subsystem defined at the lower level. Therefore, it allows to describe a system on many different levels of abstraction. A summary of Alvis graphical elements and code statements is presented in Fig. 1.

Alvis models are designed using an Alvis design toolkit including *Alvis Editor*, *Alvis Simulator* and *Alvis Compiler* tools. *Alvis Editor* is a visual modelling environment featuring design of Alvis models. *Alvis Simulator* enables step-by-step simulations of the models. *Alvis Compiler* [22] translates designed models into Haskell program. The Haskell middle-stage representation is used to generate LTS graphs (*labelled transition system* [23]) of the Alvis models. LTS graphs will be explained in more detail in the next section. They can be used to formally verify models using model checking techniques. LTS graphs are checked in terms of satisfaction of model properties described as temporal logic formulae. The original verification process included only action-based verification with μ-calculus [24] in CADP toolbox [25]. The approach presented in this paper employs nuXmv tool to allow the usage of LTL and CTL temporal logics. The modelling and verification process of Alvis models is presented in Fig. 2.

More details on this topic may be found in the manual at the website of the Alvis project.

III. Alvis state space representation

Before an Alvis LTS to nuXmv translation algorithm can be introduced, some of the key concepts regarding Alvis state space must be defined.

Definition 1. A state of an agent X is a tuple:

$$S(X) = (am(X), pc(X), ci(X), pv(X)),$$

where $am(X)$ is an agent mode, $pc(X)$ is a program counter, $ci(X)$ is a context information list, and $pv(X)$ is parameters values.

Each agent state can be described unambiguously with information contained by this four-tuple. Where necessary, to every one of am, pc, ci and pv symbols, there can be a state index added, e.g. pv_{S_i}, to indicate which state it refers to.

The *agent mode* can take one of the five following values: *Finished* (F), *Init* (I), *Running* (R) and *Taken* (T). *Finished* means that an agent has finished its work. *Init* is the default mode for agents that are inactive in the initial state. *Running* means that an agent is performing one of its statements. *Taken* means that one of the passive agent's procedures has been called and the agent is executing it. *Waiting*, for passive agents, means that the corresponding agent is inactive and waits for another agent to call one of its accessible procedures. For active agents, this mode means that the corresponding agent is waiting either for a communication with another active agent, or for a currently inaccessible procedure of a passive one.

The *program counter* points at the current statement of an agent i.e. the next statement to be executed or the statement that has been already executed by an agent but needs a feedback from another agent to be completed (e.g. a communication between agents).

The *context information list* contains additional information about the current state of an agent e.g. if an agent is in the waiting mode, ci contains information about events the agent is waiting for.

The *parameters values list* contains the current values of the agent's parameters.

Definition 2. A state of a model $\boldsymbol{A} = (D, B, \alpha^0)$, where $D = (\mathcal{A}, \mathcal{C}, \sigma)$ and $\mathcal{A} = \{X_1, ..., X_n\}$ is a tuple $S = (S(X_1), ..., S(X_n))$.

The concept of an Alvis model state is explained in Fig. 3.

Execution of any step is expressed as a transition between formally defined states of an Alvis model. States of a model and transitions among them are represented using a labelled transition system (LTS graph).

Definition 3. A Labelled Transition System is a tuple:

$$LTS = (S, A, \rightarrow, s_0),$$

where S is a set of states, A is a set of actions, $\rightarrow \subseteq S \times A \times S$ is the transition relation and s_0 is an initial state.

For an Alvis model, an LTS is a four-tuple:

$$LTS = (\mathcal{R}(S_0), \mathcal{T}, \rightarrow, S_0),$$

where $\mathcal{R}(S_0)$ is a set of states reachable from the initial state, $\mathcal{T}$ is a set of all possible steps for a given model, $\rightarrow = \{(S, t, S') : S - t \rightarrow S' \wedge S, S' \in \mathcal{R}(S_0)\}$, where $t \in \mathcal{T}$, and S_0 is an initial state. In untimed Alvis models arcs are labelled with names of individual steps performed by agents. In the timed models arcs are labelled with the sets of parallel steps.

In order to describe the translation algorithm in the next section, a few additional terms need to be introduced:

- $B(X)$ – Agent X dynamics definition (code);
- $card(B(X))$ – number of steps in $B(X)$;
- $\mathcal{N}(t)$ – a name of the t transition.

NuXmv models are basically *finite state transition systems* [15] which can be defined as *Kripke structures* [26].

Definition 4. A *finite state transition system* is a tuple $TS = (S, I, \rightarrow, L)$, where:

- S is a finite set of *states*,
- $I \subseteq S$ is the set of *initial states*,
- $\rightarrow \subseteq S \times S$ is the *transition relation*, specifying the possible transitions from state to state,
- L is the *labelling function* that labels states with *atomic propositions* that hold for the given state.

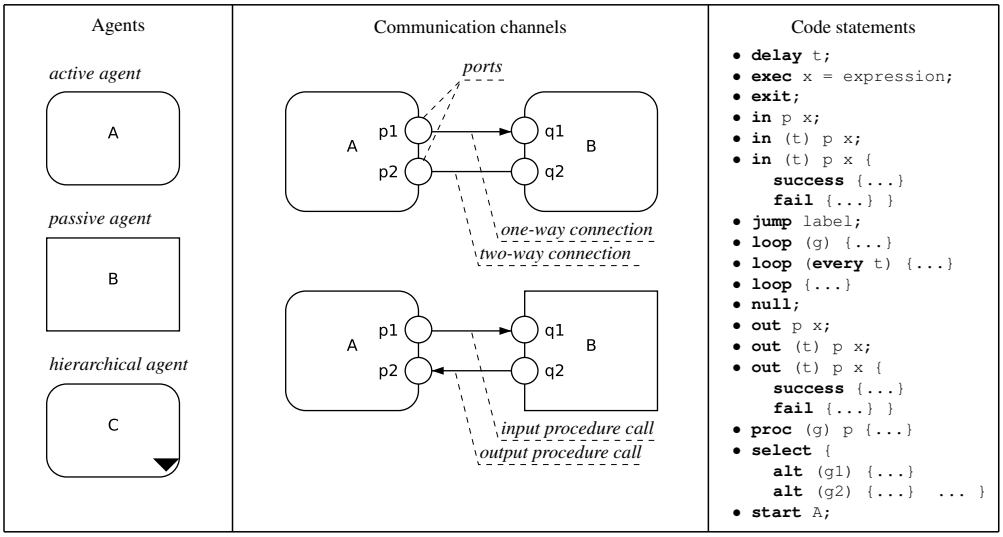


Figure 1: Elements of Alvis language.

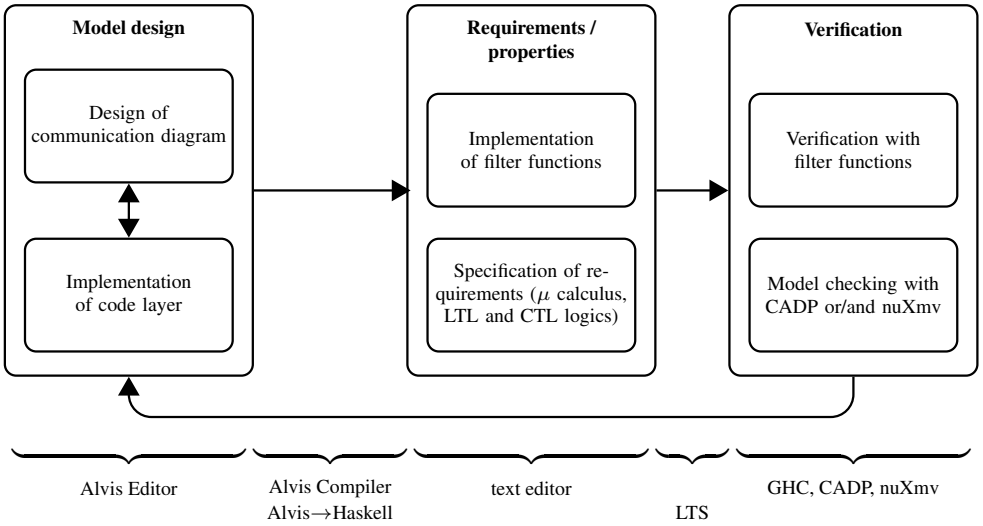


Figure 2: Alvis modelling and verification process.

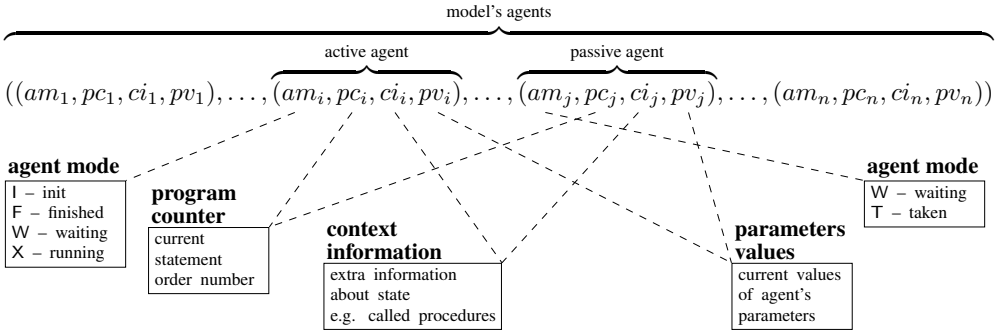


Figure 3: Representation of an Alvis model state.

IV. Alvis LTS to nuXmv translation algorithm

Finite state transition systems in nuXmv tool are modelled with a dedicated modelling language called SMV [15]. In the presented approach, a nuXmv model after translation consists of three main parts: Variables definitions (`VAR` and `IVAR`), `ASSIGN` section and specification of transitions' availability

(TRANS). The first one of them, the IVAR section, contains definition of an input variable action. It is used to contain transitions' labels.

The VAR section is used to contain definitions of standard variables. These include set of states and atomic propositions variables. The ASSIGN section is composed of three main parts. The first one initializes the state variable, the second is responsible for defining transitions between the states and the final one assigns values to the atomic propositions for specific states. The set of atomic propositions is given implicitly using variables and their domains. The last main part, the TRANS section, specifies for every state which actions are available in it. For instance, line `TRANS s = s1 -> (action = a1) | (action = a2)` determines that when the system is in state s1, the only available actions are a1 and a2.

A translation algorithm of an Alvis model LTS into the nuXmv code is presented in Fig. 1. In the adopted notation, ▷ symbol indicates generated nuXmv code, # represents string concatenation, $Label()$ produces a variable name and $Type()$ returns a variable's data type.

The first part of the algorithm is responsible for generation of a declaration of the input variable action. In the proposed approach, this variable is used to define names of the transitions between the model states as edges' labels in the nuXmv transition system. It enables using transition names in LTL formulae during the verification process. During the model simulation this additional information is a major enhancement that allows to analyse not only the state changes, but also the steps that led to them. The $action$ set is initialised with a single element representing empty action named NOP. Every transition label is added to the set then.

In the next step, the domain of the state variable s is defined. Its value denotes the current state of an Alvis model. For every reachable state ($S \in \mathcal{R}(S_0)$), the state name is added to the set of nuXmv states.

Lines 13–41 contain steps required to define variables representing elements of the model's state. This section starts with L_{am}, L_{pc}, L_{ci} and L_{pv} sets initialisation. They represent the sets of defined variable labels for agent mode, program counter, context information list, and parameters values correspondingly. Every agent has exactly one am and pc variable, while possibly multiple ci and pv variables.

For every agent of the model, variable labels are generated by concatenating agent's name with a two letter abbreviation indicating the element of the agent state's tuple it refers to. Agent mode variable is an enumeration and can have assigned one of the x, w, f, i and t values. Program counter variable is a bounded integer, ranging from 0 to $card(B(X_i))$, where X_i is the given agent. Context information variable labels, in addition to the basic naming convention, contain also information about the possible values of the original ci. They are booleans and the TRUE value implies that the given ci entry is present in the context list in the given state. Parameters values labels are generated by concatenating agent name, pv keyword, pv order number. If pv variable's type in the code layer is integer or boolean, the nuXmv variable is of the same type. In other case pv value is appended to the label and the type of the nuXmv variable is boolean. TRUE means that the given agents parameter is of the value specified as the last part of nuXmv variable label in the given state.

The next part of the algorithm starts with adding of the beginning of the ASSIGN section and initialization of the s variable with the name of the initial state. Then the transition relation switch statement is opened (line 44). $successors_{ik}$ is the set of successors of the si state, reachable through transition t_k. In the nested loops that follow, successors lists for every reachable state are generated.

The next section of the algorithm contains five similar blocks of pseudocode (lines 57–63, 64–70, 71–80, 81–95, 96–104). Each generates labelling functions for the agent state variables defined in the VAR section. Labelling functions are basically switch statements in which the proper value is assigned to the variable depending on the current state of the system. pv variables labelling functions are divided into two separate loops, depending on the type of the variable in the code layer.

The last part of the algorithm (lines 105–114) defines availability of the transitions. It is determined by the value of the si variable. TRANS line is generated for every reachable state. It contains a list of available transitions. NOP action is not listed there because it is only available when no successor exists. This situation indicates a terminal state of the system.

Algorithm 1 Alvis LTS to nuXmv translation algorithm.

1: ▷ `MODULE main`
2: ▷ `IVAR`
3: $action \leftarrow \{NOP\}$
4: **for all** $t_i \in T$ **do**
5: $\quad action \leftarrow action \cup \{Label(\mathcal{N}(t_i))\}$
6: **end for**
7: ▷ `action : {NOP,N(t0), N(t1), ...};`
8: ▷ `VAR`
9: **for all** $S_i \in \mathcal{R}(S_0)$ **do**
10: $\quad s \leftarrow s \cup \{Label(s_i)\}$
11: **end for**
12: ▷ `s : {s0, s1, ...};`
13: $L_{am} \leftarrow \emptyset$
14: $L_{pc} \leftarrow \emptyset$
15: $L_{ci} \leftarrow \emptyset$
16: $L_{pv} \leftarrow \emptyset$
17: **for all** $X_i \in \mathcal{A}$ **do**
18: $\quad l_{am} \leftarrow Label(X_i\#\texttt{am})$
19: $\quad$ ▷ `X_i#am : {x, w, f, i, t};`
20: $\quad L_{am} \leftarrow L_{am} \cup \{l_{am}\}$
21: $\quad l_{pc} \leftarrow Label(X_i\#\texttt{pc})$
22: $\quad k \leftarrow card(B(X_i))$
23: $\quad$ ▷ `X_i#pc : 0..k;`
24: $\quad L_{pc} \leftarrow L_{pc} \cup \{l_{pc}\}$
25: $\quad$ **for all** $S_j \in \mathcal{R}(S_0)$ **do**
26: $\quad\quad l_{ci} \leftarrow Label(X_i\#\texttt{ci}\#ci_{S_j}(X_i))$
27: $\quad\quad$ ▷ `l_ci : boolean;`

$L_{ci} \leftarrow L_{ci} \cup \{l_{ci}\}$
end for
for all $pv_j \in pv(X_i)$ **do**
$type \leftarrow Type(pv_j)$
if $type = Integer \vee type = Boolean$ **then**
$l_{pv} \leftarrow Label(X_i\#\texttt{pv}\#j)$
▷ `l_pv : type;`
else
$l_{pv} \leftarrow Label(X_i\#\texttt{pv}\#j\#pv(X_i))$
▷ `l_pv : boolean;`
end if
$L_{pv} \leftarrow L_{pv} \cup \{l_{pv}\}$
end for
end for
▷ `ASSIGN`
▷ `init(s) = s0;`
▷ `next(s) := case`
for all $si \in s$ **do**
for all $t_k \in \mathcal{T}$ **do**
$successors_{ik} \leftarrow \emptyset$
for all $sj \in s$ **do**
if $S_i \xrightarrow{t_k} S_j$ **then**
$successors_{ik} \leftarrow successors_{ik} \cup \{sj\}$
end if
end for
▷ `s = si & action = t_k:` $\{successors_{ik}\}$`;`
end for
end for
▷ `esac;`
for all $l_{am} \in L_{am}; (l_{am} = Label(X_i\#\texttt{am}))$ **do**
▷ `l_am := case`
for all $sj \in s$ **do**
▷ `s = sj : am_sj(X_i);`
end for
▷ `esac;`
end for
for all $l_{pc} \in L_{pc}; (l_{pc} = Label(X_i\#\texttt{pc}))$ **do**
▷ `l_pc := case`
for all $sj \in s$ **do**
▷ `s = sj : pc_sj(X_i);`
end for
▷ `esac;`
end for
for all $l_{ci} \in L_{ci}; (l_{ci} = Label(X_i\#\texttt{ci}\#ci_{S_j}(X_i))$ **do**
▷ `l_ci := case`
for all $sk \in s$ **do**
if $ci_{s_k}(X_i) = ci_{S_j}(X_i))$ **then**
▷ `s = sk : TRUE;`
end if
end for
▷ `TRUE: FALSE;`
▷ `esac;`
end for
for all $l_{pv} \in L_{pv} : Type(l_{pv}) = Integer \vee Type(l_{pv}) = Boolean; (l_{pv} = Label(X_i\#\texttt{pv}\#j))$ **do**
▷ `l_pv := case`
for all $sk \in s$ **do**
$value \leftarrow pv_{s_k}(X_i)$
if $value \neq 0 \wedge value \neq FALSE$ **then**
▷ `s = sk : value;`
end if
end for
if $Type(l_{pv}) = Integer$ **then**
▷ `TRUE: 0;`
else
▷ `TRUE: FALSE;`
end if
▷ `esac;`
end for
for all $l_{pv} \in L_{pv} : Type(l_{pv}) \neq Integer \wedge Type(l_{pv}) \neq Boolean; (l_{pv} = Label(X_i\#\texttt{pv}\#j\#pv(X_i)))$ **do**
▷ `l_pv := case`
for all $sk \in s$ **do**
$value \leftarrow pv_{s_k}(X_i)$
▷ `s = sk : value;`
end for
▷ `TRUE: FALSE;`
▷ `esac;`
end for
for all $si \in s$ **do**
$T_i \leftarrow \emptyset$
for all $t_k \in \mathcal{T}$ **do**
if $\exists_{S_j} S_i \xrightarrow{t_k} S_j$ **then**
$T_i \leftarrow T_i \cup \{t_k\}$
end if
▷ `TRANS s = si ->`
`(action = Ti_0|action = Ti_1 | ... );`
end for
end for

The main purpose of including the above algorithm in this paper is to convey the basic concept behind the translation. Therefore, as one may notice, the above algorithm is not optimal. Nonetheless, after many optimizations and enhancements, this algorithm was implemented as an additional module to the PetriNet2ModelChecker tool. This module enables automatic conversion of an LTS graph of an Alvis model stored in .dot file into nuXmv code, and therefore allows to verify any Alvis model using LTL and CTL logics in one of the top model checkers available.

V. Usability studies

The presented approach was tested against models of existing safety-critical systems. Among them, the tests were conducted on railway switch system and fire alarm contol panel. The former is the solution manufactured by Grupa ZUE S.A. [27] and employed in public transport in Krakow, Szczecin and Wroclaw [28]. The latter is a project of the SITP organization [29]. More information on this system can be found in [30]. The approach presented in this paper will be illustrated on the first one of them. A schematic of the system

is shown in Fig. 4.

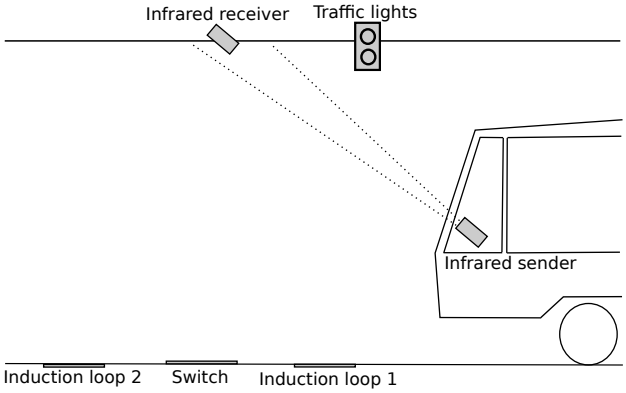


Figure 4: Railway switch system.

In every tram, there is an NP03 infra-red transmitter used to send IR signals to the switch control system. It is located on a tram driver console panel. An OP03 receiver is responsible for collecting infra-red signals and sending them to system driver. It is usually installed on overhead lines or on special poles placed before the switch. Traffic lights are providing a motorman with two pieces of information, i.e. the current direction of the tracks and a status of a switch blades lock.

A motorman's responsibility is to ensure that switch blades are set in the right direction and locked before he can drive a tram through a switch. He can change blade's direction using an NP03 transmitter. On receiving a signal, an IR receiver sends a pulse to a switch motor controller, which in turn sends a signal to blades controller to change direction of the rails. The change is possible only in the operating range of the receiver. Tram driver has limited time for choosing the expected direction, depending on the speed of the tram. If the motorman does not change the direction while the tram is in a reach zone of the IR receiver, he would have to stop the tram and manually change the direction using a special lever.

In a switch zone, there are also two induction loops installed, one before and one after switch blades. They are responsible for detecting a tram entering and leaving the crossing zone. When a tram is detected, an electrical switch lock mechanism is locking blades in the current position, in order for tram to pass safely through the switch zone. They are unlocked immediately after the tram leaves the crossing.

An Alvis model of this system was constructed. Its communication diagram is presented in Fig. 5.

Using the implementation of the algorithm presented in Section IV the system model was automatically translated into a nuXmv source file. The nuXmv representation maintains every piece of the original information about system behaviour stored in the LTS. The next step involves verification of system properties. In the presented approach they are described as LTL and CTL formulae. Three examples of such formulae are given in Listing 1.

Listing 1: Examples of LTL and CTL formulae for the railway switch system model.

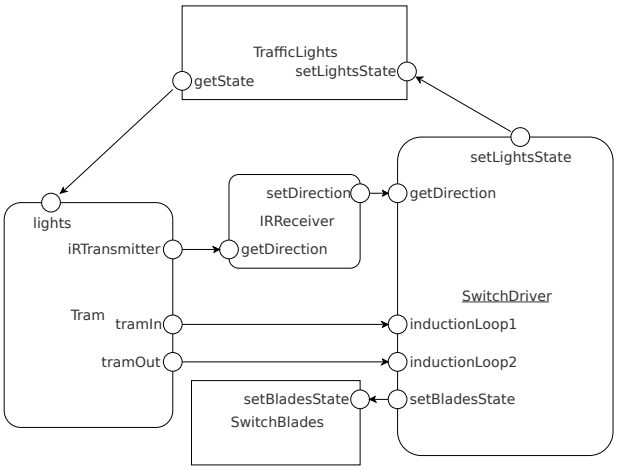


Figure 5: Railway switch system communication diagram.

```
--1) TramPassed = false U BladesLocked = true
LTLSPEC (SwitchDriver#pv3 < 2) U (SwitchBlades#pv1 <
  ↪  0)

--2) EF TramPassed = true
CTLSPEC EF (SwitchDriver#pv3 = 2)

--3) F TramPassed = true
LTLSPEC F (SwitchDriver#pv3 = 2)
```

`SwitchDriver#pv3` denotes a `tram` variable from the *SwitchDriver* agent. It can have three values: 0 when the tram is before the first induction loop, 1 when tram passed the first induction loop and 2 when the tram passed the crossing (the second induction loop). `SwitchBlades#pv1` is a `bladesState` variable. When its value is positive, the switch blades are not locked and when the value is negative, the blades are locked.

The first formula verifies whether the switch is being locked before the tram passes the crossing. This formula is crucial for the safety of railway switch mechanism. The nuXmv confirmed that it is satisfied in the model. The second formula checks whether there is a path in which a tram passes through the crossing. It is also true. The last one is similar to the previous, except it checks whether a tram always finally passes through the crossing. This one is not satisfied. The true potential of nuXmv is in providing the counterexample. If a formula is not true, nuXmv provides a path that proves it. In this case the tram does not pass through the crossing if the driver did not manage to send the signal to change the blades direction when the tram was in the zone of the infra-red receiver. Driver has to stop the tram, step out of it and manually change the direction. It is a desired behaviour of the system.

Three versions of the same model were prepared with varying complexity, each describing the system on a different level of abstraction. For each one of them, the same set of properties was verified. They were categorized into three groups: reachability, safeness and liveness properties. For each

group an average verification time was measured and presented in Table I. Tests were performed on a PC with AMD Phenom II X6 processor and 16 GB of RAM. The verification times are growing fast with complexity but even for the most complex railway switch system version, they are quite acceptable.

For the comparison, translation to nuXmv was performed also on the fire alarm control panel system. Its communication diagram is presented in Fig. 6. As this system is significantly more complex, the amount of states in the LTS is adequately larger. Although the translation itself was fast, nuXmv couldn't handle loading of such a complex system on the testing machine. The amount of memory needed to load the model exceeded available resources (RAM and swap space) and the nuXmv process was killed by the operating system. Therefore, the average verification times are not provided.

The results of the tests confirm that the presented approach is performing well for models of medium complexity. More complex ones require a lot of resources, especially RAM. Provided enough RAM or swap space, the approach can be applied to most models.

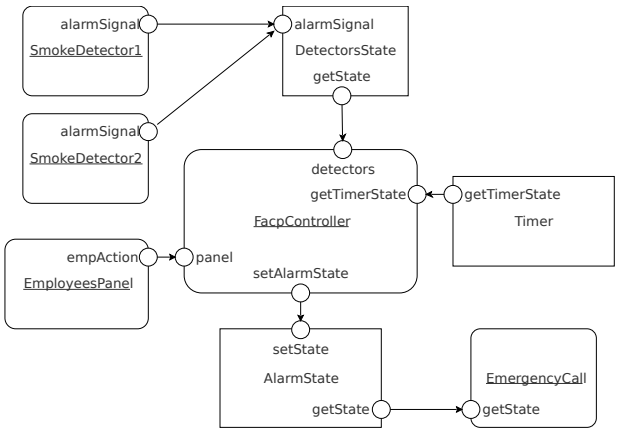


Figure 6: Fire alarm control panel communication diagram.

VI. Summary

The paper introduces state-base approach to verification of Alvis models. The presented solution enables to automatically verify properties of the modelled system using the mainstream model checking tool nuXmv. It employs formulated and implemented algorithm of Alvis LTS translation into nuXmv source code.

The usability studies of the concept were conducted on models of actual real-time safety critical systems, railway switch system and fire alarm control panel. Illustrative properties of these systems have been specified as LTL and CTL formulae and verified with nuXmv to demonstrate the capabilities and limitations of the approach. The results of the tests have been presented and summarised.

The proposed verification method proved to be handling most middle-sized models with ease, even on a regular PC. Although nuXmv supposed capabilities exceed the current needs, the solution is limited by the amount of RAM available. Complex systems require a great deal of memory to load. Compared to the amount of memory required to verify systems using action-based approach in CADP, nuXmv seems ineffective. On the other hand, state-based solution allows much more thorough verification because of the information stored about the states. The combination of both state- and action-based verification is currently the most optimal option.

Future work on state-based verification of Alvis models will focus on other possibilities. The most promising is a concept of a dedicated query language operating on the middle-stage Haskell representation.

References

[1] M. Szpyrka, P. Matyasik, and R. Mrówka, "Alvis – modelling language for concurrent systems," in *Intelligent Decision Systems in Large-Scale Distributed Environments*, ser. Studies in Computational Intelligence, P. Bouvry, H. Gonzalez-Velez, and J. Kołodziej, Eds. Springer-Verlag, 2011, vol. 362, ch. 15, pp. 315–341.

[2] M. Szpyrka, P. Matyasik, R. Mrówka, and L. Kotulski, "Formal description of Alvis language with α^0 system layer," *Fundamenta Informaticae*, vol. 129, no. 1-2, pp. 161–176, 2014. doi: 10.3233/FI-2014-967

[3] K. Jensen and L. Kristensen, *Coloured Petri nets. Modelling and Validation of Concurrent Systems*. Heidelberg: Springer, 2009.

[4] T. Murata, "Petri nets: Properties, analysis and applications," *Proceedings of the IEEE*, vol. 77, no. 4, pp. 541–580, 1989.

[5] J. Bengtsson and W. Yi, "Timed automata: Semantics, algorithms and tools," *Lecture Notes on Concurrency and Petri Nets*, vol. 3098, 2004.

[6] R. Alur and D. Dill, "A theory of timed automata," *Theoretical Computer Science*, vol. 126, no. 2, pp. 183–235, 1994.

[7] J. A. Bergstra, A. Ponse, and S. A. Smolka, Eds., *Handbook of Process Algebra*. Upper Saddle River, NJ, USA: Elsevier Science, 2001.

[8] R. Milner, *Communication and Concurrency*. Prentice-Hall, 1989.

[9] T. Bolognesi and E. Brinksma, "Introduction to the iso specification language lotos," *Comput. Netw. ISDN Syst.*, vol. 14, no. 1, pp. 25–59, Mar. 1987. doi: 10.1016/0169-7552(87)90085-7. [Online]. Available: http://dx.doi.org/10.1016/0169-7552(87)90085-7

[10] C. Baier and J.-P. Katoen, *Principles of Model Checking*. London, UK: The MIT Press, 2008.

[11] I. Grobelna, R. Wisniewski, M. Grobelny, and M. Wisniewska, "Design and verification of real-life processes with application of petri nets," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. PP, no. 99, pp. 1–14, 2016. doi: 10.1109/TSMC.2016.2531673

[12] P. Matyasik, "Alvis virtual machine," in *Computer Science and Information Systems (FedCSIS), 2014 Federated Conference on*, Sept 2014. doi: 10.15439/2014F267 pp. 1639–1645.

[13] P. Matyasik, M. Szpyrka, and M. Wypych, "Generation of Java code from Alvis model," in *International Conference of Computational Methods in Sciences and Engineering (ICCMSE 2015)*, ser. AIP Conference Proceedings, vol. 1702. Athens, Greece: AIP Publishing, March 20-23 2015. doi: 10.1063/1.4938890 pp. 100 013–1–100 013–4.

[14] R. Cavada, A. Cimatti, M. Dorigatti, A. Griggio, A. Mariotti, A. Micheli, S. Mover, M. Roveri, and S. Tonetta, "The nuXmv symbolic model checker," in *Computer Aided Verification*, ser. Lecture Notes in Computer Science, vol. 8559. Springer, 2014, pp. 334–342.

[15] A. Cimatti, E. Clarke, F. Giunchiglia, and M. Roveri, "NUSMV: a new symbolic model checker," *International Journal on Software Tools for Technology Transfer*, vol. 2, no. 4, pp. 410–425, 2000.

[16] E. Clarke, O. Grumberg, and D. Peled, *Model Checking*. Cambridge, Massachusetts: The MIT Press, 1999.

[17] E. Emerson, "Temporal and modal logic," in *Handbook of Theoretical Computer Science*, J. van Leeuwen, Ed. Elsevier Science, 1990, vol. B, pp. 995–1072.

[18] M. Szpyrka and P. Matyasik, "Formal modelling and verification of concurrent systems with XCCS," in *Proceedings of the 7th International Symposium on Parallel and Distributed Computing (ISPDC 2008)*, Krakow, Poland, July 1-5 2008, pp. 454–458.

[19] A. Burns and A. Wellings, *Concurrent and real-time programming in Ada 2005*. Cambridge University Press, 2007.

[20] B. O'Sullivan, J. Goerzen, and D. Stewart, *Real World Haskell*. Sebastopol, CA, USA: O'Reilly Media, 2008.

Table I: Performance tests for Alvis models verification in nuXmv.

Model	Number of states	RAM usage [GB]	Avarage verification time [s]		
			Reachability properties	**Safeness properties**	**Liveness properties**
Basic railway switch system	1221	0.17	0.91	1.71	1.32
Detailed railway switch system	3005	0.83	4.01	4.08	4.81
Complex railway switch system	6833	3.6	25.13	46.31	41.89
Fire alarm control panel	43624	>20	-	-	-

[21] M. Szpyrka, P. Matyasik, J. Biernacki, A. Biernacka, M. Wypych, and L. Kotulski, "Hierarchical communication diagrams," *Computing and Informatics*, vol. 35, pp. 55–83, 2016.

[22] M. Wypych, M. Szpyrka, and P. Matyasik, "Extension of Alvis compiler front-end," in *International Conference of Computational Methods in Sciences and Engineering (ICCMSE 2015)*, ser. AIP Conference Proceedings, vol. 1702. Athens, Greece: AIP Publishing, March 20-23 2015. doi: 10.1063/1.4938892 pp. 100 015–1–100 015–4.

[23] M. Szpyrka, P. Matyasik, and M. Wypych, "Generation of labelled transition systems for alvis models using haskell model representation," in *Proceedings of the 22nd International Workshop on Concurrency, Specification and Programming (CS&P 2013)*, vol. 1032. Warsaw, Poland: CEUR Workshop Proceedings, 2013, pp. 409–420.

[24] E. A. Emerson, "Model checking and the Mu-calculus," in *Descriptive Complexity and Finite Models*, ser. DIMACS Series in Discrete Mathematics and Theoretical Computer Science, N. Immerman and P. G. Kolaitis, Eds. American Mathematical Society, 1997, vol. 31, pp. 185–214.

[25] H. Garavel, F. Lang, R. Mateescu, and W. Serwe, "CADP 2006: A toolbox for the construction and analysis of distributed processes," in *Computer Aided Verification (CAV'2007)*, ser. LNCS, vol. 4590. Berlin, Germany: Springer, 2007, pp. 158–163.

[26] S. Kripke, "A semantical analysis of modal logic I: normal modal propositional calculi," *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, vol. 9, pp. 67–96, 1963, announced in *Journal of Symbolic Logic*, **24**, 1959, p. 323.

[27] ZUE S.A. (2015) Infrared-systems. Http://www.grupazue.pl/en/oferta/produkcja-urzadzen/infrared-systems.

[28] M. Gorowski. (2014) Rail transport. Http://www.transportszynowy.pl/zwrotnicetramsterowanie.php.

[29] J. Ciszewski, K. Kunecki, W. Markowski, J. Sawicki, and M. Sobecki, *SITP Guideline WP-02:2010. Fire alarm systems. The design*, 2010.

[30] J. Biernacki, A. Biernacka, and M. Szpyrka, "Action-based verification of RTCP-nets with CADP," in *International Conference of Computational Methods in Sciences and Engineering (ICCMSE 2015)*, ser. AIP Conference Proceedings, vol. 1702. Athens, Greece: AIP Publishing, March 20-23 2015. doi: 10.1063/1.4938888 pp. 100 011–1–100 011–4.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1709–1713
DOI: 10.15439/2016F289
ACSIS, Vol. 8. ISSN 2300-5963

Java-HCT: An approach to increase MC/DC using Hybrid Concolic Testing for Java programs

Sangharatna Godboley[1], Arpita Dutta[2], Durga Prasad Mohapatra[3]
National Institute of Technology Rourkela
Odisha, India
Email: sanghu1790@gmail.com[1], arpitad10j@gmail.com[2], durga@nitrkl.ac.in[3]

***Abstract*—Modified Condition / Decision Coverage (MC/DC) is the second strongest coverage criterion in white-box testing. According to DO178C/RTCA criterion it is mandatory to achieve Level A certification for MC/DC. Concolic testing is the combination of Concrete and Symbolic execution. It is a systematic technique that performs symbolic execution but uses randomly-generated test inputs to initialize the search and to allow the tool to execute programs when symbolic execution fails. In this paper, we extend concolic testing by computing MC/DC using the automatically generated test cases. On the other hand Feedback-Directed Random Test Generation builds inputs incrementally by randomly selecting a method call to apply and find arguments from among previously-constructed inputs. As soon as the input is built, it is executed and checked against a set of contracts and filters.**

In our proposed work, we combine feedback-directed test cases generation with concolic testing to form Java-Hybrid Concolic Testing (Java-HCT). Java-HCT generates more number of test cases since it combines the features of both Feedback-Directed Random Test and Concolic Testing. Hence, through Java-HCT, we achieve high MC/DC. Combinations of approaches represent different tradeoffs of completeness and scalability. We develop Java-HCT using RANDOOP, jCUTE, and COPECA. Combination of RANDOOP and jCUTE creates more test cases. COPECA is used to measure MC/DC% using the generated test cases. Experimental study shows that Java-HCT produces better MC/DC% than individual testing techniques(feedback-directed random testing and concolic testing). We have improved MC/DC by $\times$1.62 and by $\times$1.26 for feedback-directed random testing and concolic testing respectively.

I. Introduction

SOFTWARE Testing is the technique to detect bugs in software. Manual software testing accounts for 50-80% of the cost of software development. Manually created test cases are expensive, error-prone, and generally not exhaustive [2]. Therefore, automated software testing techniques have been discovered [3], [4].

There exists some controversy regarding the relative advantages of random testing and systematic testing. Some work [5], [6] suggest that random testing is same effective as systematic testing techniques. Existing work [7] found that random test case generation achieves less code coverage than systematic generation techniques. These systematic generation techniques include chaining, exhaustive generation, model checking, and symbolic execution.

Pacheco et al. [8] proposed Feedback-Directed Random testing. They have addressed random generation of unit tests for object-oriented programs. Their proposed work indicates that feedback-directed random generation retains the benefits of random testing (scalability, simplicity of implementation), and avoids redundant test cases.

Concolic testing is a systematic technique that performs symbolic execution. Concolic testing uses randomly generated test cases to start the search and to allow the tool to make progress when symbolic execution fails due to limitation of the symbolic technique (e.g. native calls) [9].

MC/DC is a criterion for code coverage and was introduced by the RTCA DO-178B standard [10]. MC/DC must satisfy the following criteria [11]:

- All the entry and exit points of the input programs must be invoked at least once.
- All possible outcomes of a decision must be affected by the changes made to each condition.
- All possible outcomes of every decision must execute.
- All the conditions in a decision must execute.

According to Majumdar et al. [2], in hybrid concolic testing, the concolic testing phase is initiated whenever random testing saturates, i.e, does not find new coverage points even after running a predetermined number of steps. Majumdar et al. [2] observed that CUTE and jCUTE tools have ultimately run up against path explosion. Concolic testing can only cover a small fraction of branches, those that can be reached using “short" executions from the initial state of the program. Therefore concolic testing requires “deep" program status to be explored.

We have implemented Java-Hybrid Concolic Testing using RANDOOP, jCUTE, and COPECA, and applied it to achieve high Modified Condition/Decision Coverage for Java programs.

The rest of the article is organized as follows: Section 2 presents the background concepts. Section 3 presents the proposed approach Java-HCT. Section 4 shows the experimental study. Section 5 compares the proposed approach with some of the existing approaches. Section 6 concludes the paper and suggests some future work.

II. Background Concepts

In this section we discuss some important background concepts, which are required to understand our work.

Definition 1: Feedback-Directed Random Testing: “It is a combination of random and systematic approach that results a test suite consisting of unit tests for the classes under test. Systematic approach deals with Feedback-Directed, i.e

as soon as an input value is built, it is executed and checked against a set of contacts and filters. The result of the execution determines whether the input is redundant, illegal or useful for generation of more input [8]."

There is a tool available for Feedback-Directed Random Testing called RANDOOP[1]. RANDOOP stands for Random Tester for object-Oriented Programs. It is a fully automatic tool and requires no input from the user, and scales to realize applications with hundreds of classes.

Definition 2: Concolic Testing: "Concolic testing is defined as a variant of symbolic execution where symbolic execution is run simultaneously with concrete executions, i.e., the program is simultaneously executed on concrete and symbolic values, and symbolic constraints generated along the path are simplified using the corresponding concrete values. The symbolic constraints are then used to incrementally generate test inputs for better coverage by combining symbolic constraints for a prefix of the path with the negation of a conditional taken by the execution [9], [15]."

JCUTE[2] is a Java concolic unit test engine based on concolic testing to execute Java programs.

Definition 3: Java-HCT: "Java-Hybrid Concolic Testing is the combination of Feedback-Directed Random Testing and Concolic testing for Java programs to result high MC/DC coverage."

Java-HCT is implemented using RANDOOP, jCUTE, and COPECA. RANDOOP and jCUTE are open source testing tools and used performing Random testing and Concolic testing respectively. We have developed the tool COPECA (COverage PErcentage CAlculator), which is plugged into RANDOOP and jCUTE to measure MC/DC%, using the generated test cases. COPECA is based on Extended Truth Table.

Definition 4: Modified Decision / Condition Coverage: "MC/DC is some kind of Predicate Coverage technique, where condition is a leaf level Boolean expression and decision controls the program flow. MC/DC% is defined as the total number of independently affected conditions (I) out of total conditions (C) present in a program [11] mathematically."

$$MCDC\% = \frac{|I|}{|C|} * 100\% \quad (1)$$

III. Proposed Approach: Java-HCT

In this section, we discuss the detailed and algorithmic description of Java-HCT followed by the proposed steps of the technique.

A. Overview

Our proposed technique Java-HCT consists of seven modules. These are i) Syntax_Converter, ii) RANDOOP, iii) jCUTE, iv) TCs Extractor, v) TCs Combiner, vi) TCs Minimizer, and vii) COPECA. These modules are shown in Fig. 1. Java-HCT accepts a Java program and produces MC/DC%.

[1]https://github.com/randoop/randoop-eclipse-plugin

[2]http://osl.cs.illinois.edu/software/jcute/

TABLE I
CHARACTERISTICS OF DIFFERENT TARGET PROGRAMS

Sl. No.	Program Name	LOC	# of Predicates	# of Conditions	# of Variables
1	SwitchTest	84	1	2	2
2	StringBuffer	1369	5	10	3
3	ScopeCheck	148	8	18	8
4	MyQuickSort	87	1	2	3
5	MathCall1	190	13	26	4
6	MyInsertionSort	70	2	6	4
7	Condition	60	4	9	3
8	FruitSales	267	23	69	4
9	InsertionSort	163	7	14	6
10	Comparison1	128	17	43	4
11	DSort1	136	10	20	2
12	GradeCalculation	103	6	12	1
13	MarketSales1	179	8	17	4
14	FruitBasket1	209	12	38	2
15	BSTree	307	6	13	3
16	SwitchTest2	104	6	16	5
17	AssertTest	75	3	7	3
18	BubbleSort	142	6	14	7
19	DSort_BST	305	3	7	3
20	CAssume	63	3	7	3
21	Demo1	76	3	8	2
22	MarketSales2	230	24	49	7
23	MathCall2	160	7	14	4
24	Selection_Sort	163	7	14	6
25	Sorting_algo	336	25	50	9
26	SwitchTest3	80	2	2	1
27	StringBuffer1	485	5	15	4
28	StudentGrades	67	5	10	1
29	Testy	53	3	6	1
30	Weight	39	1	3	3
31	Weight_Exp1	114	10	22	3
32	Weight_Exp2	77	5	13	3
33	Wildlife1	17	9	28	3
34	Wildlife2	199	13	40	3
35	Zodiac	104	18	84	10
36	WBS	321	5	10	3
37	AssertTest2	91	7	21	7
38	HelloWorld	44	2	4	2
39	IFExample	82	2	4	2
40	IFSample	95	6	12	3

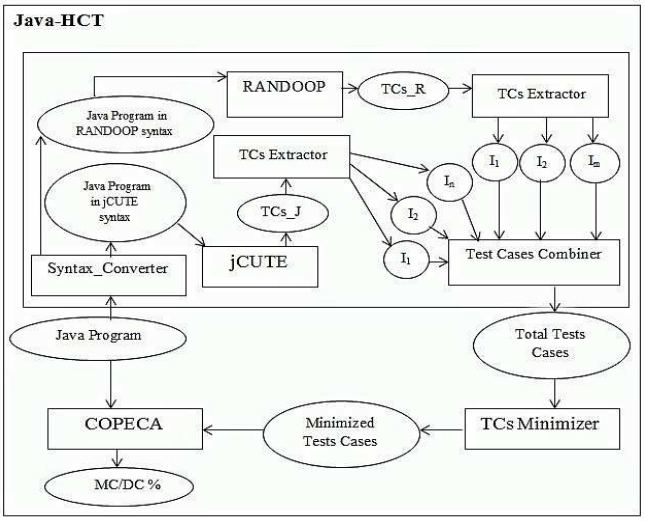


Fig. 1. Schematic representation of Java-HCT

Basically Java-HCT is the combination of RANDOOP and jCUTE which produce test cases combining. These tools RANDOOP and jCUTE are plugged into COPECA so that, the hybrid tool will be capable of computing MC/DC%. Our proposed technique provides deep as well as wide exploration of concolic execution.

B. Detailed Description

Java-Hybrid Concolic Testing is the best combination to achieve better MC/DC, and is the hybrid combination of Feedback-Directed Random testing and Java Concolic Testing. We have inspired from the core-idea proposed by Majumdar et al.[2]. They proposed a Hybrid Concolic Testing algorithm, that interleaves random testing with concolic execution to obtain both a deep and a wide exploration of program state space. They have implemented their algorithm on top of concolic tester (CUTE) and experimented to obtain high branch coverage for two large programs;*VIM 5.7* and *Red black tree*. Similarly, we extend the work of Majumdar et al.[2] by measuring MC/DC and that too for Java programs. Majumdar et al.[2] implemented their algorithm using undirected random testing and concolic testing, whereas we propose an efficient hybrid concolic testing for Java programs i.e. Feedback-Directed Random testing with concolic testing to obtain high MC/DC.

Fig. 1 shows the proposed tool for Java-Hybrid Concolic Testing (Java-HCT)[3]. Java-HCT is developed by integrating seven modules. The process starts by supplying a Java program. From Fig. 1 we can observe that, this Java program is converted into two different syntaxes using *Syntax_converter*. Since, we supply this Java program into both RANDOOP and jCUTE, it is essential to convert the original Java program into respective tool syntaxes. Now, the Java program in RANDOOP syntax is supplied to *Random tester for Object-Oriented Programs (RANDOOP)* to generate *TCs_R* automatically. Similarly, the Java program in jCUTE syntax is supplied to *Java Concolic Unit Testing Engine (jCUTE)* to generate *TCs_J* automatically. Unfortunately, TCs_R and TCs_J are not in same syntax. Therefore, *TCs Extractor* module is uses both the test suites to extract the input values those are present in TCs_R and TCs_J as described in Fig. 1. Then all extracted input values are supplied to *TCs Combiner* to produce Total test cases. Since, these test cases may be redundant and useless for MC/DC, therefore we have developed *TCs Minimizer* that accepts all the input values and checks which are essential to compute MC/DC percentage and removes rest of the those non-essential test cases. Now, the minimized test cases are supplied to *COverage PErcenatge CAlculator (COPECA)*. Since, we focus to increase MC/DC percentage, so we have developed this COPECA to measaure MC/DC percentage. COPECA accepts the minimized test cases along with the original Java program to produce MC/DC%.

[3]https://sourceforge.net/projects/java-hct/

C. Algorithmic Description

Algorithm 1 deals with the pseudocode of Java-Hybrid Concolic Testing (Java-HCT). We supply a Java program to our algorithm. Java-HCT to produce MC/DC%.

Algorithm 1 Java-HCT

Input: J (Java Program)
Output: MC/DC%

1: $J_R, J_J \leftarrow Syntax\_Converter(J)$
2: $TCs\_R \leftarrow RANDOOP(J_R)$
3: $TCs\_J \leftarrow jCUTE(J_J)$
4: $Input\_values \leftarrow TCs\_Extractor(TCs\_R, TCs\_J)$
5: $Total\_TCs \leftarrow TCs\_Combiner(Input\_values)$
6: $Minimized\_TCs \leftarrow TCs\_Minimizer(Total\_TCs)$
7: $MC/DC\% \leftarrow COPECA(J, Minimized\_TCs)$
8: **return** $MC/DC\%$

In Line 1 of Algorithm 1, the Syntax_Converter takes a Java program as input and produces a Java program in RANDOOP syntax (J_R) and a Java program in jCUTE syntax (J_J) as outputs. Line 2 shows the execution of RANDOOP tool. RANDOOP takes J_R as input and generates test cases (TCs_R) as output. Line 3 shows the execution of jCUTE tool. J_J as input and generates test cases (TCs_J) from jCUTE as output. Now, these two generated test case sets (TCs_R, TCs_J) are forwarded to Test Cases Extractor (TCs Extractor) modules to separate each input values as presented in Line 4.

Line 5 shows the execution of Test cases Combiner (TCs Combiner). This Combiner module collects all the input values created from TCs Extractor and generate a single set called Total Test Cases (Total TCs). Line 6 shows minimization of the test cases generated through Test Cases Minimizer (TCs Minimizer). This module produces the Minimized Test Cases (Minimized TCs).

Line 7 deals with the computation of MC/DC% through COPECA. COPECA takes the original Java program along with the Minimized TCs as input. Line 8 returns the final MC/DC% as output.

IV. Experimental Study

In this section we discuss our experimental setup, the result analysis, and threats to validity.

A. Setup

The experimental programs are ran on a computer system with 4GB of memory (RAM) Intel(R) Core(TM)i5 CPU 650 @ 3.20 GHz 3.19 GHz and 32-bit operating system.

B. Result Analysis

Table I deals with the characteristics of forty input Java programs. Column 3 shows the size of programs in Lines of codes (LOCs). Columns 4,5,6, show the Predicates, Conditions, and Variables respectively.

Table II presents the generated test cases and MC/DC% for RANDOOP, jCUTE, and Java-HCT. Column 3 shows the

TABLE II
RESULTS ON EXECUTION OF COPECA

Sl. No.	Program Name	RANDOOP TCs	jCUTE TCs	Total TCs	Minimized TCs	MC/DC_1%	MC/DC_2%	MC/DC_3%	Inc_1	Inc_2
1	SwitchTest	20	8	28	4	50	50	100	50	50
2	StringBuffer	12	9	21	20	40	50	80	40	30
3	ScopeCheck	20	25	45	30	77.77	83.33	100	23.23	16.67
4	MyQuickSort	5	5	10	3	100	100	100	0	0
5	MathCall1	70	10	80	45	16.66	46.15	69.23	52.57	23.08
6	MyInsertionSort	13	6	19	11	0	50	83.33	83.33	33.33
7	Condition	26	7	33	15	44.44	66.67	88.88	44.44	22.21
8	FruitSales	114	12	126	112	31.88	42.02	56.52	24.64	14.5
9	InsertionSort	28	10	38	20	71.42	78.57	85.71	14.29	7.14
10	Comparison1	93	27	120	81	27.90	41.86	58.13	30.23	16.27
11	DSort1	39	4	43	28	50	75	85	35	10
12	GradeCalculation	23	5	28	18	33.33	50	75	16.7	25
13	MarketSales1	43	8	51	23	52.94	64.70	88.23	35.29	23.53
14	FruitBasket1	271	8	279	59	39.47	50	60.52	21.05	10.52
15	BSTree	86	5	91	24	23.07	69.23	84.61	61.54	15.38
16	SwitchTest2	29	14	42	30	12.5	18.75	31.25	18.75	12.5
17	AssertTest	31	7	38	9	57.14	57.14	100	42.86	42.86
18	BubbleSort	43	8	56	21	35.71	42.85	64.28	28.57	21.43
19	DSort_BST	36	8	44	9	42.85	28.57	57.14	14.29	28.57
20	CAssume	73	6	79	10	71.42	85.71	100	28.58	14.29
21	Demo1	44	4	48	12	62.5	75	87.5	25	12.5
22	MarketSales2	313	11	324	78	69.38	73.46	73.46	4.08	0
23	MathCall2	38	11	49	20	57.14	64.28	71.42	14.28	7.14
24	Selection_Sort	53	9	62	18	35.71	42.85	50	14.29	7.15
25	Sorting_algo	343	9	352	73	28	50	70	42	20
26	SwitchTest3	5	11	16	4	50	100	100	50	0
27	StringBuffer1	15	7	22	23	86.66	86.66	100	13.34	13.34
28	StudentGrades	103	8	111	20	30	50	80	50	30
29	Testy	19	3	22	11	50	66.66	83.33	33.33	16.67
30	Weight	10	4	14	5	33.33	33.33	66.66	33.33	33.33
31	Weight_Exp1	25	10	35	33	95.45	95.45	95.45	0	0
32	Weight_Exp2	26	8	34	18	100	100	100	0	0
33	Wildlife1	40	6	46	32	7.14	17.85	53.57	46.43	35.72
34	Wildlife2	50	10	60	59	10	40	50	40	10
35	Zodiac	190	63	253	131	5.95	16.66	27.86	21.91	11.2
36	WBS	20	7	27	18	0	20	30	30	10
37	AssertTest2	42	13	55	40	38.09	66.67	76.19	38.1	9.52
38	HelloWorld	10	5	15	5	100	100	100	0	0
39	IFExample	12	7	19	7	50	100	100	50	0
40	IFSample	24	13	37	5	75	83.33	100	25	16.67

test cases generated by Feedback-Directed Random Tetsing. RANDOOP is the tool that generates these test cases. Column 4 presents the test cases generated by Java Concolic Unit Testing Engine (jCUTE). Column 5 shows the total test cases of RANDOOP and jCUTE. TCs Minimizer accepts these total test cases and only selects essential test cases according to MC/DC criterion. Column 6 presents the number of minimized test cases. Columns 7,8,9 deal with the MC/DC percentages achieved by RANDOOP, jCUTE, and Java-HCT respectively. These percentages are defined below:

Definition 5: MC/DC_1%: This MC/DC percentage is computed through RANDOOP and COPECA.

Definition 6: MC/DC_2%: This MC/DC percentage is computed through jCUTE and COPECA.

Definition 7: MC/DC_3%: This MC/DC percentage is computed through RANDOOP, jCUTE and COPECA or Java-HCT.

Column 10 and 11 deal with the increase in MC/DC. Column 10 is named as Inc_1 and shows the difference between MC/DC_1% and MC/DC_3% using Eq.2, whereas Column 11 named as Inc_2 shows the difference between MC/DC_2% and MC/DC_3% using in Eq.3.

$$\text{Inc\_1} = \text{MC/DC\_3\%} - \text{MC/DC\_1\%} \quad (2)$$

$$\text{Inc\_2} = \text{MC/DC\_3\%} - \text{MC/DC\_2\%} \quad (3)$$

We have experimented forty Java programs. We computed the values of Inc_1 and Inc_2 for these programs which are 29.91% and 16.26% (on average) respectively. According to the observation of our experimental study, Java-HCT achieved better MC/DC by $\times$ 1.62 as compared to RANDOOP and by $\times$ 1.26 as compared to jCUTE.

V. COMPARISON WITH RELATED WORK

Majumdar et al. [2] presented a hybrid concolic testing for C programs. They have proposed an algorithm that interleaved random testing with concolic testing to achieve both a deep and a wide exploration of program state space. They had implemented their algorithm on top of CUTE tool and applied it to achieve better branch coverage for two large C based applications. For the same testing budget, almost they obtained $4\times$ branch coverage and $2\times$ branch coverage of random testing and concolic testing, respectively. We inspired from Majumdar et al. [2]'s core idea and proposed a new technique called Java-Hybrid Concolic Testing, which is implemented in Java language. Java-HCT is the combination of Feedback-Directed Random Testing and Java Concolic Testing.

Ganai et al. [12] and Ho et al. [13] proposed a technique of VLSI design validation where a combination of formal (symbolic execution or BDD based reachability) and random simulation engines were used to improve the design coverage

for big scale designs. Our proposed approach combines the Feedback-Directed Random Testing and Java Concolic Testing for Java programs to obtain better MC/DC.

Pacheco et al. [14], [8] presented a technique that improved random test generation by incorporating feedback obtained from executing test cases as they were created. Their proposed approach produced a test suite consisting of Java unit tests for the classes to be tested. Their experimental study showed that, use of feedback-directed random test generation was far better than systematic and undirected random test generation in term of coverage and error detection. In our approach, we used this improved random testing with the combination of Java concolic testing to obtain high MC/DC.

Sen et al. [9] proposed concolic testing in Java version called jCUTE and it is available online. jCUTE automatically selects the input values both symbolically and concretely, simultaneously. In our proposed work, we used this jCUTE tool to form Java-Hybrid Concolic Testing and obtained increased MC/DC.

Godefroid et al. [15] proposed an improved random testing technique by providing Directed fashion (Systematic way) combined with symbolic execution to generate test input values. In our proposed work, we used feedback-directed random testing instead of only directed because feedback-directed provides better code coverage. According to Pacheco et al. [8] RANDOOP is better in completeness and scalability, as compare to other approaches like DART. So, we have chosen feedback-directed technique to use.

Godboley et al. [16] proposed an approach to improve distributed concolic testing. They have proposed an approach for code transformation that supported to enhance MC/DC by generating extra test cases for C programs. Godboley et al. [17], [18], [19] has also developed transformation techniques for object oriented Java programs. They have also proposed green computation of testing tools.

VI. Conclusion and Future Work

To improve existing concolic testing and obtain high Modified Condition/Decision Coverage (MC/DC), we proposed a novel technique called Java-Hybrid Concolic Testing (Java-HCT). This technique is called as hybrid because it is the combination of two testing techniques Feedback-Directed Random Test and Concolic Testing. We experimented Java-HCT for forty Java programs and found there is an increase of 29.91% and 16.26% (on average), when compared to feedback-directed random testing and concolic testing respectively. We have improved MC/DC by $\times 1.62$ and by $\times 1.26$ in comparison to feedback-directed random testing and concolic testing, respectively.

In our future work, we will extend the proposed work by plugging with some transformation techniques to obtain better results.

References

[1] Csallner C, and Yannis S. 2004. *JCrasher: an automatic robustness tester for Java*, Software: Practice and Experience, Volume(34), Number(11), 10.1002/spe.602 pages 1025–1050.

[2] Majumdar R, and Sen K, May 2007. "Hybrid concolic testing," In proceedings of 29th International Conference on *Software Engineering 2007*, doi: 10.1109/ICSE.2007.41. ISSN 0270-5257 pages. 416–426.

[3] Bird D.L., and Munoz C.U., 1983."Automatic generation of random self-checking test cases," *IBM Systems Journal*, vol. 22, no. 3, pages. 229–245. doi:10.1147/sj.223.0229. 1983.

[4] Gupta N, Mathur A.P., and Soffa M.L., 1998. "Automated test data generation using an iterative relaxation method," In *Proceedings of the 6th ACM SIGSOFT International Symposium on Foundations of Software Engineering*, New York, NY, USA: ACM, doi: 10.1145/288195.288321. ISBN 1-58113-108-9 pages. 231–244. [Online]. Available: http://doi.acm.org/10.1145/288195.288321

[5] Xia S, Vito B.D., and Muñoz C., 2005."Automated test generation for engineering applications," In *Proceedings of the 20th IEEE/ACM International Conference on Automated Software Engineering*, New York, NY, USA: ACM. doi: 10.1145/1101908.1101951. ISBN 1-58113-993-4, pages. 283–286. [Online]. Available: http://doi.acm.org/10.1145/1101908.1101951

[6] Xie T., Notkin D., and Marinov D, 2004. "Rostra: a framework for detecting redundant object-oriented unit tests," In Proceedings of the 19th International Conference on *Automated Software Engineering*. doi: 10.1109/ASE.2004.1342737. ISSN 1938-4300, pages. 196–205.

[7] Visser W, Păsăreanu C.S., and Khurshid S.,2004. "Test input generation with java pathfinder," In *Proceedings of the 2004 ACM SIGSOFT International Symposium on Software Testing and Analysis*, New York, NY, USA: ACM, doi: 10.1145/1007512.1007526. ISBN 1-58113-820-2 pages. 97–107. [Online]. Available: http://doi.acm.org/10.1145/1007512.1007526

[8] Pacheco C., Lahiri S.K., Ernst M.D., and Ball T,2007. "Feedback-directed random test generation," In *29th International Conference on Software Engineering. ICSE 2007.*. doi: 10.1109/ICSE.2007.37. ISSN 0270-5257, pages. 75–84.

[9] Sen K., and Agha G,2006. "CUTE and jCUTE: Concolic Unit Testing and Explicit Path Model-Checking Tools," *Computer Aided Verification: 18th International Conference, CAV 2006, Seattle, WA, USA. Berlin, Heidelberg: Springer Berlin Heidelberg,* pages. 419–423. ISBN 978-3-540-37411-4. [Online]. Available: http://dx.doi.org/10.1007/11817963_38

[10] Ammann P, Offutt J, and Huang H, 2003."Coverage criteria for logical expressions," In *Proceedings of the 14th International Symposium on Software Reliability Engineering*, ISSRE '03. Washington, DC, USA: IEEE Computer Society. ISBN 0-7695-2007-3. pages. 99–108.

[11] Kelly H.J., Dan V.S., John C.J., Leanna R.K., 2001. "A practical tutorial on modified condition/decision coverage," Tech. Rep. Nasa.

[12] Ganai M.K., Aziz A., and Kuehlmann A.,1999. "Enhancing simulation with bdds and atpg," In *Proceedings of the 36th Annual ACM/IEEE Design Automation Conference*. DAC '99. New York, NY, USA: ACM. doi: 10.1145/309847.309965. ISBN 1-58113-109-7 pages. 385–390. [Online]. Available: http://doi.acm.org/10.1145/309847.309965

[13] Ho P.H., Shiple T., Harer K., Kukula J., Damiano R., Bertacco V, Taylor J, and Long J, 2000. "Smart simulation using collaborative formal and simulation engines," In *Int. Conf. on Computer Aided Design (ICCAD)*, pages. 120–126.

[14] Pacheco C., Lahiri S.K., Ernst M.D., and Ball T, 2006. "Feedback-directed random test generation," In *Technical Report MSR-TR-2006-125, Microsoft Research,*, pages. 75–84.

[15] Godefroid P., Klarlund N., and Sen K., 2005. "Dart: Directed automated random testing," In *Proceedings of the 2005 ACM SIGPLAN Conference on Programming Language Design and Implementation*. PLDI '05. New York, NY, USA: ACM, 2005. doi:10.1145/1065010.1065036. ISBN 1-59593-056-6 pp. 213–223. [Online]. Available: http://doi.acm.org/10.1145/1065010.1065036

[16] Godboley S., Mohapatra D.P., Das A., and Mall R., 2016."An Improved Distributed Concolic Testing", Software: Practices and Experiences, DOI: 10.1002/spe.2405

[17] Godboley S., Dutta A., Mohapatra D.P., Das A. and Mall R., 2016."Making a concolic tester achieve increased MC/DC.," Innovations in Systems and Software Engineering, pp.1-14, DOI:10.1007/s11334-016-0284-8 .

[18] Godboley S., Panda S., Dutta A. and Mohapatra D.P., 2016."An Automated Analysis of the Branch Coverage and Energy Consumption Using Concolic Testing.," Arabian Journal for Science and Engineering, pp.1-19, DOI:10.1007/s13369-016-2284-2.

[19] Godboley S., Dutta A., Besra B. and Mohapatra D.P., 2015, October. "Green-JEXJ: A new tool to measure energy consumption of improved concolic testing.," *In proceedings of Green Computing and Internet of Things (ICGCIoT),IEEE*, pp. 36-41,DOI: 10.1109/ICGCIoT.2015.7380424.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1715–1718
DOI: 10.15439/2016F170
ACSIS, Vol. 8. ISSN 2300-5963

A Development Process Based on Variability Modeling for Building Adaptive Software Architectures

Ngoc-Tho Huynh
IRISA / Université Européenne de Bretagne / TELECOM Bretagne, Brest, France
tho.huynh@telecom-bretagne.eu

Maria-Teresa Segarra
IRISA / Université Européenne de Bretagne / TELECOM Bretagne, Brest, France
mt.segarra@telecom-bretagne.eu

Antoine Beugnard
IRISA / Université Européenne de Bretagne / TELECOM Bretagne, Brest, France
antoine.beugnard@telecom-bretagne.eu

***Abstract*—Adaptive software is a class of software which is able to dynamically modify at runtime its own internal structure and hence its behavior in response to changes in its operating environment. Adaptive software development has been an emerging research area of software engineering in the last decade. Many existing approaches use techniques issued from software product lines (SPLs) to develop adaptive software. They propose tools, frameworks or languages to build adaptive software architectures (ASAs) but do not guide developers on the process of using them. In this paper, we propose an adaptive software architecture development process to guide developers building an ASA. One of the important activities of this development process is software specification based on models. In our process, we propose to use the models and basic tools of Common Variability Language (CVL, proposed as an OMG standard) to generate an ASA and a subprocess to specify these models.**

I. Introduction

MAINTENANCE phase is one important stage of a software development process. Maintenance aims at evolving and updating software to meet new requirements or to satisfy new conditions in the software execution context. In order to be maintained and evolved, software is usually stopped, then updated, and finally restarted. However, in certain circumstances, stopping the software is unacceptable, e.g., cloud gaming, medical, and finance systems as stopping the software has consequences for business, or even dangerous for humans.

One solution to solve this problem is to add dynamic reconfiguration mechanisms to modify the software architecture at runtime. Several works are interested in determining when the architecture reconfiguration can occur [1], [2]. Other works are interested in tools and methods to develop an ASA [3], [4], [5]. Particularly, they allow specifying variation points in the software architecture. A variation point is a particular point in the architecture specification where choices can be realized. In SPL, variation points are specified in a variability model. Such a model describes commonality and variability of the product line. Commonality represents common parts of all products in the family. Variability represents the parts that may be different in different software products [6]. A product contains all the common parts and the choices made on all the variation points. Once all the variation points are resolved (a choice has been made for all of them), the variability model is said to be configured.

Many existing works use techniques issued from SPLs to develop adaptive software [7], [8]. These works use a variability model to specify the software variability and propose the mapping between the variability model and the software architecture. When the variability model is configured, the corresponding component architecture is deduced. During software execution, a new configuration of the variability model can be decided and the corresponding software architecture deduced. Then, by calculating the differences between the current and the new architecture, reconfiguration actions are identified. However, these approaches do not specify a software development process to guide developers to build the adaptive software architecture.

To deal with the above limitation, we are working on a development process to guide developers to specify information needed to generate an ASA. In this paper, we focus on identifying the information to be specified in a development process. All along the development process we use CVL metamodels [9] to manage variability and propose a subprocess to describe how to specify this variability.

The remainder of the paper is structured as follows. Section II describes the CVL approach. Our general development process for building an ASA is presented in Section III. Section IV focuses on the variability modeling stage of our process and how a variability model is configured in order to generate an adaptive software. Related work is discussed in Section V then the paper concludes.

II. Common Variability Language

CVL [9] is a domain-independent language, and also an approach for specifying and configuring variability. We use CVL in our approach for two main reasons:

- CVL has been proposed as an OMG standard and we think it will be largely used in the near future.

- It proposes a MOF-based variability language which means that any MOF-based product model can be easily extended with variability information using CVL.

An overview of the CVL approach is depicted in Figure 1. The base model is used to specify the elements of the architecture that does not contain any information about variability. Variability information is specified in the variability model. In order to generate the configuration of a specific product, the resolution model consists of VSpecResolutions each determining a decision for a VSpec.

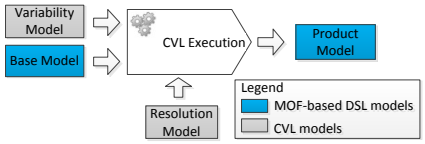


Fig. 1. CVL approach

In CVL, a variability model consists of three main parts:

- **VSpec tree.** A VSpec tree consists of VSpecs which are similar to features in feature models (FMs) [10]. There are four types of VSpecs: *Choice*, *Variable*, *VClassifier* and *CompositeVSpec*. A *Choice* allows to specify binary selections (true/false). A *Variable* should be used to specify a parameter which value may change. A *VClassifier* allows to specify a min and max number of instances of the VSpec. Finally, *CompositeVSpec* is used for modularity purposes.
- **Variation points.** A variation point links a VSpec to the corresponding elements in the base model.
- **OCL constraints.** CVL supports the definition of OCL constraints among VSpecs of a VSpec tree that cannot be directly captured by hierarchical relations.

Let *CVL model* denote the three models: the base model, the variability model, and the resolution model.

CVL offers tools and meta-models to specify variability of a product family but it does not offer a method to specify the variability model and the base model. Additionally, the product is generated without runtime variability.

III. A DEVELOPMENT PROCESS FOR BUILDING AN ADAPTIVE SOFTWARE ARCHITECTURE

In order to help engineers on developing adaptive software, we propose a development process that includes variability as a first-class stage. Although our process is based on CVL models and meta-models, other tools and frameworks may be used. The process we propose encompasses the variability specification issue to generate the architecture of the product. On the other hand, a reconfiguration process (at runtime) is out of the scope of this paper, but appears in the process in order to give an overview of the whole picture.

Our process is based on SPL engineering. The SPL engineering distinguishes two phases: domain engineering and application engineering [11] (see Figure 2).

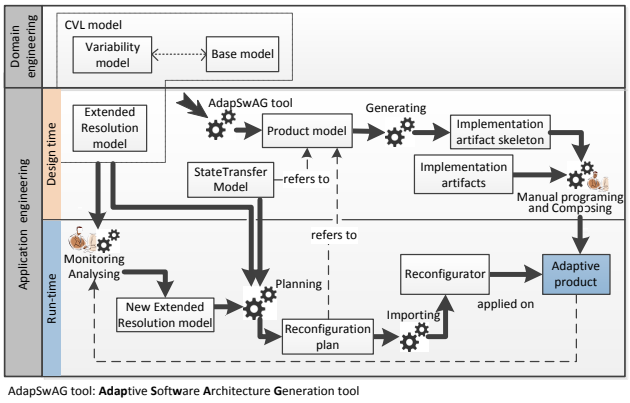


Fig. 2. General Adaptive Software Architecture Development Process

Domain engineering is the process responsible for defining the commonality and variability of the product line. The variability model and the base model are artifacts of this phase. The base model in our approach corresponds to a component-based architecture model specified by using an architecture description language (ADL).

Application engineering refers to the process of actually combining the artifacts obtained during the domain engineering phase, in order to generate a software product. In our approach, this phase is partitioned into two subprocesses: one at design time and another one at runtime.

At design time, a resolution model is specified to configure the variability model. This resolution model is extended from the CVL resolution meta-model for generating the necessary elements in an ASA. Generating the architecture is realized by our **Adap**tive **S**oft**w**are **A**rchitecture **G**eneration (AdapSwAG) tool. As depicted in Figure 2, the input of this tool is the CVL model and its output is a particular architecture (the product model). Compared to the CVL execution of the CVL approach, the software architecture generated by AdapSwAG tool includes part of its initial variability, and is specified in the same language as the base model. Based on the product model and a given target platform, a text generation module generates implementation artifacts skeleton. Each of them corresponds to a component specified in the product model and consists of component implementations, a component specification, and a configuration file. Once the implementation artifacts skeleton is generated, implementation artifacts that are either available or should be developed are integrated into it to build operational components. This task is performed by an application engineer. The result is an adaptive product that can be executed in the target platform.

If the product embeds variability, then, its architecture may change at runtime, and the components state affected by the changes should be migrated to the new ones. As the semantic of this state is component-specific, the state migration task is never completely automated. Thus, engineers have to specify a state transfer model that gives actions to effectively migrate state between components. The state transfer model represents the state mapping between previous components and the new

ones in the product model. It is specified at design time and used at runtime.

At runtime, because of new user requirements or execution environment changes, a change decision may be made. In this case, a new resolution model (new configuration) must be specified. This configuration may be either defined by engineers or computed thanks to analysis activities by a decision module integrated into the adaptive product. The current resolution model, the new one, and the state transfer model are processed to generate a reconfiguration plan that refers to the product model. Finally, the reconfiguration plan is injected into the reconfigurator to execute the reconfiguration actions. The moment when these actions are executed at runtime is important since the components must be driven to safe state (e.g., quiescent [1] or tranquility [2]), but this is out of the scope of this paper.

In this paper, we focus on the domain engineering phase and the design subprocess of the application engineering and give details on how to specify the VSpec tree, the base model, and the variation points to generate an ASA.

IV. Variability Modeling and Configuration

Variability modeling plays an important role in our approach. This section presents strategies to specify variability models according to CVL meta-model.

A. Variability Modeling Process

There may be different strategies to specify the variability and the base models.

- "Variability-driven process" (top-down approach): in this strategy the VSpec tree is specified first from information collected by engineers such as documentation or already existing products. Following this model, a base model may be built. Finally, variation points may be specified. This strategy allows specifying variability at high level of abstraction towards the diversity of components at concrete level.
- "Architecture-driven process" (bottom-up approach): in this strategy the base model is specified first. Then, VSpec tree is deduced and finally, variation points identified. This strategy allows specifying components at concrete level towards high level abstraction.
- "VSpec tree - base model independent process" (hybrid approach): in this strategy the VSpec tree and the base model are independently specified. Once the VSpec tree and the base model are specified, variation points can be identified to do the mapping between them. The advantage of this strategy is to allow independent specifications. Unlike the two first strategies, there is no guarantee to have a variation point for each VSpec.

As our approach is focused on guiding developers to identify information for generating adaptive products, we consider the first strategy as the process to be followed by engineers. In this strategy, variability modeling plays an important role. We focus on this task to represent the changes that an ASA may undergo. Based on the variability model, the base model can be created appropriately. The adopted strategy is depicted in Figure 3. In order to specify the VSpec tree and the base model, engineers need to identify variability and commonality. To do this, they can use documentation of existing applications of a product line, domain experts may also participate on the specification (step 1 in Figure 3). Once commonality and variability are identified, they are represented as a VSpec tree (step 2 in Figure 3). In step 3, a base model is defined with respect to the VSpec tree and the set of commonality and variability in the first step. Finally, step 4, variation points are defined to map VSpecs in the VSpec tree and elements in the base model. Steps 1 and 2 have to be manually realized by variability engineers. Step 3 should be realized by a software architecture expert. Step 4 can be manually (by an expert) or semi-automatically done thanks to similar characteristics of elements in the two models. To this end, in our development process, the domain engineering process encompasses all activities of this variability modeling process.

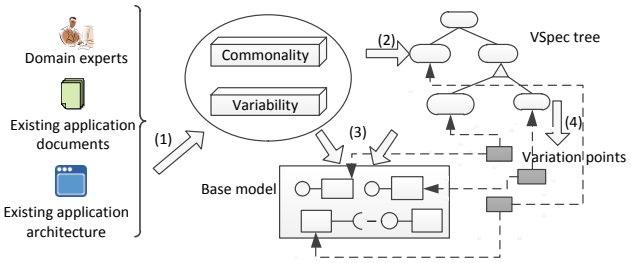


Fig. 3. "Variability-driven process" strategy

B. Product model generation

The first activity in Figure 2 shows the stage of the development process where the variability model is configured and the product model is generated. As previously mentioned, the product model is generated by configuring the variability model based on the resolution model. The task is realized by the AdapSwAG tool. It considers the CVL model as its input. Its output is a product model that contains the necessary components for adaptation. The AdapSwAG tool generates this product model by applying the following rules:

1) A component in the base model will be present and activated in the product if the corresponding *Choice* VSpec is resolved by a True value in the VSpecResolution.
2) A component in the base model will be present in the product at runtime if the corresponding *Choice* VSpec is not resolved by a True value in the VSpecResolution.
3) An attribute in a component in the architecture is assigned the value that is specified to the *value* attribute of its VSpecResolution.
4) Connections between components that are activated in the product are maintained in the product model.

Unlike the CVL approach, the generated product model contains all components in the base model and only connections between components which corresponding *Choice* VSpecs are resolved by True values in the resolution model.

V. Related work

SPL engineering has proven to be a well-suited methodology for developing a diversity of software products and software-intensive systems [11]. Several approaches are based on SPL engineering to develop adaptive software. Parra et al. [3] propose an approach to build a dynamic SPL. In this approach, activities such as analysis, composition generation, transformation, and runtime reconfiguration are separated into domain engineering and application engineering processes as in SPL engineering. Moreover, the role of developers in the process is defined such as application architect, platform architect, and asset developer. Lee et al. in [4] propose an approach to develop dynamically reconfigurable products. They introduce the activities of product line development, e.g., feature and feature binding analysis, behavior and functional specifications of feature binding units, product line architecture, and component design. They provide guidelines for designing dynamically reconfigurable architectures. Authors in [8] represent a process to automatically build a dynamic SPL from a feature model based on the assumption that a feature can be modeled as a component. This process consists of four steps: defining the core and dynamic architecture, adding the configurator, and defining the initial product. The reconfiguration action at runtime is simple to activate or deactivate components. The MADAM approach [5] is based on SPL techniques to build adaptive systems. They propose five steps to develop adaptive applications as follows: identifying fixed and varying user needs and resource constraints, designing the architecture, implementing the components identified by the architecture and deriving runtime plan objects, designing the utility function and property predictors for the components and composition.

Almost all these approaches use a feature model to specify variability. The feature model is configured to generate an initial product. However, they do not define how to specify variability and build an ASA. In our approach, an adaptive software development process based on SPL engineering with a set of specific steps has been proposed to guide engineers building an ASA. Activities to specify the variability model and the base model in our approach are considered as steps in the domain engineering process. The other activities are separated into two subprocesses in the application engineering process. Moreover, our approach identifies an explicit activity for maintaining the software state integrity based on specifying a state transfer model at design time.

VI. Conclusion

In this paper, we have presented a variability-driven development process for building an ASA. The process is based on CVL tools and meta-models to model variability of the software architecture but identifies an ordered set of tasks to be performed by engineers from variability specification up to the software architecture.

In order to validate our process, we have implemented a simple client-server application. The server is implemented with two different versions and can dynamically switch. In order to represent the variability of this application, a VSpec tree in CVL that conforms to the CVL meta-model is finally specified. Its architecture (base model) is specified by using ACME - an ADL. In order to generate a product, a resolution model is specified. Then, AdapSwAG tool generates a product model as an ASA. In this example, we have used the iPOJO component model and the Apache CXF framework to develop components. Based on their specification, a text generation module is implemented by using Xpand generator framework to generate implementation artifacts skeleton that includes packages. Each of them corresponds to a component specified in the product model and consists of component implementations, a component specification represented in XML documents, and a configuration file to create iPOJO components. To this end, the generated packages can be manually completed by an engineer, or the implementation artifacts developed independently by an engineer are integrated into them to build the iPOJO components of the adaptive product. Due to space constraints we can not give more details of our implementation, but the example is available at https://github.com/nthohuynh/.

References

[1] J. Kramer and J. Magee, "The evolving philosophers problem: Dynamic change management," *IEEE Transaction on Software Engineering*, vol. 16, no. 11, pp. 1293–1306, 1990. doi: 10.1109/32.60317

[2] Y. Vandewoude, P. Ebraert, Y. Berbers, and T. D'Hondt, "Tranquility: A low disruptive alternative to quiescence for ensuring safe dynamic updates," *IEEE Transaction on Software Engineering*, vol. 33, no. 12, pp. 856–868, 2007. doi: 10.1109/TSE.2007.70733

[3] C. Parra, X. Blanc, A. Cleve, and L. Duchien, "Unifying design and runtime software adaptation using aspect models," *Science of Computer Programming*, vol. 76, no. 12, pp. 1247 – 1260, 2011. doi: 10.1016/j.scico.2010.12.005

[4] J. Lee and K. C. Kang, "A feature-oriented approach to developing dynamically reconfigurable products in product line engineering," in *Proceedings of the 10th International on Software Product Line Conference*, ser. SPLC '06, Washington, DC, USA, 2006. doi: 10.1109/SPLINE.2006.1691585. ISBN 0-7695-2599-7 pp. 131–140.

[5] S. Hallsteinsen, E. Stav, A. Solberg, and J. Floch, "Using product line techniques to build adaptive systems," in *Proceedings of the 10th International on Software Product Line Conference*, ser. SPLC '06. Washington, DC, USA: IEEE Computer Society, 2006. doi: 10.1109/SPLINE.2006.1691586. ISBN 0-7695-2599-7 pp. 141–150.

[6] R. Capilla, J. Bosch, and K.-C. Kang, "Systems and software variability management: Concepts, tools and experiences," in *Systems and Software Variability Management*, 2013. doi: 10.1007/978-3-642-36583-6. ISBN 978-3-642-36582-9

[7] G. Pascual, M. Pinto, and L. Fuentes, "Self-adaptation of mobile systems driven by the common variability language," *Future Generation Computer Systems*, 2014. doi: 10.1016/j.future.2014.08.015

[8] P. Trinidad, A. R. Cortés, J. Peña, and D. Benavides, "Mapping feature models onto component models to build dynamic software product lines," in *1st SPLC Workshop on Dynamic Software Product Line (DSPL)*, Kyoto, Japan, 2007, pp. 51–56.

[9] O. Haugen, A. Wkasowski, and K. Czarnecki, "Cvl: Common variability language," in *Proceedings of the 17th International Software Product Line Conference*, ser. SPLC '13. New York, NY, USA: ACM, 2013. doi: 10.1145/2491627.2493899. ISBN 978-1-4503-1968-3 pp. 277–277.

[10] K. Kang, S. Cohen, J. Hess, W. Novak, and A. Peterson, "Feature-oriented domain analysis (FODA) feasibility study," Software Engineering Institute, Carnegie Mellon University, Pittsburgh, PA, Tech. Rep. CMU/SEI-90-TR-021, 1990.

[11] K. Pohl, G. Böckle, and F. van der Linden, "Software product line engineering: Foundations, principles, and techniques." Springer-Verlag Berlin Heidelberg, 2005. doi: 10.1007/3-540-28901-1. ISBN 978-3-540-28901-2 p. 467.

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1719–1726
DOI: 10.15439/2016F117
ACSIS, Vol. 8. ISSN 2300-5963

Efficient Data-Race Detection with Dynamic Symbolic Execution

Andreas Ibing
Chair for IT Security, TU München
Boltzmannstrasse 3, 85748 Garching, Germany

***Abstract*—This paper presents data race detection using dynamic symbolic execution and hybrid lockset / happens-before analysis. Symbolic execution is used to explore the execution tree of multi-threaded software for FIFO scheduling on a single CPU core. Compared to exploring the joint scheduling and execution tree, the combinatorial explosion is drastically reduced. An SMT solver is used to control a debugger's machine interface for adaptive dynamic instrumentation to drive program execution into desired paths. Data races are detected in concrete execution with available static binary instrumentation using hybrid analysis. State interpolation using unsatisfiable cores is employed for path pruning, to avoid exploration of paths that do not contribute to increasing branch coverage. An implementation in Eclipse CDT is described and evaluated with data race test cases from the Juliet C/C++ test suite for program analyzers.**

***Index Terms*—race detection; symbolic execution; interpolation; branch coverage.**

I. Introduction

A DATA race means, that there are concurrent accesses from different threads to the same variable, of which at least one is a write. Data race bugs are introduced in multi-threaded software when the developer forgets to lock a resource, that is shared between threads. Because races are observed only for certain thread interleavings depending on the scheduler's decisions, they are difficult to reproduce and sometimes called 'Heisenbugs'. Exactly locating feasible data races is known to be NP hard [1].

Data race detection is typically implemented as dynamic analysis with binary instrumentation. It follows happens-before analysis with vector clocks [2], or lockset analysis [3], or a hybrid algorithm [4]. These algorithms introduce some false positive and / or false negative detections. A prominent example is ThreadSanitizer [5], which is integrated with the GNU and Clang C compilers. It is reported to slow down execution speed by at least a factor of ten. Therefore, it is typically applied as dynamic analysis with a manually written test-suite. The achievable code coverage is then limited to execution coverage of the test suite. Therefore, data race detection in this approach also depends on the size of the available test suite, which is limited by cost constraints.

Symbolic execution [6] automatically performs a systematic program path exploration and enables program coverage independent of a test suite. Program input is treated as symbolic variables, and operations on the variables are translated into logic equations. The feasibility of program paths (feasible with any program input) is then decided with a Satisfiability Modulo Theories (SMT, [7]) solver. Symbolic execution is often applied as dynamic analysis in the form of 'concolic' (concrete and symbolic) execution [8], [9]. The program is executed with concrete input, while symbolic constraints are collected on the program path. The input for the next path is generated by the SMT solver, so that the program takes the desired path. Concolic execution offers the possibility for consistent concretization of formulas (fallback to concrete value). This is useful if certain constructs can not be handled with the solver. Concretization does not introduce false positive path satisfiability decisions or error detections, but in general introduces false negatives.

The prominent symbolic execution tools DART [8], CUTE [9] and KLEE [10] currently do not feature data race detection. Symbolic execution tools that do support race detection are jCUTE [11], Con2colic [12] and LCT [13]. They use a solver to search paths through the program's joint execution and scheduling tree, i.e., the solver determines both program input and thread scheduling. The combinatorial explosion can be partly mitigated with partial order reduction [14]. The resulting race detection with symbolic execution is considerably more complex and slower compared to symbolic execution of single-threaded code.

This paper presents data race detection using dynamic symbolic execution and hybrid lockset / happens-before analysis. Symbolic execution is used to explore the execution tree of multi-threaded software for FIFO scheduling on a single CPU core. Complexity and scaling of symbolic execution are improved by interpolation based path pruning.

The remainder of this paper is organized as follows: Section II motivates and describes the algorithm, Section III depicts its implementation. In Section IV, the implementation is evaluated with data race test cases from the Juliet test suite [15]. Related work is reviewed in Section V. Results of the experiments are discussed in Section VI.

II. Algorithm

A. Motivation

The algorithm is motivated by the following aspects:

- The operating system scheduler can be used in concolic execution for speed-up. Code execution is faster than interpretation. Symbolic execution needs a reproducible execution tree, also for multi-threaded software. This can be achieved by restricting the scheduling to one CPU core

and to reproducible scheduling independent of system load.
- Race detection works faster during concrete execution. Event tracing and instrumentation do not need a symbolic interpreter.
- The analysis of a program path should continue after the detection of a potential data race, in order to detect further errors along path extensions. This means, that actual races should be avoided while still detecting them. Program behaviour without races is independent of scheduling. Data races can be prevented by using FIFO scheduling on one CPU core. With this scheduling, potential data races can still be detected (using happens-before, lockset or hybrid analysis).
- State interpolation can be used to improve scaling of symbolic execution. Unsatisfiable branches can be interpolated by computing unsatisfiable equation cores [16]. The interpolation can be used to prune paths, that are redundant with respect to coverage. Unsatisfiable cores can be backtracked by approximate weakest-precondition computation. This requires depth-first path exploration.

An advantage of this approach is that bugs other than races (e.g., buffer overflows) can be found just like with symbolic execution of single-threaded code. Only one representative thread interleaving is analyzed per executed program path. With FIFO scheduling on one CPU core, there is also the possibility to apply standard code coverage criteria like in single-threaded execution.

B. Dynamic Symbolic Execution

Code execution is faster than interpretation, and especially faster than translation into logic formulas. Therefore, only as few code locations as needed are interpreted symbolically, otherwise the code is executed concretely. These locations are the definition of input-dependent variables (symbolic variables) and input-dependent branch decisions (dependent on a symbolic variables). Because it is context-sensitive which variables are symbolic, adaptive dynamic instrumentation is used. The program under analysis is executed concretely. If a program location needs formula generation, the analysis switches to symbolic interpretation and generates the constraint formula. Then, further dependent locations are marked for symbolic interpretation, and the concrete execution is continued.

1) Reproducible Execution Tree for Multi-Threaded Programs: Multi-threaded software can be described as state transistion system with a combined scheduling and execution tree. Considering a deterministic scheduling algorithm without outside parameters (independent of system load etc.), the execution tree for this scheduling is yielded. Here, FIFO scheduling on one CPU core is used. This avoids data races because only one thread at a time is active, and it is not preempted. Computations in a thread between calls to the scheduler become atomic. Data race detection is still possible. The execution further becomes reproducible: when restarted with the same program input, the identical thread interleaving is yielded.

2) Execution Tree Exploration: Symbolic program input is configurable. It can comprise command line parameters and system call return values. In the execution of the first program path, pre-configured stardard return values (that are always valid) from the symbolic system calls are used for concrete execution. Then, the solver is used to generate concrete input values for the next path from the collected constraints of the previous path. The last symbolic branch condition is negated, if the negation is not yet covered in this context. If the resulting equation system is unsatisfiable, the branch decision is backtracked. This path exploration is depth-first search. Without state interpolation and path pruning, it explores all satisfiable program paths.

3) Interpolation Based Path Pruning: The state interpolation uses unsatisfiable core (unsat-core) computation with serial constraint elimination as described in [16]. Given an unsatisfiable conjunction of formulas, an unsat-core is a subset of the formulas whose conjunction is still unsatisfiable. If input generation for a path is unsatisfiable, an unsat-core is computed from the path constraint. Unsat-cores are backtracked during depth-first path exploration. A constraint is removed from the unsat-core when the control flow graph (CFG) node is backtracked, for which the constraint was generated.

a) Pruning: Prune formulas are the backtracked unsat-cores. When a control flow decision node is backtracked, the conjunction of the branch prune formulas is used. Branch targets are used as potential prune points. When symbolic execution reaches a branch for that a prune formula has been computed, then the solver is used to decide whether the current path is redundant and can be pruned. If the current path's path constraint implies the prune formula, then the path is pruned. This approach can prune paths from different contexts [16]. The implication means that all branches, that were unsatisfiable in the previous context, are also unsatisfiable in the current context. Extensions of the path would therefore not contribute to increasing branch coverage [17].

C. Dynamic Race Detection

Data races are detected with hybrid dynamic analysis during concrete execution, using the ThreadSanitizer algorithm [5]. This subsection shortly reviews the algorithm. Happens-before analysis with vector clocks is combined with lockset analysis to reduce the number of false negative detections. The hybrid detection has false positive detections. False positives can be eliminated by adding annotations to the program. Binary instrumentation is used to trace the relevant events [5]:

- memory access events: read and write
- synchronization events: locking and happens-before arcs. Locking events are write lock, read lock, write unlock and read unlock. Happens-before events are signal and wait.

The events are traced as state machines in shadow memory. The state machines can be run as pure happens-before or as hybrid analysis, with configurable context tracing information (speed versus information).

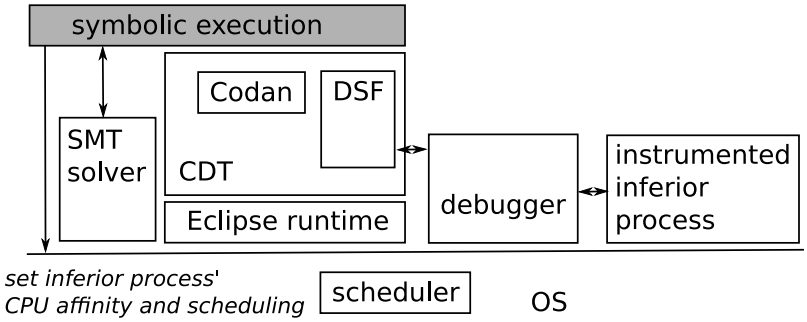


Fig. 1. Overview

III. Implementation

An overview of the main components is shown in Figure 1. These comprise the Eclipse runtime with plug-ins for the C/C++ development tools (CDT). CDT contains a code analysis framework (Codan) and debugger services framework (DSF). The instrumented program under test is controlled through DSF using a debugger (here the GNU debugger `gdb`). The debug inferior process is scheduled by the operating system scheduler. Symbolic execution is implemented as Eclipse plug-in on top of CDT. Logic formulas are decided using the Z3 SMT solver [18].

A. Debugger based concolic execution of multi-threaded code

The implementation extends the dynamic symbolic execution engine for single-threaded code presented in [17]. It uses the C/C++ parser from Eclipse CDT and the control flow graph builder from Codan. The debugger services framework is an abstraction layer over debuggers' machine interfaces. The program under test is executed in a debugger.

a) Selective symbolic interpretation: Only when the debugger hits a breakpoint, then a constraint is generated for the respective code location. The breakpoints are updated so that the debugger breaks on usage and definition of symbolic variables, then the debugger continues. At a breakpoint, the CFG node is resolved for the thread that stopped. Tree-based translation is used to generate the logic constraint. The debugger is queried for values of concrete variables where needed. Variables can become symbolic (by assignment of a symbolic value) and concrete (by assignment of a concrete value).

b) Static pre-analysis: Initial breakpoint locations are determined with static analysis before starting the symbolic execution. These locations are the definition of symbolic program input and input-dependent branches (branch locations that are input-dependent on any branch). In addition, the analysis overapproximates the set of pointers that might on any program path point to a symbolic target. If a target becomes symbolic during symbolic execution, a breakpoint is set on usage and definition of all pointers, that might point to this target. The static pre-analysis could be called 'maybe symbolic' analysis.

c) Single-core FIFO scheduling: The implementation currently runs on Linux, which supports different scheduling algorithms at the same time for different processes. Differing from the standard scheduler `SCHED_OTHER`, for the program under test the FIFO scheduler (`SCHED_FIFO`) is used. The CPU affinity is restricted to one CPU core. The corresponding Linux commands are `chrt` (to set the scheduling) and `taskset` (for CPU affinity).

d) Translation into SMT logic: The tree-based translation is implemented with CDT's abstract syntax tree (AST) visitor class. A control flow graph node references an AST subtree, that is traversed to generate a logic constraint. The translation uses bit-vectors and arrays. The solver is neither aware of multiple threads, nor of any scheduling. It just gets a conjunction of constraints that were collected along the program path.

e) Backtracking unsat-cores and path pruning: Unsat-cores are computed with serial constraint deletion as in [16]. There is one pass through the collected constraints on the path beginning from program start. Each constraint is tested once: if it can be removed and the rest remains unsatisfiable, then it is removed. Backtracking considers only CFG nodes that have been interpreted (where the debugger stopped). If a CFG node is backtracked, then any constraint generated for this node is removed. When a decision node is backtracked, the conjunction of the formulas from the branches is used [16]. When backtracking reaches a branch node, a prune formula is connected with this location. The prune formula is the conjunction of backtracked unsat cores. When a branch target is reached in (forward) symbolic execution, it is checked, whether there is a prune formula. If yes, then it is checked with the solver whether the path constraint implies the prune formula (i.e., whether the negation of the implication is unsatisfiable). In this case, the path can not contribute to increase branch coverage, and therefore is pruned.

B. Race detection in concrete execution using ThreadSanitizer

The implementation uses compiler instrumentation with `ThreadSanitizer` [5], which is featured by the GNU C compiler. The program under test is linked statically with the `ThreadSanitizer` library, and a breakpoint is set on the race detection error report function. If this breakpoint is hit,

```
1  #define N_ITERS 1000000
2  void CWE366_Race_Condition_Within_Thread__int_byref_12_bad() {
3    if(global_returns_t_or_f())      {
4      std_thread thread_a = NULL, thread_b = NULL;
5      int val = 0;
6      if (!std_thread_create(helper_bad, (void*)&val, &thread_a)) {
7        thread_a = NULL;
8      }
9      if (!std_thread_create(helper_bad, (void*)&val, &thread_b)) {
10       thread_b = NULL;
11     }
12     if (thread_a && std_thread_join(thread_a)) std_thread_destroy(thread_a);
13     if (thread_b && std_thread_join(thread_b)) std_thread_destroy(thread_b);
14     printIntLine(val);
15   } else {
16     std_thread thread_a = NULL, thread_b = NULL;
17     int val = 0;
18     if (!std_thread_lock_create(&g_good_lock)) { return; }
19     if (!std_thread_create(helper_good, (void*)&val, &thread_a)) {
20       thread_a = NULL;
21     }
22     if (!std_thread_create(helper_good, (void*)&val, &thread_b)) {
23       thread_b = NULL;
24     }
25     if (thread_a && std_thread_join(thread_a)) std_thread_destroy(thread_a);
26     if (thread_b && std_thread_join(thread_b)) std_thread_destroy(thread_b);
27     std_thread_lock_destroy(g_good_lock);
28     printIntLine(val);
29 } }
30 static void helper_bad(void *args) {
31   int *p_val = (int*)args;
32   for (int i = 0; i < N_ITERS; i++)     {
33     *p_val = *p_val + 1;
34 } }
35 int global_returns_t_or_f() {
36   return (rand() % 2);
37 }
```

Fig. 2. Example 'bad' function from [15] that contains a data race in line 33

the stack is traced back to a source file location, where the race is reported.

`ThreadSanitizer` supports dynamic annotations (C makros), which can be used if standard Posix threads are not used. They can also be used to eliminate false positive detections and to hide benign races [5]. Parts of the code can be marked as safe by the tool user. `ThreadSanitizer` can be run as happens-before or hybrid analysis. Also in pure happens-before mode it can report the involved locks. The slow-down by the instrumentation is reported as factor 20 – 50, and up to several hundred MB can be consumed for shadow memory [5].

IV. Experiments

a) Test cases and test setup: The implementation is evaluated with the data race test cases from the Juliet suite [15] for common weakness CWE-366 'race condition within a thread'. The test cases are 38 small artificial programs with 5-7 threads each. They contain 'good' functions (without data race) as well as 'bad' functions (that contain a data race) in order to measure false positive and false negative detections. There are two sets of 19 programs each. One set contains data races on global variables, the other contains data races on stack variables with access through pointers. Both sets cover the same 19 different data and control flow variants, that include conditional branches, loops, goto statements etc. The tests are run as JUnit plug-in tests with Eclipse 4.5.1

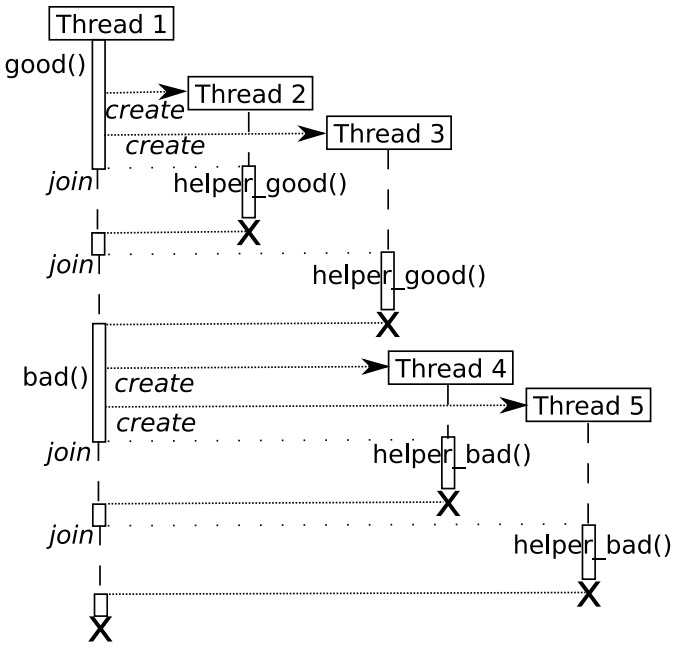


Fig. 3. Example, FIFO scheduling on one CPU core

and `gdb` version 7.10 on Linux kernel 4.2.0, on a Core i7-4650U CPU. The programs under test are run as unoptimized

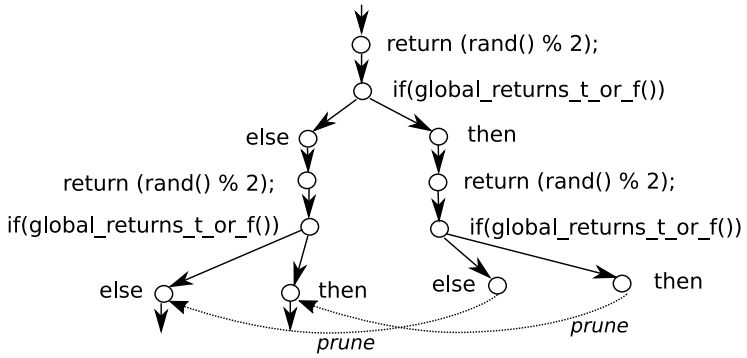


Fig. 4. Example, traversed part of execution tree

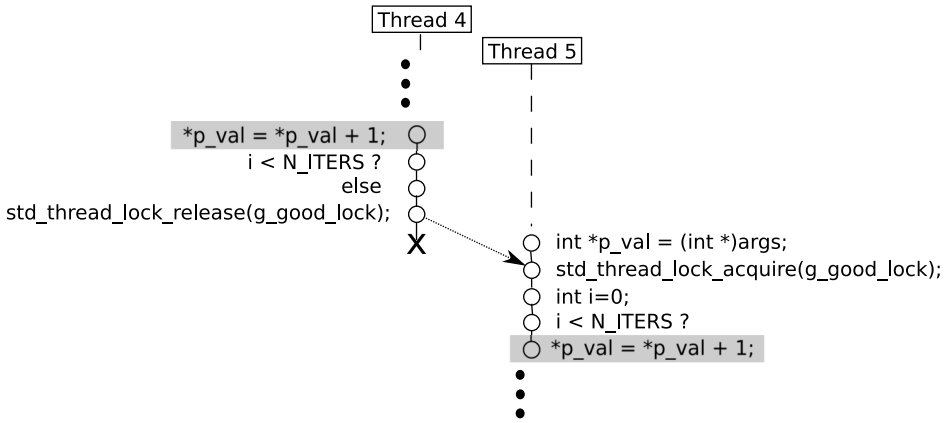


Fig. 5. Happens-before analysis, no data race for example 'bad' function if the `else` branch is taken (Figure 2 line 16)

code with default `ThreadSanitizer` settings without any annotations.

b) Example: An example 'bad' function is shown in Figure 2. It contains a data race on a stack variable with access by reference in line 33. The control flow depends on random input that is returned by a global function. The `rand()` call in line 36 generates this program input, i.e., the debugger breaks at this call, and the return value of `rand()` is treated as unconstrained symbolic variable. Before this 'bad' function, the program executes a similar 'good' function with proper locking in both branches, that also executes a loop with 10^6 iterations. The test program contains many branches, but few of them are input dependent. The loops in the 'good' and 'bad' functions are executed concretely. Because they contain only input-independent variables and branches, there is no need for symbolic interpretation. The symbolic execution finds four satisfiable program paths that non-deterministically depend on two `rand()` calls, and of which two paths exhibit the data race. FIFO scheduling on one CPU core for this program is illustrated in Figure 3. This figure looks the same for any of the four paths. The part of the execution tree (under this scheduling), that is traversed by symbolic execution, is illustrated in Figure 4. It only shows the locations where the debugger stopped. Two paths are followed to program end, the other two are pruned as indicated in the figure. The data race is accurately detected by `ThreadSanitizer`. An example part of the happens-before analysis is illustrated in Figure 5, for the `else` branch of the example function, where proper locking is used. The shaded memory access events are separated by a locking arc.

c) Results: All data races in the 38 test programs are accurately detected without false positives or false negatives. The (wall-clock) analysis runtimes are shown in Figure 7. The horizontal axis indicates the Juliet flow variant number. The average analysis runtime is below 2s. Error reporting through the Codan framework in the Eclipse GUI is shown in Figure 6.

V. Related Work

Exactly locating feasible data races is known to be NP hard [1]. The main approaches for practical race detection are usage of the happens-before relation with vector-clocks [2], or usage of locksets [3], or a combination of them. These approaches detect races when they might occur, i.e., not only with the current thread interleaving, but also with different interleavings. Main analysis methods are static analysis, model checking, dynamic analysis and symbolic execution.

a) Static Analysis: offers a possibility to detect races without false negatives. A method that requires programmer annotations and is based on type inference is described in [19]. The tools `RELAY` [20] and `LOCKSMITH` [21] apply lockset

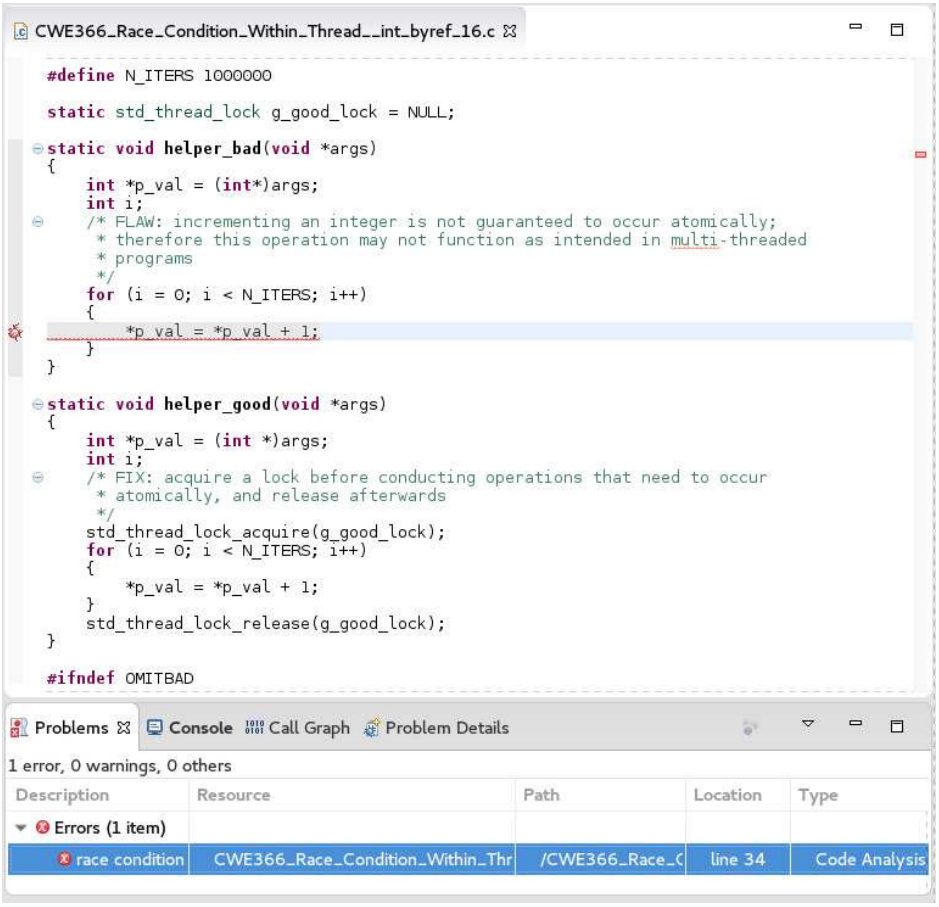


Fig. 6. Error reporting

analysis statically with data flow analysis. According to Palsberg [22], ”the best existing static technique” is implemented in `Chord` [23], but it ”reports a large number of false positives that would be daunting to examine by hand”.

b) Model Checking: In symbolic model checking, a program is translated into a logic formula, and properties are checked with an SMT solver. In theory, model checking allows for accurate race detection. It explores the symbolic state-space, that is the combined execution and scheduling tree. In practice, model checking does not scale well due to combinatorial explosion. One practical approach is bounded model checking [24], [25], [26], where the combined execution and scheduling tree is pruned with limits for the number of context switches and loop unrollings. Another way of pruning the tree is partial order reduction [27], [28], [14], [29], which prunes away irrelevant thread interleavings. Dynamic partial order reduction [14] traces the happens-before relation for thread interactions to find backtracking points for branching [14]. Optimal dynamic partial order reduction [29] explores a minimum number of representative thread interleavings.

c) Dynamic detection at runtime: is a practical way for race detection. It does not need a constraint solver. One technique is to instrument memory accesses and thread interaction with binary instrumentation and check for races using the happens-before relation [2] with vector-clocks. Happens-before analysis may have false negative detections depending on the scheduling. It is more sophisticated than lockset analysis, but scales worse with an increasing number of threads. `LiteRace` applies sampling, i.e., it monitors only a subset of all memory accesses. It can detect a majority of races by monitoring a small number of accesses [30]. `FastTrack` is an optimized implementation of happens-before analysis with reduced complexity [31]. `Pacer` [32] combines sampling with `FastTrack`. `DataCollider` [33] implements memory access sampling with hardware breakpoints and watchpoints and applies it to kernel code. Lockset analysis is used in `Eraser` [3]. It instruments memory accesses and traces thread locksets and variable locksets. If a variable access is not protected by a lock, then a warning is issued. Lockset analysis is lightweight and scales well. On the downside, it leads to more false positive detections than happens-before analysis. Hybrid race detection is presented in [4] as a two-pass solution. First, locksets are used to find problematic variables. Then, happens-before analysis is applied only to those variables. In [34], the DJIT algorithm is presented, which is a variation of happens-before analysis. `MultiRace` [35] combines DJIT with locksets to reduce false positives. A location’s lockset is reset

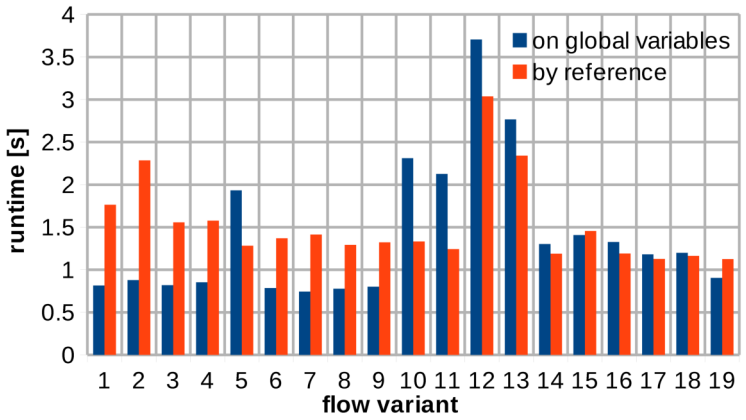


Fig. 7. Analysis runtime for data race tests from Juliet suite

at synchronization barriers. `ThreadSanitizer` [5] applies static binary instrumentation for happens-before and lockset analysis and is integrated with several current C compilers. Dynamic race detection is integrated in the managed runtime environments `RaceTrack` [36] and `Goldilocks` [37]. In [38], [39], it is proposed to integrate hardware acceleration for race detection into CPUs.

d) Symbolic execution: The prominent symbolic execution tools `DART` [8], `CUTE` [9] and `KLEE` [10] currently do not feature race detection. `jCute` [11] determines program input and thread schedule with the solver to explore different paths and interleavings, and it detects races when they occur. `Con2colic` also determines input and schedule with the solver. It implements a heuristic to first achieve branch coverage, and then explore an increasing number of context switches. Also `Con2colic` can detect a race when it occurs (no happens-before or lockset analysis). `LCT` [13] implements concolic execution with dynamic partial order reduction. In [22], the tool `Racageddon` is described. It starts with race candidates that have been found with an existing hybrid technique. It then uses concolic execution to search for input and schedule that lead to a real race (to remove false positives). `WHOOP` [40] considers races between pairs of entry points to driver code and runs a symbolic lockset analysis with SMT solver. In [16], [41], it is described that a program path can be pruned if the context implies a previously computed interpolant for the same program location. A combination of partial order reduction with interpolation based path pruning is described in [42].

e) This paper: differs from previous work on race detection with symbolic execution both in that it factors out the scheduling, and in that it applies hybrid data race detection during concrete execution. It extends own prior work [17] (on dynamic symbolic execution of single-threaded code with interpolation based path pruning) with support for multi-threaded execution and with data race detection.

VI. Discussion

Symbolic execution with FIFO scheduling on one core is used to automatically drive concrete execution into program paths of interest. The scheduling effectuates a reproducible execution tree for multi-threaded code. FIFO scheduling avoids triggering data races. Races are detected during concrete execution by instrumenting the program under test with the available `ThreadSanitizer`. The analysis is comparatively fast through concrete scheduling and concrete race detection. Path pruning is used based on interpolation of unsatisfiable branches with unsatisfiable cores. Implication checking with SMT solver assures that only paths are pruned, that can not contribute to increasing branch coverage. It is also possible to run the analysis in a virtual machine like `qemu`, that contains a `gdb` server.

Acknowledgement

This work was funded by the German Ministry for Education and Research (BMBF) under grant 01IS13020.

References

[1] R. Netzer and B. Miller, "What are race conditions?: Some issues and formalizations," *ACM Letters on Programming Languages and Systems*, pp. 74–88, 1992. [Online]. Available: http://dx.doi.org/10.1145/130616.130623

[2] L. Lamport, "Time, clocks, and the ordering of events in a distributed system," *Communications of the ACM*, vol. 21, no. 7, pp. 558–565, 1978. [Online]. Available: http://dx.doi.org/10.1145/359545.359563

[3] S. Savage, M. Burrows, G. Nelson, P. Sobalvarro, and T. Anderson, "Eraser: A dynamic data race detector for multi-threaded programs," *ACM Trans. Computer Systems*, vol. 15, no. 4, pp. 391–411, 1997. [Online]. Available: http://dx.doi.org/10.1145/268998.266641

[4] R. O'Callahan and J. Choi, "Hybrid dynamic data race detection," in *ACM Symposium on Principles and Practice of Parallel Programming*, 2003. [Online]. Available: http://dx.doi.org/10.1145/966049.781528

[5] K. Serebryany and T. Iskhodzhanov, "ThreadSanitizer: data race detection in practice," in *Workshop on Binary Instrumentation and Applications*, 2009, pp. 62–71. [Online]. Available: http://dx.doi.org/10.1145/1791194.1791203

[6] J. King, "Symbolic execution and program testing," *Communications of the ACM*, vol. 19, no. 7, pp. 385–394, 1976. [Online]. Available: http://dx.doi.org/10.1145/360248.360252

[7] L. de Moura and N. Bjorner, "Satisfiability modulo theories: Introduction and applications," *Communications of the ACM*, vol. 54, no. 9, 2011. [Online]. Available: http://dx.doi.org/10.1145/1995376.1995394

[8] P. Godefroid, N. Klarlund, and K. Sen, "DART: Directed automated random testing," in *Conference on Programming Language Design and Implementation*, 2005, pp. 213–223. [Online]. Available: http://dx.doi.org/10.1145/1064978.1065036

[9] K. Sen, D. Marinov, and G. Agha, "CUTE: A concolic unit testing engine for C," in *European Software Engineering Conference and International Symposium on Foundations of Software Engineering*, 2005, pp. 263–272. [Online]. Available: http://dx.doi.org/10.1145/1095430.1081750

[10] C. Cadar, D. Dunbar, and D. Engler, "KLEE: Unassisted and automatic generation of high-coverage tests for complex systems programs," in *USENIX Symp. Operating Systems Design and Implementation*, 2008.

[11] K. Sen and G. Agha, "CUTE and jCUTE: concolic unit testing and explicit path model-checking tools," in *Int. Conf. Computer Aided Verification*, 2006, pp. 419–423. [Online]. Available: http://dx.doi.org/10.1007/11817963_38

[12] A. Farzan, A. Holzer, N. Razavi, and H. Veith, "Con2colic testing," in *ESEC/FSE Joint Meeting on Foundations of Software Engineering*, 2013, pp. 37–47. [Online]. Available: http://dx.doi.org/10.1145/2491411.2491453

[13] K. Kähkönen, O. Saarikivi, and K. Heljanko, "LCT: A parallel distributed testing tool for multithreaded Java programs," *Electronic Notes in Theoretical Computer Science*, pp. 253—259, 2013.

[14] C. Flanagan and P. Godefroid, "Dynamic partial-order reduction for model checking software," in *ACM Symposium on Principles of Programming Languages*, 2005, pp. 110–121. [Online]. Available: http://dx.doi.org/10.1145/1047659.1040315

[15] T. Boland and P. Black, "Juliet 1.1 C/C++ and Java test suite," *IEEE Computer*, vol. 45, no. 10, 2012. [Online]. Available: http://dx.doi.org/10.1109/MC.2012.345

[16] J. Jaffar, A. Santosa, and R. Voicu, "An interpolation method for CLP traversal," in *Int. Conf. Principles and Practice of Constraint Programming (CP)*, 2009, pp. 454–469. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-04244-7_37

[17] A. Ibing, "Dynamic symbolic execution with interpolation based path merging," in *Int. Conf. Advances and Trends in Software Engineering*, 2016.

[18] L. deMoura and N. Bjorner, "Z3: An efficient SMT solver," in *Tools and Algorithms for the Construction and Analysis of Systems (TACAS)*, 2008, pp. 337–340. [Online]. Available: http://dx.doi.org/10.1007/978-3-540-78800-3_24

[19] M. Abadi, C. Flanagan, and S. Freund, "Types for safe locking: Static race detection for Java," *ACM Trans. Programming Languages and Systems*, vol. 28, no. 2, pp. 207–255, 2006. [Online]. Available: http://dx.doi.org/10.1145/1119479.1119480

[20] J. Voung, R. Jhala, and S. Lerner, "RELAY: static race detection on millions of lines of code," in *ACM Symp. Foundations of Software Engineering (ESEC-FSE)*, 2007. [Online]. Available: http://dx.doi.org/10.1145/1287624.1287654

[21] P. Pratikakis, J. Foster, and M. Hicks, "LOCKSMITH: Practical static race detection for C," *ACM Trans. Programming Languages and Systems*, vol. 33, 2011. [Online]. Available: http://dx.doi.org/10.1145/1889997.1890000

[22] M. Eslamimehr and J. Palsberg, "Race directed scheduling of concurrent programs," in *ACM Symposium on Principles and Practice of Parallel Programming*, 2014, pp. 301–314. [Online]. Available: http://dx.doi.org/10.1145/2692916.2555263

[23] M. Naik, "Effective static race detection for java," Ph.D. dissertation, Stanford University, 2008.

[24] S. Qadeer and J. Rehof, "Context-bounded model checking of concurrent software," in *TACAS*, 2005. [Online]. Available: http://dx.doi.org/10.1007/978-3-540-31980-1_7

[25] I. Rabinovitz and O. Grumberg, "Bounded model checking of concurrent programs," in *Int. Conf. Computer Aided Verification (CAV)*, 2005. [Online]. Available: 10.1007/11513988_9

[26] L. Cordeiro and B. Fischer, "Verifying multi-threaded software using SMT-based context-bounded model checking," in *ACM Int. Conf. Software Eng.*, 2011, pp. 331–340. [Online]. Available: http://dx.doi.org/10.1145/1985793.1985839

[27] P. Godefroid, "Partial-order methods for the verification of concurrent systems - an approach to the state-explosion problem," *Lecture Notes in Computer Science*, vol. 1032, 1996. [Online]. Available: http://dx.doi.org/10.1007/3-540-60761-7

[28] V. Kahlon, C. Wang, and A. Gupta, "Monotonic partial order reduction: An optimal symbolic partial order reduction technique," in *CAV*, 2009. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-02658-4_31

[29] P. Abdulla, S. Aronis, B. Jonsson, and K. Sagonas, "Optimal dynamic partial order reduction," in *ACM Symposium on Principles of Programming Languages*, 2014. [Online]. Available: http://dx.doi.org/10.1145/2535838.2535845

[30] D. Marino, M. Musuvathi, and S. Narayanasamy, "LiteRace: Effective sampling for lightweight data-race detection," in *ACM Conf. Programming Language Design and Implementation*, 2009. [Online]. Available: http://dx.doi.org/10.1145/1542476.1542491

[31] C. Flanagan and S. Freund, "FastTrack: Efficient and precise dynamic race detection," in *PLDI*, 2009. [Online]. Available: http://dx.doi.org/10.1145/1542476.1542490

[32] M. Bond, K. Coons, and K. McKinley, "PACER: proportional detection of data races," in *ACM Conf. Programming Language Design and Implementation*, 2010. [Online]. Available: http://dx.doi.org/10.1145/1806596.1806626

[33] J. Erickson, M. Musuvathi, S. Burchhardt, and K. Olynyik, "Effective data-race detection for the kernel," in *USENIX Symposium on Operating Systems Design and Implementation*, 2010.

[34] A. Itzkovitz, A. Schuster, and O. Mordehai, "Towards integration of data race detection in DSM systems," *J. Parallel and Distributed Computing*, vol. 59, pp. 180–203, 1999. [Online]. Available: http://dx.doi.org/10.1006/jpdc.1999.1574

[35] E. Pozniansky and A. Schuster, "MultiRace: Efficient on-the-fly data race detection in multithreaded C++ programs," *J. Concurrency and Computation: Practice and Experience*, vol. 19, no. 3, pp. 327–340, 2007. [Online]. Available: http://dx.doi.org/10.1002/cpe.v19:3

[36] Y. Yu, T. Rodeheffer, and W. Chen, "RaceTrack: Efficient detection of data race conditions via adaptive tracking," in *ACM Operating Systems Review*, 2005. [Online]. Available: http://dx.doi.org/10.1145/1095809.1095832

[37] T. Elmas, S. Qadeer, and S. Tasiran, "Goldilocks: a race-aware Jave runtime," *Communications of the ACM*, vol. 53, no. 11, pp. 85–92, 2010. [Online]. Available: http://dx.doi.org/10.1145/1839676.1839698

[38] P. Zhou, R. Teodorescu, and Y. Zhou, "HARD: Hardware-assisted lockset-based race detection," in *Int. Symp. High-Performance Computer Architecture*, 2007. [Online]. Available: http://dx.doi.org/10.1109/HPCA.2007.346191

[39] J. Devietti, B. Wood, K. Strauss, L. Ceze, D. Grossman, and S. Qadeer, "RADISH: always-on sound and complete race detection in software and hardware," in *Int. Symp. Computer Architecture*, 2012, pp. 202–212. [Online]. Available: http://dx.doi.org/10.1145/2366231.2337182

[40] P. Deligiannis, A. Donaldson, and Z. Rakamaric, "Fast and precise symbolic analysis of concurrency bugs in device drivers," in *Int. Conf. Automated Software Eng.*, 2015. [Online]. Available: http://dx.doi.org/10.1109/ASE.2015.30

[41] K. McMillan, "Lazy annotation for program testing and verification," in *Int. Conf. Computer Aided Verification (CAV)*, 2010, pp. 104–118. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-14295-6_10

[42] D. Chu and J. Jaffar, "A framework to synergize partial order reduction with state interpolation," in *Hardware and Software: Verification and Testing*, 2014, pp. 171–187. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-13338-6_14

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1727–1736
DOI: 10.15439/2016F173
ACSIS, Vol. 8. ISSN 2300-5963

Managing Big Clones to Ease Evolution: Linux Kernel Example

Kuldeep Kumar
Department of Computer Science and Information Systems,
BITS-Pilani, Pilani Campus, India
kuldeep.kumar@pilani.bits-pilani.ac.in

Stan Jarzabek
Faculty of Computer Science
Bialystok University of Technology,
Poland
s.jarzabek@pb.edu.pl

Daniel Dan
Info-Software Systems
ST Electronics Pte. Ltd.,
Singapore
ddan8807@gmail.com

***Abstract*—Successful software is often enhanced and adapted to the needs of new users. During evolution, a software system grows in size, becomes more complex, and costly to maintain. In this paper, we point to big clones—large granular duplicated program structures such as files or directories—as one of many reasons why this happens. Using the Linux kernel as an example, we show that big clones arise in the Linux kernel despite careful architecture design and a systematic approach for managing variability. We propose a solution to avoid these big clones by representing them as generalized templates in ART (Adaptive Reuse Technique). ART templates are constructed on top of the Linux code, without conflicts with the state-of-art techniques and tools used to manage the Linux kernel. Benefits include simplification of the Linux kernel due to non-redundancy, easier comprehension, and traceability of the change impact during evolution. The proposed technique is general and the Linux example discussed in this paper also illustrates general phenomena.**

I. Introduction

CHANGES in user features, design decisions, and platforms arise naturally during software evolution. Sometimes, these changes are contained at the architectural level [1]. But, most often the impact of changes spreads widely throughout the code. It has been reported that after years of evolution software systems grow in size, their structure decays, and become more and more difficult to maintain [2]. Evolution may lead to cloning [3]. New system versions are generally built by cloning (copy-paste-modify practice) code from the earlier versions. However, less cloning happens in advanced Software Product Line (SPL) solutions [4] where reuse and evolution are aided by systematic variability management rather than by cloning. There is a large body of research on reasons why clones arise—both within and across system versions—and whether clones are good or bad [5][6][7]. These studies show that designers may intentionally create certain clones to fulfill some design goals (e.g., for performance, readability, or yet other reasons) [5]. Other clones may result from careless design and can be refactored [6][8], and yet others may not play any useful role, but cannot be eliminated using conventional design techniques [7]. Nevertheless, cloning is a reality and there is need to deal with it [9]. No matter if clones are good or bad, it is beneficial to know where clones are in programs. It is particularly true for big clones such as duplicated files or directories. Big clones happen even if software evolution is systematically managed with variability management techniques [10]. In the paper, we use the Linux kernel to illustrate why this happens, and how we can manage big clones.

The Linux kernel is among the largest well documented evolving systems systematically managed with variability techniques [11]. In that sense, a family of the Linux-kernel versions forms an SPL whose reusable core assets include a carefully designed architecture, systematically identified and documented configuration options (SPL features), a code base managed with the C preprocessor (cpp) as a main variability technique, Kconfig, and other tools and techniques. The reason why we find big clones in the Linux kernel—and, we believe, in many other evolving systems—is that commonly used variability management techniques fail to avoid them in a convenient way.

Using generics (or C++ templates), we can non-redundantly represent similar classes differing in type parameters [12]. Big clones found in industrial systems need not be classes, but files or directories–program structures of any kind and size that differ in arbitrary ways, not just in type parameters. In this paper, we use Adaptive Reuse Technique (ART: https://sourceforge.net/projects/vclang/) to represent big clones as generalized templates. Like generics or C++ templates, ART templates can be instantiated in variant forms. Unlike other templates, ART templates can be built for groups of similar program structures of any kind (e.g., files or directories) that differ in variety of ways typically found in real systems. ART is an enhanced, lightweight and XML-free version of XVCL [13].

We briefly introduce variability management in the Linux kernel in Section II. In Section III, we discuss examples of big clones in Linux kernel version 3.10. Section IV describes how ART blends into Linux kernel development and use cycles. Sections V and VI describe explain how we manage big clones with ART. We evaluate the benefits and trade-offs of the proposed solution in Section VII. Related work and conclusions end the paper.

II. Variability Management in the Linux Kernel

Despite technological advancements in programming technologies, preprocessors are still indispensable. Preprocessing solves some niche problems better than other techniques do. One such problem area is variability management in software evolution and reuse. To some extent

we can manage variability at the level of software architecture [1]. But from architecture variability leaks to the code, and here that preprocessors along with configuration files and other similar techniques [14] become handy.

The Linux kernel developers applied clean architectural design, cpp, build tools, shell scripts, and other tools to facilitate adaptation of the Linux kernel to the specifics of target computers. Close to eleven thousand configuration options control the adaptation process. These options correspond to cpp parameters that navigate execution of cpp directives (such as #ifdef's) that select code relevant to the target computer of user's choice. The high-level configuration tool Kconfig maps these configuration options to the chains of relevant cpp directives embedded in the Linux code. Having selected required options, Kconfig automatically triggers execution of cpp directives and selects compilation units via the make utility to build a custom Linux kernel for a specific computer. Users do not have to understand the details of the adaptation mechanism. While the complexities of cpp instrumentation are hidden from the users, developers who maintain and extend the Linux kernel must understand code instrumented with cpp.

III. Motivating Example: Big Clones in the Linux

In the Linux kernel, the Journaling Block Device (JBD) provides an interface for the file system journaling. There are two directories namely /jbd and /jbd2 implementing this functionality, with /jbd2 being an evolutionary branch of /jbd. /jbd2 compatibly extends /jbd with new features such as support for 64-bit computers, check-summing of journal transactions, and asynchronous transaction commit block write.

Each directory consists of six files shown in Fig. 1. Much similarity in functionality and code (Table I) among files corresponding by names suggests that /jbd2 files were created by copying and modifying /jbd files. Fig. 2 sketches code snippets highlighting the code similarities and differences between the two checkpoint.c files.

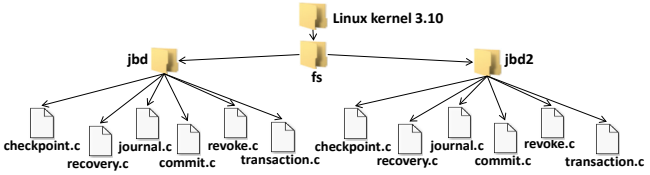


Fig. 1. Motivating example: cloned directories /jbd and /jbd2

Identical Code Fragments : ~554 LOC

```
51: static inline void __buffer_unlink(struct journal_head *jh)
52: {
53:         transaction_t *transaction = jh->b_cp_transaction;
54:
55:         __buffer_unlink_first(jh);
56:         if (transaction->t_checkpoint_io_list == jh) {
57:                 transaction->t_checkpoint_io_list = jh->b_cpnext;
58:                 if (transaction->t_checkpoint_io_list == jh)
59:                         transaction->t_checkpoint_io_list = NULL;
60:         }
61: }
```

```
51: static inline void __buffer_unlink(struct journal_head *jh)
52: {
53:         transaction_t *transaction = jh->b_cp_transaction;
54:
55:         __buffer_unlink_first(jh);
56:         if (transaction->t_checkpoint_io_list == jh) {
57:                 transaction->t_checkpoint_io_list = jh->b_cpnext;
58:                 if (transaction->t_checkpoint_io_list == jh)
59:                         transaction->t_checkpoint_io_list = NULL;
60:         }
61: }
```

Code Fragments with Parametric Changes: ~47 LOC

```
128: while (__log_space_left(journal) < nblocks) {
129:         if (journal->j_flags & JFS_ABORT)
130:                 return;
131:         spin_unlock(&journal->j_state_lock);
132:         mutex_lock(&journal->j_checkpoint_mutex);
```

```
124: while (__jbd2_log_space_left(journal) < nblocks) {
125:         if (journal->j_flags & JBD2_ABORT)
126:                 return;
127:         write_unlock(&journal->j_state_lock);
128:         mutex_lock(&journal->j_checkpoint_mutex);
```

Code Modification: ~12 LOC

```
333: set_buffer_jwrite(bh);
334: bhs[*batch_count] = bh;
335: __buffer_relink_io(jh);
336: jbd_unlock_bh_state(bh);
337: (*batch_count)++;
338: if (*batch_count == NR_BATCH) {
339:         spin_unlock(&journal->j_list_lock);
340:         __flush_batch(journal, bhs, batch_count);
```

```
311: journal->j_chkpt_bhs[*batch_count] = bh;

312: __buffer_relink_io(jh);
313: transaction->t_chp_stats.cs_written++;
314: (*batch_count)++;
315: if (*batch_count == JBD2_NR_BATCH) {
316:         spin_unlock(&journal->j_list_lock);
317:         __flush_batch(journal, batch_count);
```

Code Insertion: ~29 LOC

```
306: spin_unlock(&journal->j_list_lock);
```

```
276: transaction->t_chp_stats.cs_forced_to_close++;
277: spin_unlock(&journal->j_list_lock);
278: if (unlikely(journal->j_flags & JBD2_UNMOUNT))
279:         /* The journal thread is dead; so starting and
281:          * waiting for a commit to finish will cause
282:          * us to wait for a _very_ long time.*/
284:         printk(KERN_ERR "JBD2: %s: "
285:                 "Waiting for Godot: block %llu\n",
286:                 journal->j_devname,
287:                 (unsigned long long) bh->b_blocknr);
```

Code Deletion: ~95 LOC

```
520: journal_update_sb_log_tail(journal, first_tid, blocknr,
521:         WRITE_FLUSH_FUA);
522: spin_lock(&journal->j_state_lock);
523: /* OK, update the superblock to recover the freed space.
524:  * Physical blocks come first: have we wrapped beyond the end of
525:  * the log? */
526: freed = blocknr - journal->j_tail;
```

```
460: __jbd2_update_log_tail(journal, first_tid, blocknr);
```

Fig. 2. Motivating example: code snippets of cloned file /jbd/checkpoint.c (left) and /jbd2/checkpoint.c (right)

TABLE I. SIMILARITY AMONG FILES IN DIRECTORIES /JBD2 AND /JBD

File Name	Total LOC in corres-ponding jbd/jbd2 files	Identical LOC	LOC with parametric differences	Modified LOC	Inserted LOC	Deleted LOC
checkpoint.c	782/705	554	47	12	29	95
commit.c	1002/1192	523	93	35	364	218
journal.c	2122/2146	1266	287	29	690	229
recovery.c	594/862	420	52	12	234	0
revoke.c	740/769	544	94	3	25	0
transaction.c	2229/2348	1346	130	56	516	399

The directories /jbd and /jbd2 exemplify the situations that can benefit from ART as they cannot be effectively handled by other techniques. The reasons why we find such situations in the Linux kernel are functional similarities among different subsystems, extensions to the existing functionalities, adaptation of the existing subsystem code for the new one (incremental development), evolutionary development, and decentralized development [15][16][17].

IV. ART FOR THE LINUX KERNEL

In this section, we show how ART blends with the Linux-kernel development and uses cycles (Fig. 3). A *Linux Developer*, a member of an open-source community evolves the Linux kernel, e.g., by adding new devices into it. The *Linux SysAdmin* adapts the kernel for her computer using tools such as Kconfig.

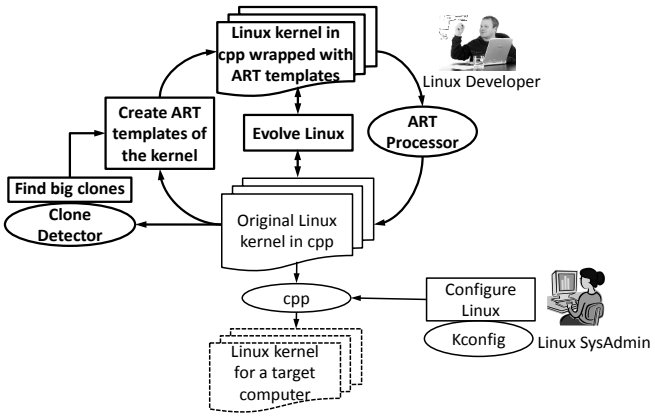


Fig. 3. An overview of Adaptive Reuse Technique (ART)

Big clones can be identified in the Linux kernel with aid of a suitable clone detector (we used Clone Miner [18]). The Linux Developer builds ART templates for big clones on top of the Linux code managed by cpp. From that point onwards, big clones are maintained via ART templates. ART templates do not affect the work of the Linux SysAdmin.

The ART Processor converts ART templates back to the original Linux code. The ART Processor instantiates templates in the same way as the C Preprocessor expand cpp directives. For example, for a template representing a group of similar files, the ART Processor generates code for these files based on specifications of deltas—differences between the template and each of these files. The generated Linux code is in the original form, and can be processed as normal by Kconfig, cpp, or make tool.

ART-template view of the Linux kernel and the original Linux kernel can be used together in two independent cycles of maintaining and using the kernel.

V. TYPES OF CLONES THAT WE HANDLE WITH ART

We categorized big clones in the Linux kernel based on their granularity.

A. Similar Directories

In the example of Section III, /jbd and /jbd2 play the same role, with /jbd2 being an evolutionary branch of /jbd addressing a new computer architecture and its capabilities. Each of the two directories contains six files, with much similarity between files corresponding by names. We found five other cases in Linux kernel following the pattern of /jbd and /jbd2, with the number of files in these directories varying between 19 and 46. In some cases, a directory contained one or more files that do not have similar counterparts in the cloned directory.

B. Similar Files

We found many cases of similar files within the same directory, as well as across directories. A common reason for replicating a file in the same directory is to make a certain existing functionality available for yet other computer architecture, device, or tool. An example is drivers for different brands of touchscreen devices—in directory /drivers/input/touchscreen, 10 files share the same structure and much code. Two directories having almost similar purpose (vide our motivating example) may contain similar files. Sometimes, the same or similar file may be required in two or more directories, even if these directories have not enough code similarity. For example, functionality for handling extended user attributes is needed in directories /fs/ext2, /fs/ext3 and /fs/ext4, therefore file "xattr_user.c" that defines this functionality appears in all three directories.

C. Duplicated Code Fragments

At times, creating templates for duplicated code fragments can be useful too, provided these fragments are long enough, play some specific role (e.g., represent some meaningful function), or recur in many places in programs. For example, code fragments in Fig. 4 implement a device specific queue handling procedure for different wireless network adapters. An instance of this code fragment occurs once in each of the files "rt2400pci.c", "rt2500pci.c", "rc2800pci.c" and "rt61pci.c", and twice in each of the files "rt2500usb.c", "rc2800usb.c" and "rt73usb.c".

VI. CONSTRUCTION AND PROCESSING OF ART TEMPLATES

In this section, we explain how we represented big clones as ART templates. We start with a brief overview of how ART works, followed by the explanation of how we build ART templates, illustrated with the Linux kernel example.

rt73usb.c

```
static void rt73usb_start_queue(struct data_queue *queue) {
        struct rt2x00_dev *rt2x00dev = queue->rt2x00dev;
        u32 reg;
        switch (queue->qid) {
        case QID_RX:
                rt2x00usb_register_read(rt2x00dev, TXRX_CSR0, &reg);
                rt2x00_set_field32(&reg, TXRX_CSR0_DISABLE_RX, 0);
                rt2x00usb_register_write(rt2x00dev, TXRX_CSR0, reg);
                break;
        case QID_BEACON:
                rt2x00usb_register_read(rt2x00dev, TXRX_CSR9, &reg);
                rt2x00_set_field32(&reg, TXRX_CSR9_TSF_TICKING, 1);
                rt2x00_set_field32(&reg, TXRX_CSR9_TBTT_ENABLE, 1);
                rt2x00_set_field32(&reg, TXRX_CSR9_BEACON_GEN, 1);
                rt2x00usb_register_write(rt2x00dev, TXRX_CSR9, reg);
                break;
        default:
                break;
        }
}
```

rc2800usb.c

```
static void rt2800usb_start_queue(struct data_queue *queue) {
        struct rt2x00_dev *rt2x00dev = queue->rt2x00dev;
        u32 reg;
        switch (queue->qid) {
        case QID_RX:
                rt2x00usb_register_read(rt2x00dev, MAC_SYS_CTRL, &reg);
                rt2x00_set_field32(&reg, MAC_SYS_CTRL_ENABLE_RX, 1);
                rt2x00usb_register_write(rt2x00dev, MAC_SYS_CTRL, reg);
                break;
        case QID_BEACON:
                rt2x00usb_register_read(rt2x00dev, BCN_TIME_CFG, &reg);
                rt2x00_set_field32(&reg, BCN_TIME_CFG_TSF_TICKING, 1);
                rt2x00_set_field32(&reg, BCN_TIME_CFG_TBTT_ENABLE, 1);
                rt2x00_set_field32(&reg, BCN_TIME_CFG_BEACON_GEN, 1);
                rt2x00usb_register_write(rt2x00dev, BCN_TIME_CFG, reg);
                break;
        default:
                break;
        }
}
```

Fig. 4. Sample code fragments from rt73usb.c and rc2800usb.c (differences highlighted)

A. An Overview of ART

For each group of clones, we distill common code into ART templates and mark the locations where clones differ one from another with ART commands (*italicized* for clarity in the description below). Fig. 5 outlines the overall solution, which consists of an ART-template hierarchy in which templates at the lower-level serve as building blocks for the higher-level templates. The ART templates are linked together by *#adapt* commands. The top-most template, called the specification file (SPC), specifies how to adapt other templates lower in the hierarchy to accommodate required variations. The ART Processor checks the templates for their conformance to the ART grammar definitions. It then traverses the template hierarchy in the depth-first order, starting with the SPC and performs adaptations by executing the ART commands embedded in the SPC and other ART templates. During traversal, each ART template adapts other templates from its sub-hierarchy. At the end, the ART Processor produces the required cloned instances.

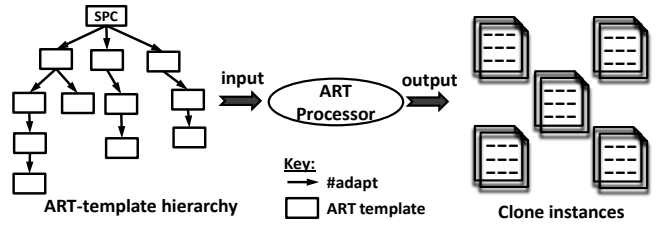


Fig. 5. An overview of the ART-template solution

Fig. 6 depicts steps in template processing. The ART Processor starts by reading the SPC (step-1). It fetches the ART commands step-by-step in the order in which they appear in the SPC (step-2). Whenever it hits an *#adapt* command (step-3), the processing will switch immediately to the adapted template (step-4) and switch back when the adapted template finishes its processing. Within a template, each ART command is processed one after another, in the same way as in the SPC. For the other commands, the Processor executes the ART command and builds the output (step-4') incrementally. Once the Processor reaches the end of the SPC (step-5), it generates the required source code files (step-6); if not, the ART Processor fetches the next ART command from the SPC (step-6').

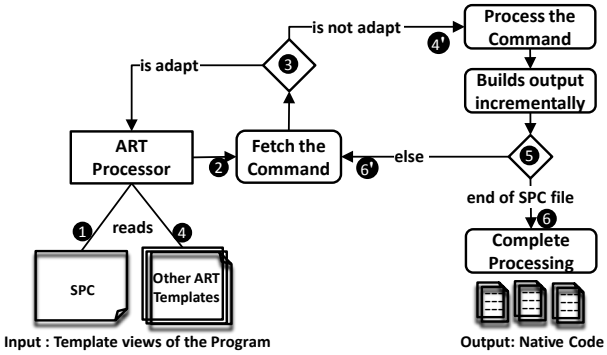


Fig. 6. Traversal mechanism of the ART Processor

B. ART-template Construction Mechanism

Despite a large fraction of the code common to all the clone instances (i.e., identical code fragments in the corresponding clone instances), as shown in Fig. 2, the three main types of differences among corresponding clone instances are parametric differences (code with parametric changes), alternatives (code modifications), and extras (code insertions and deletions).

The first task during the ART-template construction process is to identify these similarities and differences among corresponding clone instances. Once the corresponding similarities and differences are identified, ART templates record exact locations of these variation points at which the clone instances differ. ART commands can be used systematically to mark these variation points. Identical code fragments can be used directly as-it-is in the corresponding ART templates. ART variables treat parametric differences. The ART command *#select* allows choosing one among pre-defined alternatives (options), and *#insert* into *#break* mechanism handles additions and deletions of extra code.

C. Example: Template Construction for JBD

Fig. 7 shows the structure of ART solution for the JBD files. Each pair of similar files (e.g., checkpoint.c in /jbd and /jbd2) is represented by a template (e.g., checkpoint.art). The associated template checkpoint.spc specifies the differences between the two source files as deltas from checkpoint.art. The top-most template jbdX.spc navigates the process of instantiating the templates to form the Linux source files in their original form (i.e., instrumented with cpp).

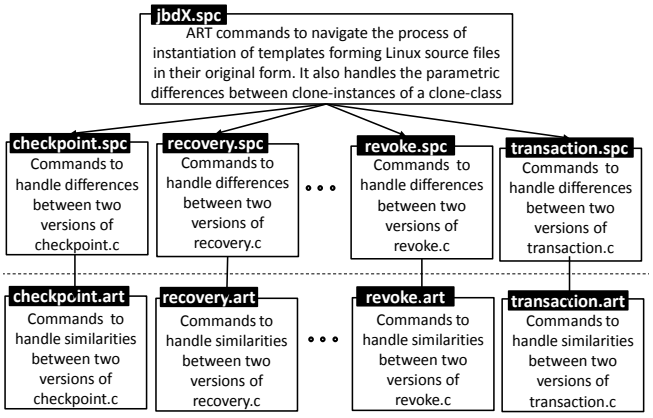


Fig. 7. Constructing ART templates: JBD example

Fig. 8 shows the details of ART templates. In jbdX.spc, ART variables are declared using *#set* commands (lines 1–6). Variable "dirName" is assigned two values, "jbd" and "jbd2" (line 2) that control the *#while* loop (line 7). The loop executes twice, with the value of "dirName = jbd" in the first iteration, and the value of "dirName = jbd2" in the second one. The Variable "fileName" is set to six values, each representing a file name (line 3).

The ART variable "action" helps represent lines:

*spin*_unlock(&journal->j_state_lock); //in jbd/checkpoint.c

*write*_unlock(&journal->j_state_lock); //in jbd2/checkpoint.c

in a single line in checkpoint.art (line 4):

*?@action?*_unlock(&journal->j_state_lock);

The two values of "action" are defined by:

#set action = "spin", "write" // line 4 in jbdX.spc

The generation loop defined in line 7:

#while dirName, action,…, tagByte

is controlled by a list of variables that cater for all parametric differences between the two checkpoint.c files. The command *#output* (line 9) instructs the ART Processor to create a directory and to place any further output into this directory (if the output file or directory is not specified, the ART Processor emits the code to an automatically generated default file named "defaultOutput"). Expression "?@fileName?" is used to fetch the value of an ART variable filename (line 9).

Similar to cpp's #include directive, an *#adapt* command (line 10 in jbdX.spc) instructs the ART Processor to include the designated template to the output. In addition, the *#adapt* command also tells the Processor to customize the designated template and assemble the customized result into the output. For example, given two ART templates *t* and *t'*, the statement "*#adapt t*" in template *t'* suspends processing of the current template (i.e., *t'*), and transfers processing to the template *t*. The ART Processor applies all the customizations specified under template *t*. Commands below *#adapt* in template *t'* indicate customizations to be applied after the template *t* is processed.

Variation points at which the two corresponding files (e.g., checkpoint.c) in /jbd and in /jbd2 directories differ are marked using ART commands—references to the ART variables, *#select*, *#break*, and possibly other commands. ART variables control selection of the code in case of alternative differences. This is illustrated as "*#select* dirName" in the template checkpoint.spc (line 4). *#option* (line 5 and 10 in checkpoint.spc) controls the variable values.

File checkpoint.c in one directory contains some extra lines as compared to checkpoint.c in another directory. These extra lines are specified using *#insert* commands in various "*#select* dirName" options. "*#insert* process_buffer" (line 11 in checkpoint.spc) propagates the code to "*#break* process_buffer" in checkpoint.art (line 12). *#insert-before* and *#insert-after* (line 6–9 in checkpoint.spc) add their code before or after the code contained in the matching *#break* (line 7 in checkpoint.art). While *#select* instruments a template with known variations, *#break* allows for extensions to a template in unexpected ways in the specific context of adaptation, without affecting others. These provisions for unexpected evolutionary changes give ART templates flexibility and stability.

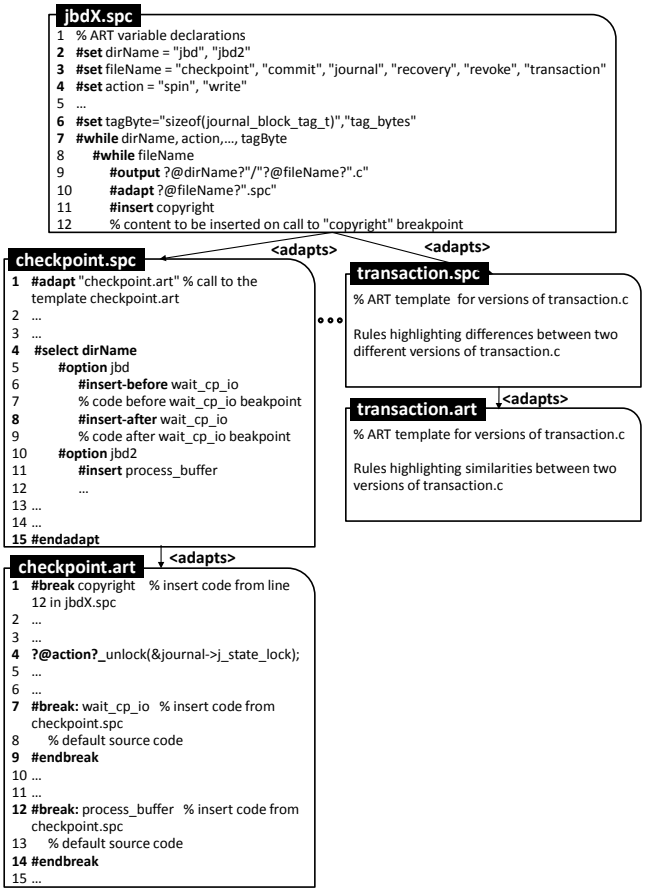


Fig. 8. Code snippet of ART templates for the JBD example

The ART Processor generates Linux code traversing the template hierarchy and emitting the code for the six files in the /jbd and /jbd2 directories from their respective templates. After that, the Linux code can be configured with Kconfig, and processed with cpp in the usual way. Template views expose the fact that the two directories and corresponding files in them are similar to each other, and also explicate every detail of similarities and differences among them. This information is implicit in the Linux code. Explicating it using

ART can be useful in further evolution of the JBD file system (Section VII).

D. Other Similarity Patterns at the Directory Level

Other cases of similar directories may not follow such a regular similarity pattern as in /jbd and /jbd2. For example, in the directories /drivers/infiniband/hw/qib and /drivers/infiniband/hw/ipath, in addition to similar files, /drivers/infiniband/hw/qib contains some extra files that do not have a counterpart in /drivers/infiniband/hw/ipath. Still, there is enough similarity in the concept and code between /drivers/infiniband/hw/ipath and /drivers/infiniband/hw/qib to build an ART-template solution for these two directories. The scheme used for building ART templates for /jbd and /jbd2 is also applicable in these situations, as templates manage pairs of similar files only and the remaining other files remain intact in the directories.

E. Constructing Templates for Similar Files

In this case, we deal with the similar files found in the same directory and the similar files in different directories, bearing in mind that directories as a whole are not considered good candidates for representing them as templates. For each such situation, we can create ART templates for similar files if we think that exposition of similarities and differences among these files can aid developers in reuse, program understanding, maintenance, and evolution of the Linux kernel. The solution follows the similar scheme as shown in Fig. 7 and Fig. 8.

VII. Evaluation

ART and its predecessor XVCL have been applied in industrial projects as a variability technique to manage reuse in product lines of web portals, and command and control systems [19][20]. In these projects, the productivity impact of applying the technique was measured and evaluated. An industry partner also participated in the Linux study described in this paper. In this section, we evaluate our ART solution for Linux, complementing it with lessons learned from other industrial projects with ART.

A. Reusing Templates within a Version of the Linux kernel

In a large system such as the Linux kernel, it is common to find clones within subsystems or modules, as well as across subsystems or modules. Each clone group can be managed by ART templates as long as such a non-redundant representation is deemed useful. Therefore, ART solution takes form of template hierarchies (Fig. 9) that explicates the location of clones and the exact nature of similarities and differences among replicated program structures. This knowledge is generally useful in understanding program design.

The example in Fig. 9 shows how ART templates reveal implicit couplings among bigger structures that contain repetitions. The same functionality defined in the templates commonConnectDisconnect.art and serioDriverStructure.art is needed in /touchscreen/common.art and /joystick/common.art. Templates for these two directories explicitly show the fact that this functionality is needed in both "touchscreen" and "joystick" drivers. If such implicit dependency among program modules is not documented, it may be overlooked during program evolution that may lead to errors.

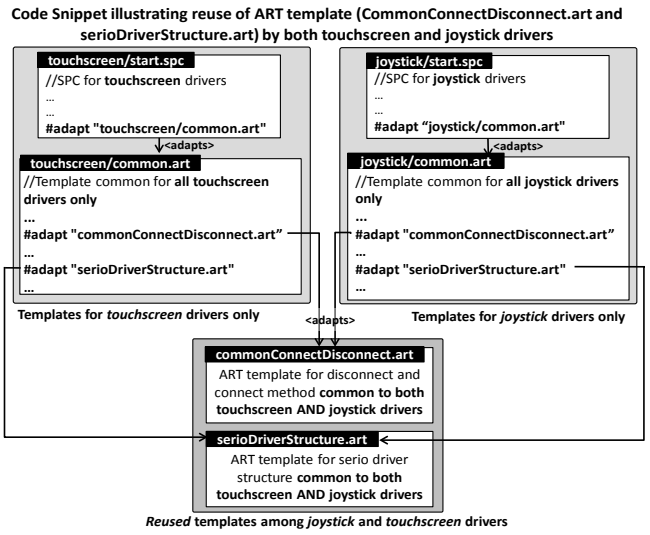


Fig. 9. Template reuse

B. Reusing Templates across Versions of the Linux kernel

Template reuse interconnects ART-template solutions developed for different groups of clones from the bottom, as shown in Fig. 9. It is also useful to interconnect partial ART-template solutions from the top, by introducing higher-level umbrella templates that trigger ART processing of some or all templates in the solutions.

Umbrella templates help developers manage multiple versions of the Linux kernel from the common base. A case study performed on 136 stable versions of the Linux kernel shows clone coverage of approximately 67% [21]. The coverage was found to be even higher between two consecutive versions due to small changes in successive releases of the kernel. Using umbrella templates, as shown in Fig. 10, we represented the commonalities between two versions, together with the version-specific code in different templates.

C. Handling Evolutionary Changes

Evolution often brings forward changes to the requirements and related code. For example, there might be a need to add a new directory /jbd3, or add more files to the JBD directories. ART has provisions to accommodate evolutionary changes to the templates (e.g., adding jbd3), without affecting existing code derived from the templates (e.g., jbd and jbd2).

Assuming that the new directory /jbd3 also contains six files that are similar to their counterparts in the /jbd and /jbd2, we need to make the following changes to the templates shown in Fig. 8:

jbdX.spc:

#set dirName = "jbd", "jbd2", **"jbd3"**

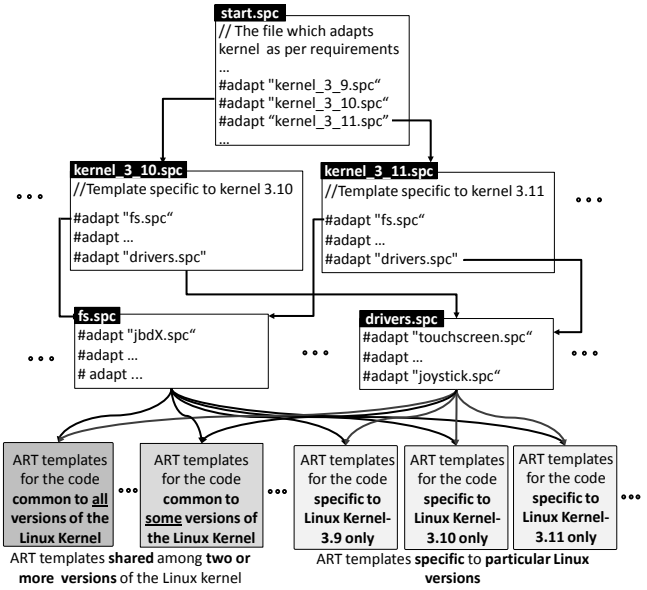


Fig. 10. Umbrella templates for an overall ART solution

```
#set fileName = "checkpoint",…, "recovery"
…
#while dirName , action,…, tagByte
#while filename
#output ?@dirName?"/"?@fileName?".c"
   #adapt: ?@fileName?".spc"
   …
checkpoint.spc:
   #adapt: "checkpoint.art"
   #select dirName
   #option jbd3
   …
% Customization to other templates with regards to jbd3
checkpoint.art:
   % Customizations to checkpoint.art specific to jbd3
% Customizations to other templates considering jbd3
```

In case of new variation points between the template and the file in /jbd3, we place new *#break* commands in the template. These new *#break* commands will cater for the differences specific to /jbd3, injected by *#insert*s in "*#option* jbd3" without affecting /jbd or /jbd2.

D. Aid in Program Understanding and Maintenance

Ease of comprehending program relations that matter during maintenance: ART templates enhance important relationships among program elements that matter to programmers trying to understand and modify the code. Instead of dealing with each directory or file separately, programmers can comprehend them in groups, and see the commonalities and differences among members of each group. It reduces ripple effects and the risk of update anomalies. In this way, if one wants to change a file, it is easy to check whether the changes also affect other files. For example, as illustrated in Fig. 7, similarities and differences are explicitly visible for jbd and jbd2 file systems. Such relations are generally hidden in conventional programs. Making them visible and easily tractable improves program maintenance. It also makes impact of changes easy to comprehend (as shown in Section VII.C).

Non-redundancy: ART templates eliminate redundant code from the software systems. For example, in our Linux experiment, ART templates reduced the size of code with redundancies by 30-50%. As both code and comments are important in software maintenance and program understanding, the upside of the proposed technique is that it is possible to manage both duplicated code and comments using it. ART allows a clean separation of various sources of changes that affect program during evolution. ART templates reduce the number of points at which affected changes must be made. Changes done to one template consistently propagate to all the contexts in which that template is adapted. Even if the changes are not uniform, adaptations can be made at specific variation points using the ART commands without directly modifying the code fragments. The ART template hierarchy explicitly reflects the impact of changes on the program structure. We can easily trace how different features affect the code.

Enhancing program understanding and conceptual integrity: According to Brooks [22], program understanding and conceptual integrity are among the most important considerations in system design. Big clones often embody domain-specific abstractions or design concepts. By formally capturing these abstractions and concepts, ART templates aid in program understanding and enhance conceptual integrity of the design.

Creating templates can be considered as refactoring at the meta-level: In some cases, developers seek to improve certain program qualities but due to some unavoidable reasons cannot achieve this at the code level. In such cases, we can do that at the level of meta-level templates. We benefit from non-redundancy at the meta-level templates, while still keeping repetitions in programs (as it is often desirable or unavoidable [7][23]).

Formally representing multiple design views: Program modules often belong to many logical groups that matter to developers at different times. Each logical partitioning reflects a certain aspect of program design that matters at a given time in the development in a given context. For example, for a given business function in business software, the modules for user interface, business logic, and database are usually implemented in different system partitions. Logically these modules belong to each other and sometimes we must know which modules implement a given business function completely. But, only one logical partitioning can be formally represented in a program physical structure. ART provides means to overlay programs with a web of meta-structures formally defining these logical partitions linked to code and without conflicts with the code.

Other Benefits: ART makes it easy for the programmers to do any program modifications and extensions at the template level. So there is no need to modify the code, and templates always remain in sync with the code as programs evolve.

In addition, with reference to the Linux kernel, aiding Linux Developers without affecting Linux SysAdmins is one cornerstone of the proposed solution. ART templates are non-intrusive, i.e., they do not affect the way cpp is normally used, but it improves understandability and maintainability of the programs instrumented with cpp. Also, the Linux code can be re-generated in the original form, and processed as normal by Kconfig, cpp or make tools. It provides a two-way view to understand and maintain the kernel—one with state-of-art variability technique (i.e., cpp), and another using ART.

Like cpp, ART manipulates code in an unrestrictive way, with no concern for the syntactic or semantic rules of the underlying programming language. Freeing from language constraints makes ART powerful to represent the groups of similar structures of arbitrary kind differing in arbitrary ways in generic forms.

E. Trade-offs and Threats to Validity of Results

The flexibility of manipulating the code in unrestricted way comes at the price of not being able to quarantine the correctness of the generated code. Unrestrictive program manipulation decreases the type-safety of the program. Also, the trade-off is between the benefits and cost of learning the new technique. ART syntax is very simple and consists of only few constructs (such as *#adapt*, *#insert-break* mechanism, *#while*). Yet, building quality ART templates require skilled experts, so that the benefits of ART outweigh the burden of learning and adopting it.

The benefit of ART depends on the degree of redundancy in a software system that cannot be fixed by simple refactoring. The bigger the size of software systems, the higher the likelihood of redundancies and evolutionary changes, and hence more will be the benefits of using ART. It follows that families of similar systems should be prime candidates for ART template views, as there is much similarity among components of such systems. Thus, the proposed technique seems to have more direct relevance in the SPL context, where we have the role of domain engineer who is responsible for building reuse-based productivity solutions that serve many systems in long run. ART templates belong to that category of solutions.

VIII. Related Work

We discuss related work on cloning in the Linux kernel and techniques that help programmers achieve non-redundancy, including XVCL (the predecessor of ART).

A. Cloning in the Linux Kernel

Cloning in the Linux kernel has been extensively studied in the literature [15][16][17][21][24] mainly focusing on the detection of small cloned code fragments. In the Linux v2.4.0, Casazza et al. [17] reported cloning of 15.5% between arch and drivers subsystems. They also reported 13.6% cloning between arch and kernel subsystems. Other study showed that file subsystem had 12% clone coverage in the Linux v2.4.19 [16]. In the Linux v2.6.37.6, 8% code similarity between drivers (/sound and /drivers directories) has been reported [15]. An empirical study of cloning among SCSI drivers is done by [24]. We found these cloning rates to be lower than those reported in similar studies for web applications [19] (60%-90%) or class libraries [7] (68%).

Compared to other studies, this paper aims at the detection and analysis of big clones instead of small cloned code fragments that are detected and analyzed by other studies.

B. Managing Redundancies in Software Systems

Simple-minded development often leads to cloning (copy-paste-modify practice). As mentioned earlier, cloning may also be done for good reasons [5]. Still, non-redundancy has been always considered an important quality of well-designed software. The Software Engineering principle of generality encourages avoiding repetitions and building parameterized software solutions that can be reused in many contexts. Macros were an early attempt to make programs adaptable to various contexts. Goguen popularized the ideas of parameterized programming [25]. Among programming language features, type parameterization [12] (called generics in Ada, Eiffel, Java and C#, and templates in C++), higher-order functions, and inheritance can help avoid repetitions in certain situations. Design techniques such as iterators, design patterns, table-driven design (e.g., in compiler-compilers), and modularization with information hiding are supportive in building generic programs. The Standard Template Library (STL) is a premier example of engineering benefits gained by generality [26]. Techniques have also been proposed to lift sufficient code similarity from the code to the architectural level [27][28].

ART uses templates and code generation to achieve non-redundancy. ART templates can represent any groups of clones (e.g., files or directories) with arbitrary differences among them (as opposed to only type-parametric differences in C++ templates or Java generics).

C. ART versus cpp + scripts and XVCL

One can also achieve non-redundancy by parameterizing and wrapping the code with cpp, shell scripts, and make files. An example of that can be found in the JDK buffer library described in [7]. SUN developers used cpp, scripts, and make files to build a non-redundant representation from which actual buffer classes are derived. A quick inspection of the code reveals that such representation may serve only its author and cannot be considered as a viable method to engineer programs.

Sample ART templates shown in the paper may also look complex. At the first glance they do. But, the fact is that ART is governed by only five important constructs (i.e., *#adapt*, *#output*, *#insert-break* mechanism, *#while*, and *#select*) that are neatly integrated to form a method that can be learned easily. Experience with XVCL (the ART predecessor) demonstrates that large code can be effectively managed achieving non-redundancy in the program areas where it matters [19]. Despite user-defined syntax, ART further

improves the user experience by providing the following improvements to XVCL:

Easy to learn: XVCL is a dialect of XML and uses XML trees and a parser for processing. ART parts with XML syntax and processing. It offers a cpp based flexible and more readable user-defined syntax. Just because cpp is so widely used, learning ART is not so tedious.

More generalized: Contrary to XVCL, developers can easily blend ART with the programming technologies of their choice. It is because the developer can define their own syntax and hence can avoid conflicts with the base languages.

Expanding the customization options under #adapt command: In XVCL, the only command that you can place under *adapt* is *insert*. ART allows to use any command under *#adapt*. We found using *#set*, *#while* and *#select* commands under *#adapt* to be particularly very useful.

Robust structure instead of unreadable loops: In XVCL, *while* loops using many multi-value variables can be quite confusing. ART introduces a structure called *set-loop* which gives the possibility to store and use more multi-value variables together as one loop descriptor data structure.

More flexible: ART is more flexible than XVCL, as it allows the adaptation of a file even though the file might not contain any ART commands. Such adaptation would simply copy the adapted file to the output stream.

D. Comparison with Other Techniques

Companies today often develop and maintain custom versions of the same software system for different customers using SPL [4]. The core idea is to manage the system family as a whole from a base of core assets designed for ease of adaptation in various reuse contexts.

In the SPL context, ART attempts to capture and streamline the end-to-end process of adapting software from the specifications of variant features (e.g., in Linux called configuration options) to the architectural structures and the code. ART templates can manipulate any textual file independent of their contents. So, it can also manage variability in documentation and test cases, keeping all textual SPL core assets in sync with evolving code.

Techniques proposed in research to manage variability in SPL are mostly based on the principle of separation of concerns (SoC), introduced by Dijkstra in early 1980's [29]. The goal of SoC is to deal with concerns one by one, independently from other concerns. When applied at the level of design and implementation, SoC attempts to compose software from components implementing different concerns. Concerns that nicely fit into conventional modules are easy to deal with. The challenge is to tackle cross-cutting concerns that are tightly coupled with the rest of a program, and cannot be easily modularized in a conventional way. There have been attempts to bring SoC down to the design and implementation levels. Aspect-oriented programming (AOP) [30], multi-dimensional SoC (MDSOC) from IBM [31], feature-oriented programming (FOP) [32], and colored IDE (CIDE) [33] are among the most widely published of such techniques. Among these techniques, AOP has been widely used. In AOP, various computational aspects are programmed separately and weaved at specified join points into the base program. AOP can separate a range of programming aspects such as synchronization, persistence, security transaction management, authentication/authorization, and others. Separated aspects can be easily modified and added/deleted to/from program modules. Because of that, a number of authors proposed AOP as a variability technique in the SPL. A study to test this hypothesis revealed difficulties in using AspectJ to deal with features that have chaotic impact on the base code [34]. While AOP deals with big chunks of functionalities (i.e., aspects) reasonably, it lacks a mechanism to handle variations at the lower levels of granularity. ART on the other hand, can handle variations at any levels of granularity. Walkingshaw et al. [35] provided a systematic and broader perspective on variational data structures. Properties related to program customizations are encapsulated in these variability-aware data structures.

IX. Conclusions

A study of industrial systems has shown that around 50% of small cloned code fragments tend to be contained in big clones [36]. While big clones are certainly intentional, they contribute to the increased program size and complexity. Therefore, big clones create a useful window from which to understand and manage clones at all levels of granularity.

In this paper, we presented a technique for managing big clones with non-redundant templates built with ART. ART templates manage big clones without conflicts with programming languages and other techniques used for managing variability. We demonstrated the technique with examples from the Linux kernel that uses cpp (among other techniques) to manage variability.

In various similarity groups, by unifying clones into non-redundant templates, ART eliminated 30-50% of the code. Non-redundant views revealed by ART templates improve program understanding. Program relations that have to do with the impact of changes are important in program understanding, maintenance and evolution, but remain mostly implicit in conventional programs. ART templates expose and explicate some of these program relations. For example, when maintaining duplicated code we often must know where such duplicates are and how they are different, in order to decide if and how each of them should be modified. ART makes such information more visible and tractable, reducing the risk of unexpected errors when changing programs.

ART blends without conflicts with the underlying programming language and any other techniques used to manage variability in a software system. Therefore, we can use ART to handle big clones, while other techniques (e.g., cpp and Kconfig in the Linux kernel) deal with other aspects of the overall variability management problem. Such seamless integration is necessary to allow the developers to painlessly inject ART templates into the projects in mature stages of evolution when big clones start emerging. ART syntax is user-defined to make such injection easy, without

affecting already existing software solutions and people who work with them. In the Linux context, ART can be viewed as an extension of cpp where ART commands syntactically resemble cpp directives, and can be incrementally learned as extensions that enhance reuse capabilities of cpp. The Linux Developers work with ART templates of the program, while the ART Processor generates the Linux code in its original form for the Linux SysAdmins.

Any new technique brings some overhead, requires learning and skillful application. ART is no different from other techniques in this respect. ART templates are not created for quick gains during development, but for long-term gains during software evolution and reuse. ART aims to benefit long-lived systems that undergo extensive evolutionary changes, or need to be tailored to the needs of multiple customers.

Acknowledgment

We are thankful to Ulf Pettersson, Technical Director, Info-Software Systems, ST Electronics Pte. Ltd., Singapore for applying ART in his projects and providing us with invaluable feedback.

References

[1] P. Clements and D. Muthig, (Editors) Proceedings Workshop on Variability Management–Working with Variation mechanisms, in *SPLC*, 2006, IESE-Report No 152.06/E Version 1.0, Germany, October 15, 2006

[2] C. L. Goues, S. Forrest, and W. Weimer, "The case for software evolution," in *FoSER,* 2010, pp. 205–210, http://dx.doi.org/10.1145/1882362.1882406

[3] R. Koschke, "Identifying and removing software clones", in *Software Evolution*, Springer Berlin Heidelberg, 2008, pp. 15–36, http://dx.doi.org/10.1007/978-3-540-76440-3_2

[4] P. Clements and L. Northrop, *Software product lines: practices and patterns*. Addition-Wesley, 2002

[5] C. Kapser and M. W. Godfrey, ""Cloning considered harmful" considered harmful," in *WCRE*, 2006, pp. 19–28, http://dx.doi.org/10.1007/s10664-008-9076-6

[6] G. P. Krishnan and N. Tsantalis, "Unification and refactoring of clones," in *CSMR-WCRE*, 2014, pp. 104–113, http://dx.doi.org/10.1109/CSMR-WCRE.2014.6747160

[7] S. Jarzabek and L. Shubiao, "Eliminating redundancies with a "composition with adaptation" meta-programming technique," in *ESEC/FSE*, 2003, pp. 237–246, http://dx.doi.org/10.1145/949952.940104

[8] S. Schulze, S. Apel, and C. Kästner, "Code clones in feature-oriented software product lines," in *GPCE*, 2010, pp. 103–112, http://dx.doi.org/10.1145/1942788.1868310

[9] R. Koschke, "Frontiers of software clone management," in *FoSM*, 2008, pp. 119–128, http://dx.doi.org/10.1109/FOSM.2008.4659255

[10] Y. Dubinsky, J. Rubin, T. Berger, S. Duszynski, M. Becker, and K. Czarnecki, "An exploratory study of cloning in industrial software product lines," in *CSMR*, 2013, pp. 25–34, http://dx.doi.org/10.1109/CSMR.2013.13

[11] R. Lotufo, S. She, T. Berger, K. Czarnecki, and A.Wąsowski, "Evolution of the Linux kernel variability model," in *SPLC*, 2010, pp. 136–150, http://dx.doi.org/10.1007/978-3-642-15579-6_10

[12] R. Garcia, J. Jarvi, A. Lumsdaine, J. G. Siek, and J. Willcock, "A comparative study of language support for generic programming," in *OOPSLA*, 2003, pp. 115–134, http://dx.doi.org/10.1145/949305.949317

[13] S. Jarzabek, P. Bassett, H. Zhang, and W. Zhang, "XVCL: XML-based variant configuration language", in *ICSE*, 2003, pp. 810–811, http://dx.doi.org/10.1109/ICSE.2003.1201298

[14] P. Ye, X. Peng, Y. Xue, and S. Jarzabek, "A case study of variation mechanism in an industrial product line," in *ICSR*, 2009, pp. 126–136, http://dx.doi.org/10.1007/978-3-642-04211-9_13

[15] A. Kadav and M. M. Swift, "Understanding modern device drivers," in *ASPLOS*, 2012, pp. 87–98, http://dx.doi.org/10.1145/2150976.2150987

[16] C. Kapser and M. W. Godfrey, "Toward a taxonomy of clones in source code: A case study," in *ELISA*, 2003, pp. 67–78

[17] G. Casazza, G. Antoniol, U. Villano, E. Merlo, and M. Di Penta, "Identifying clones in the Linux kernel," in *SCAM*, 2001, pp. 90–97, http://dx.doi.org/10.1109/SCAM.2001.972670

[18] H. A. Basit and S. Jarzabek, "A data mining approach for detecting higher-level clones in software," *IEEE Trans. Softw. Eng.*, vol. 35, no. 4, pp. 497–514, 2009, http://dx.doi.org/10.1109/TSE.2009.16

[19] U. Pettersson and S. Jarzabek, "Industrial experience with building a web portal product line using a lightweight, reactive approach," in *ESEC/FSE*, 2005, pp. 326–335, http://dx.doi.org/10.1145/1081706.1081758

[20] S. Jarzabek, U. Pettersson, and H. Zhang, "University-industry collaboration journey towards product lines," in *ICSR*, 2011, pp. 223–237, http://dx.doi.org/10.1007/978-3-642-21347-2_17

[21] S. Livieri, Y. Higo, M. Matsushita, K. Inoue, "Analysis of the Linux kernel evolution using code clone coverage," in *MSR*, 2007, pp. 22, http://dx.doi.org/10.1109/MSR.2007.1

[22] F. P. Brooks, Jr., "No silver bullet essence and accidents of software engineering," *IEEE Computer*, vol. 20, no. 4, pp. 10–19, 1987, http://dx.doi.org/10.1109/MC.1987.1663532

[23] M. Kim, V. Sazawal, D. Notkin, and G. Murphy, "An empirical study of code clone genealogies," in *ESEC/FSE*, 2005, pp. 187–196, http://dx.doi.org/10.1145/1081706.1081737

[24] W. Wei and M. W. Godfrey, "A study of cloning in the Linux SCSI drivers," in *SCAM*, 2011, pp. 95–104, http://dx.doi.org/10.1109/SCAM.2011.17

[25] J. A. Goguen, "Parameterized programming," *IEEE Trans. Softw. Eng.*, vol. SE-10, no. 5, pp. 528–543, 1984, http://dx.doi.org/10.1109/TSE.1984.5010277

[26] D. R. Musser, G. J. Derge, and A. Saini, *STL tutorial and reference guide: C++ programming with the standard template library*. Addison-Wesley Professional, 2009

[27] T. Mende, R. Koschke, and F. Beckwermert, "An evaluation of code similarity identification for the grow-and-prune model," *J. of Soft. Maint. & Evol.*, vol. 21, no. 2, pp. 143–169, 2009, http://dx.doi.org/10.1002/smr.402

[28] P. Frenzel, R. Koschke, A. P. J. Breu, and K. Angstmann, "Extending the reflexion method for consolidating software variants into product lines," in *WCRE*, 2007, pp. 160–169, http://dx.doi.org/10.1109/WCRE.2007.28

[29] E. W. Dijkstra, "On the role of scientific thought," in *Selected Writings on Computing: A Personal Perspective*, ed: Springer, 1982, pp. 60–66, http://dx.doi.org/10.1007/978-1-4612-5695-3_12

[30] G. Kiczales, J. Lamping, A. Mendhekar, C. Maeda, C. V. Lopes, J. M. Loingtier, and J. Irwin, "Aspect-oriented programming," in *ECOOP*, 1997, pp. 220–242, http://dx.doi.org/10.1007/BFb0053381

[31] P. Tarr, H. Ossher, W. Harrison, and S. Sutton, "N degrees of separation: multi-dimensional separation of concerns," in *ICSE*, 1999, pp. 107–119, http://dx.doi.org/10.1145/302405.302457

[32] D. Batory, J. N. Sarvela, and A. Rauschmayer, "Scaling step-wise refinement," *IEEE Trans. Soft. Eng.*, vol.30, no. 6, pp. 355–371, 2004, http://dx.doi.org/10.1109/TSE.2004.23

[33] C. Kästner, S. Apel, and M. Kuhlemann, "Granularity in software product lines," in *ICSE*, 2008, pp. 311–320, http://dx.doi.org/10.1145/1368088.1368131

[34] C. Kästner, S. Apel, and D. Batory, "A case study implementing features using AspectJ," in *SPLC*, 2007, pp. 223–232, http://dx.doi.org/10.1109/SPLC.2007.5

[35] E. Walkingshaw, C. Kästner, M. Erwig, S. Apel, and E. Bodden, "Variational data structures: exploring tradeoffs in computing with variability," in *ONWARD!*, 2014, pp. 213–226, http://dx.doi.org/10.1145/2661136.2661143

[36] H. A. Basit, U. Ali, S. Haque, and S. Jarzabek, "Things structural clones tell that simple clones don't," in ICSM, 2012, pp. 275–284, http://dx.doi.org/10.1109/ICSM.2012.6405283

Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1737–1746
DOI: 10.15439/2016F404
ACSIS, Vol. 8. ISSN 2300-5963

ReSA Tool: Structured Requirements Specification and SAT-based Consistency-checking

Nesredin Mahmud*, Cristina Seceleanu*, Oscar Ljungkrantz†
*Mälardalen University, Sweden, {nesredin.mahmud, cristina.seceleanu}@mdh.se
†Volvo Group Trucks Technology, Sweden, oscar.ljungkrantz@volvo.com

***Abstract*—Most industrial embedded systems requirements are specified in natural language, hence they can sometimes be ambiguous and error-prone. Moreover, employing an early-stage model-based incremental system development using multiple levels of abstraction, for instance via architectural languages such as EAST-ADL, calls for different granularity requirements specifications described with abstraction-specific concepts that reflect the respective abstraction level effectively.**

In this paper, we propose a toolchain for structured requirements specification in the ReSA language, which scales to multiple EAST-ADL levels of abstraction. Furthermore, we introduce a consistency function that is seamlessly integrated into the specification toolchain, for the automatic analysis of requirements logical consistency prior to their temporal logic formalization for full formal verification. The consistency check subsumes two parts: (i) transforming ReSA requirements specification into boolean expressions, and (ii) checking the consistency of the resulting boolean expressions by solving the satisfiability of their conjunction with the Z3 SMT solver. For validation, we apply the ReSA toolchain on an industrial vehicle speed control system, namely the Adjustable Speed Limiter.

I. Introduction

MOST often, the development of dependable automotive systems that are nowadays increasingly complex [1] relies on intricate requirements, given the nature of the system that has to interact with the environment. Therefore, the importance of establishing non-ambiguous and consistent requirements is even higher than for closed systems. Despite this acknowledged situation, current specification methods and tools [2][3][4] lack adequate support to formally analyze the logical consistency of high-level natural language requirements, in order to improve the quality of their specification.

Moreover, to be able to manage the complexity of automotive embedded systems during development, incremental model-based design approaches that assume multiple levels of abstraction are becoming appealing to industry. Among others, dedicated architectural languages, such as Electrical and Software Technology - Architectural Description Language (EAST-ADL) [5] are good candidates for such approaches. In EAST-ADL, an automotive system's structure and function are modeled at multiple levels of abstraction, that is, vehicle, analysis, design, implementation levels, and each abstraction level employs distinct concepts worth considering during requirements specification. For instance, the vehicle level of EAST-ADL abstraction describes the high level function of the system. Therefore, it would be inappropriate to use concepts from the design level, such as ports, signals, hardware elements, to describe requirements at the vehicle level, since such details usually hinder communication with non-technical stakeholders. Consequently, the requirements specifications need to be adapted to the appropriate levels of abstraction.

In this paper, we propose an Eclipse-based tool chain for structured requirements specification in ReSA [6], which scales to multiple EAST-ADL levels of abstraction. ReSA is an ontology-based requirements specification language tailored to automotive embedded systems development, which uses requirements boilerplates to structure the specification in natural language. Furthermore, we propose a consistency-check function that seamlessly integrates into the tool chain, for the automated consistency check of requirements using Z3 SMT solver [7]. The consistency checking is a preliminary task during elicitation and specification of requirements that paves the way for formal verification at later stages of software development. Our approach for consistency checking does not require a behavioral, or architectural model of the system, which might increase its attractiveness to industry as there is often the case that no system models exist for industrial systems. Checking for requirements consistency has been widely used in the field of requirements engineering, e.g., to describe consistent use of terms (words, phrases), logical consistency of requirements statements, or consistency between requirements and subsequent refinements [8][9]. The term can also refer to checking against type errors, or circular definitions [10]. In this paper, the consistency checking refers to checking the logical consistency of ReSA requirements specifications, in Z3.

Consistency checking of requirements specification helps detect possible logical errors at early stages of software development, and reduce the communication cost between manufacturers and suppliers [11]. However, checking for logical consistency of requirements expressed in natural language is not an easy task, mainly because: (i) unconstrained natural language is inherently ambiguous when it comes to reasoning, (ii) substantial assumptions used during requirements specification are hidden, and (iii) the size and complexity of requirements specifications are considerable.

In this work, we reduce the problem of checking the logical consistency of ReSA requirements to a boolean satisfiability problem, hence we propose algorithms for transforming the ReSA specification into boolean expressions, encode the latter into Z3 assertions, and perform consistency check using the Z3 SMT solver. The remainder of the paper is organized as follows. In section II, we recall the main features of ReSA,

978-83-60810-90-3/$25.00

the EAST-ADL levels of abstraction, xText grammar, and the boolean satisfiability problem. We introduce the ReSA toolchain in section III, after which we describe our consistency checking steps in section IV. The applicability of the tool is shown in section V, where we specify and check the consistency of sample requirements from an industrial use case, called the Adjustable Speed Limiter (ASL). We compare to related work in section VI, before concluding the paper in section VII.

II. Preliminaries

In this section, we overview the ReSA language, and its adaptation to EAST-ADL levels of abstraction, as well as the xText grammar, and the basic boolean satisfiability problem.

A. Overview of ReSA

ReSA [6] is an ontology-based requirements specification language tailored to automotive embedded systems development. The language (i) renders natural language terms (words, phrases), and syntax, (ii) uses an ontology that defines concepts and syntactic rules of the specification, and (iii) uses requirements boilerplates to structure specification.

1) Requirements Specification Ontology: A snippet of the ontology specification is shown below.

$$[System\ *x1][ActOnPara\ *x2][Para\ *x3] \quad (1)$$

$$(Is\text{-}fb\ ?x1\ ?x2)(Is\text{-}fb\ ?x2\ ?x3) \quad (2)$$

This ontology snippet defines requirements specification concepts (1), and syntactic rules between instances of concepts (2). The specification states that an instance of *System* precedes both an instance of *ActOnPara*, and an instance of *Para* in a requirement specification, e.g., `ASL:system shall control:ActOnPara vehicle speed:para`, is a valid example that conforms to the ontology specification.

2) Requirements Boilerplate: The language uses *requirements boilerplates (or boilerplates)* [12] in order to structure a requirement. A boilerplate is a reusable specification template, which is constructed from variable, and fixed syntactic elements, e.g., `if <button> is <pressed> then <system> shall be <state> within <10><ms>`, where syntactic elements within pairs of angle brackets are variable syntactic elements, and the rest are fixed syntactic element. Table I displays the boilerplate elements of the language.

TABLE I: The ReSA Language Boilerplates

Boilerplate	Description
Simple	Instantiates a simple statement, and contains a modal verb, such as, *shall*, e,g., `system shall be activated.`
Proposition	Similar to *Simple*, except it is a proposition (or an assertive statement) [13, p.435], e.g., `button is pressed.`
Complex	Instantiates a complex statement, and is constructed from a Simple, a Proposition boilerplate, and an adverbial conjunctive (such as while, when, until). For example, `the error shall be reported while the fault is present`
Compound	Instantiates a compound statement, and is composed of two or more *Simple* or *Proposition* boilerplates and the logical operators, AND/OR, e.g., `system shall be activated and driver shall be notified.`
Conditional	Instantiates a conditional statement. The boilerplate can be instantiated to a different variant of conditional statements, i.e., *if*, *if-else*, *if-elseif*, or *if-elseif-else*, and conditional nesting.
Prepositional Phrase	Instantiates a prepositional phrase, and can be used to describe timing properties, occurrence of events, other complements to the subject of a main phrase. e.g., `within 5ms, by the driver`

B. EAST-ADL Levels of Abstraction

The ReSA language can be tailored to express requirements at multiple levels of abstraction in the development of automotive systems. This helps achieving a consistent specification style across several abstraction levels. We show this for automotive embedded systems development based on EAST-ADL. EAST-ADL [14] is a model-driven approach to the development of complex automotive embedded systems. It covers a wide range of development aspects, such as analysis, design, implementation, verification&validation. The language uses various levels of abstraction to conceptualize a system with different degrees of detail, that is, vehicle, analysis, design, and implementation levels. We briefly describe the levels of abstraction in light of requirements modeling.

- Vehicle level: a vehicle is modeled using interconnected vehicle features, that satisfy high level requirements.
- Analysis level: the vehicle feature is refined using analysis level functions, that are design, and hardware independent. These functions satisfy the refined version the high level requirements specified at the vehicle level.
- Design level: the analysis level functions are refined using design level functions, that are enriched with periodic triggering, and execution time constraints. These functions satisfy the refined version of requirements specified at the analysis level.
- Implementation: the design level requirements are refined, and are satisfied by AUTOSAR [15] implementation, which we don't discuss it in this paper.

The specialization of the ReSA language to express requirements for EAST-ADL's levels of abstraction is done by specializing the ReSA concepts to appropriate concepts found in EAST-ADL. Table II shows an example of the specialization of the *System* concept at vehicle, analysis, and design levels of EAST-ADL levels of abstraction.

C. XText Grammar

ReSA is implemented in xText Eclipse framework, a powerful, and popular Integrated Development Environment (IDE) for the development of Domain Specific Languages (DSL), and programming languages. The main component of the framework is the xText grammar language [16]. Among other constructs, the xText grammar contains the declaration of an xText file header (1-3), and parser rules (4-7). Line (1) states

TABLE II: Concept specialization for *System* concept

Vehicle-level	Analysis-level	Design-level
VehicleFeature (VF)	AnalysisFunctionType (AFT) FunctionalDevice (FD)	DesignFunctionType (DFT) BasicSoftwareFunction (BSF) LocalDeviceManager (LDM) HardwareFunctionType (HFT)

the grammar's name to be a valid java extension; (2) states the reuse of common terminal rules, e.g. rules for string, whitespace; (3) generates EPackage for the implementation of the grammar with the name *resaDSL* located at the stated Uniform Resource Identifier (URI). Rules (4-7) state the different parser rules in Extended Backus-Naur Form (EBNF) notation [17][18].

```
(1) grammar org.volvo.resadsl.ResaDsl
(2) with org.eclipse.xtext.common.Terminals
(3) generate resaDsl
    "http://www.volvo.org/resadsl/ResaDsl"
...
(4) UnAssignedPRule: AssignedRule;
(5) AssignedPRule: feature = STRING;
(6) DataTypePRule: 'dType ' name = ID;
(7) CrossRefPRule: feature = [DataTypePRule];
```

D. The Boolean Satisfiability Problem (SAT)

The consistency of ReSA specifications can be reduced to a satisfiability problem of boolean expressions (propositional formulas) [6]. A requirement specification in ReSA is constructed from one or more propositions connected by logical operators (*and*, *or*, *implies*, *not*), and parentheses. SAT techniques can be used to determine if the conjunction of ReSA requirements are satisfiable.

The *satisfiability problem* (SAT) [19] is defined as follows: given a propositional formula ϕ = f($x_1, \ldots, x_n$), over a set of boolean variables $x_1, \ldots, x_n$, decide whether or not there exists a truth assignment to the variables such that ϕ evaluates to true. SAT problem instances are usually expressed in a standard form called *conjunctive normal form* (CNF). A propositional logic formula is said to be in CNF if it is a conjunction (*and*) of disjunctions (*or*$_s$) of literals. A literal is either x, or its negation $\neg x$, for a boolean variable x. The disjunctions are called clauses.

Theorem 1 (Inconsistency of requirements specifications): Let $\Psi = \psi_1, ..., \psi_n$ denote the system requirements specification, where each of the formulas (ψ_1, ..., ψ_n) encodes requirements. We say that the set is inconsistent if the following implication is satisfied: $\psi_1 \wedge \psi_2 \wedge \ldots \psi_n \Rightarrow False$.

In order to check the consistency of requirements specification, one has to disprove Theorem 1 by showing its negation *true*, that is, find a counterexample that satisfies the CNF of the requirements specification, Ψ. In this paper, we check the consistency of ReSA requirements via the Z3 tool [7]. Z3 is an efficient Satisfiability Modulo Theories (SMT) solver developed at Microsoft Research, which integrates several decision procedures for verification.

In the following consecutive sections, we describe the main contribution of the paper regarding the tool implementation, including its architecture, and consistency checking.

III. The ReSA Toolchain

The ReSA toolchain is an Eclipse-based implementation of our requirements specification language [6]. The toolchain supports contextual content completion, and text validation features. Furthermore, it seamlessly integrates a function for checking the logical consistency of requirements using the Z3 SMT solver [7]. The toolchain also supports specifying requirements at different levels of software development, using appropriate concepts valid at a specific level of abstraction. We specialize this approach for EAST-ADL, with respect to the vehicle, analysis, and design level of abstraction. The Graphical User Interface of the toolchain is shown in Figure 1, displaying demo projects for ASL, both EAST-ADL generic specification, as well as EAST-ADL abstraction level aware specification. The toolchain is available for download from the web link: `https://github.com/nasmdh/ReSA-Tool-0.0.git`

A. The Toolchain Architecture

Figure 2 shows the architecture of the ReSA toolchain. It consists of requirements specification and consistency checking of requirements. The specification part is basically the ReSA specification editor (a.k.a. *Resa App*), and a domain model. During writing requirements specifications, domain elements can be accessed from the domain model, but also model elements can be populated during specification. Such approach allows the consistent use of terms among different requirements engineers, reduces typographic errors, and maintains a knowledge base for later system refinements. The consistency checking part consists of a consistency checking plugin that calls the Z3 SMT solver. The result of the consistency checking is returned to the editor perspective.

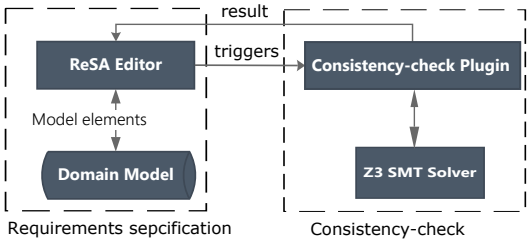


Fig. 2: The ReSA toolchain architecture

1) ReSA Specification Framework: Figure 3 shows the framework for specifying requirements with our ReSA tool. The framework consists of the Hierarchical Grammar, the ReSA Application, and the System Model. The Hierarchical Grammar is composed of a generic grammar, Gs, and grammar definitions for each EAST-ADL abstraction level, indicated by Gv, Ga, Gd, for vehicle, analysis, and design

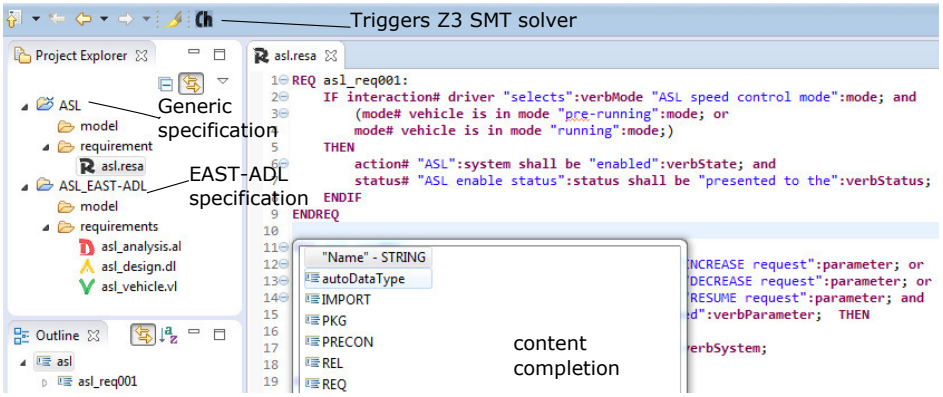


Fig. 1: The ReSA toolchain user interface

levels of abstraction, respectively. The grammar definitions for the EAST-ADL levels of abstraction are specializations of the generic grammar (indicated by the relation `<specialize>`), that is, concepts and syntactic rules are adapted to suit the specification at each levels. Through the `<import>` relation, concepts, and rules from the top level grammar are imported to the low level grammar, which enables referring to higher level concepts from lower level abstractions.

The ReSA Application is an implementation of the Hierarchical Grammar, and an editor for ReSA, indicated by the relation `<implements>`. The file extension **.resa* implements the application for the generic grammar, whereas file extensions **.vl*, **.al*, **.dl* represent the applications for vehicle, analysis, and design levels, respectively, and implement their corresponding grammar definition. The System Model provides access to the model elements of the application, during the specification at the respective abstraction level.

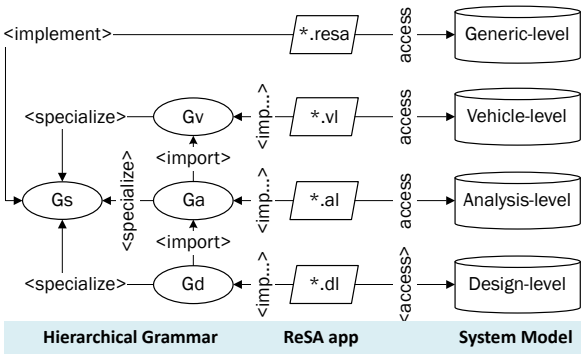


Fig. 3: Grammar architecture, and support for EAST-ADL

B. Implementation

We have implemented the toolchain in the xText Eclipse Framework. The framework provides an xText editor for grammar specification using the xText grammar language, and generates a start-up IDE based on Eclipse, which includes Parser, Compiler, Linker, and textual editor [16]. In this subsection, we go through the implementation of the ReSA grammar, and its adaptation to EAST-ADL.

1) Generic Grammar: This grammar defines the generic rules of constructing requirements specification in automotive systems. It uses automotive concepts to typeset domain elements, and action verbs associated to instances of concepts. The grammar defines the syntax of the boilerplates, and the requirements specification that is built from the boilerplates.

a) Boilerplate Rules: The following grammar rules define how a requirement specification is structured using boilerplates. Line (1) defines an unassigned rule that delegates rules to the compound boilerplate (2), and the conditional boilerplate (4). Lines (2) and (4) define a left-refracturing to handle the left-recursive nature of compound, and conditional boilerplates. Line (4) defines a rule for the different cases of conditional boilerplates, i.e., if, if-else, if-elseif, if-elseif-else, and nested-if.

```
(1) Boilerplate : Compound | Conditional;
(2) Compound:
    cx=Simple ({cmOp.left=current} biOp=LgOp
                rt=Compound)?;
(3) Condition:
    pr=Proposition ({cnOp.left=current}
                biOp=LgOp rt=Condition)?;
(4) Conditional:
    'IF'cnl=Condition 'THEN'({cnlOp.left =
                current} rt=Conditional)?
    then=Compound?
    ('ELSE' else=Compound | elseif=Elseif)?
    'ENDIF';
```

b) Syntactic Element Rules: The following grammar snippet states rules for constructing boilerplates elements. Rule (1) creates *datatypes*, that is, *system* and *state*; rule (2), (3) create syntactic elements. The syntactic elements can be typed inline, e.g., `"ASL":system`, or referred from a model; rule (4) creates the fixed syntax element *shall be*, and finally rules (5) and (6) create Simple, and Proposition boilerplates using the above parser rules, respectively. For example, Simple boilerplate, such as, `<term:system> shall be <term:state>`, and Proposition boilerplate, such as, `<term:system> is <term:state>`.

```
(1) System: name=STRING; State: name=STRING;
(2) System_Rule: System | system=[System];
(3) State_Rule: State | state=[State];
(4) PModal: "shall be" | "shall be able to";
(5) Simple:
    (not)? (sub=System_Rule modal=PModal
            obj=State_Rule";" |...);
(6) Proposition:
    (not)? (sub=System_Rule "is"
            obj=State_Rule";" |...);
```

2) Specialization of Generic Grammar to EAST-ADL: The requirements specification is adapted to EAST-ADL using specialization of types, and syntactic rules defined in the generic grammar. In the xText framework, such functionality can be supported using *grammar mixin*, a feature that allows the reuse of previously defined grammars. Rules (1), (2), (3) show how the datatype *System* is specialized at vehicle, analysis, and design levels, respectively. Furthermore, a rule in the generic grammar can be extended to cover more requirements specification scenarios, using the keyword *super*, e.g., (4) shows the extension of the *Main* rule with *MainDesign*, which includes a signal related specification, e.g., `<term:signal> shall be "received on":verb <term:port>`.

```
(1) System: VF;
(2) System: AF | FD | FP
(3) System: DF | BSFT | HFT
(4) Main returns reSADSL::Main:
    super | MainDesign
    MainDesign:
        sub=Signal_Rule modal=PModal
        verb=Verb_Rule  obj=InPort_Rule?';'
```

If we consider the previous example, that is, ”ASL”:system shall be ”activated” :state, the counterpart specification of this requirement at Vehicle-level is, ”ASL”:VF shall be ”activated”:state.

IV. Automated Consistency checking

In this paper, consistency checking refers to checking the logical consistency of ReSA requirements specifications without the use of a formal architectural model. This is in contrast to the use of the term to describe the consistent use of terms (i.e., words and phrases) in a specification [8][9], or type checking, or identification of circular definitions [10]. Since manual inspection of requirements specification is cumbersome, and sometimes impractical for finding inconsistencies, computer-assisted (automated) methods and tools, such as model checking, and theorem proving, have gained popularity in embedded systems. However, most of these methods, and tools require a formal specification language, such as LTL, CTL, which is expensive to use in industry due to the associated cost employing formal methods. We address this challenge by using the ReSA language, a relatively readable, and close to natural language, and a seamless integration of the consistency checking (with z3) in the requirements specification toolchain.

A. Consistency Checking Approach

Our consistency checking approach does not require a behavioral,or architectural model, instead the input is simply requirements specification document written in ReSA. Since, such models are not readily available in practice, our approach is appealing and useful to industry. The problem of consistency checking is reduced to a satisfiability problem as follows. The requirements, Req_i, are expressed using propositional formulas, and the conjunction of these requirements, $\bigwedge_i Req_i$, is checked for satisfiability, $M \models \bigwedge_i Req_i$, where M is an interpretation (assignment of the proportional variables that satisfies) $\bigwedge_i Req_i$. The propositional formulas are mostly expressed by conditional statements, $(P \Rightarrow Q)$ that hold globally in the system, where P, Q are propositional formula, which contains $\wedge$, $\vee$, $\rightarrow$, and $\neg$ logical operators [6].

The consistency-check is briefly described as follows:

- input: ReSA requirement specification.
- step1: ReSA requirement specification is transformed into a boolean expression of propositions (or propositional formula); check Section IV-B.
- step2: Boolean expressions are encoded into the SMT-LIB2 format [20], with each of the expressions as an assertion.
- step3: Z3 SMT solver is triggered to check the satisfiability of the expressions; check Section IV-C.
- output:The user is notified of the consistency check result.

B. ReSA-to-boolean Transformation

Algorithm 1 shows a function that transforms a ReSA requirement specification into a propositional formula. A ReSA specification can be treated as a composition of propositions, logical operators, *(and, or)*, and fixed syntactic elements, like *if...else*. The propositions are instantiations of the `Simple`, or `Proposition` boilerplates. In the ReSA requirement of Example 1, `<Btn1: inDevice> is <pressed: actOnInDev>` is an instantiation of the *Proposition* boilerplate, and `<ASL: system> shall be <activated: state>` is an instantiation of the *Simple* boilerplate:

Example 1:

```
if <Btn1: inDevice> is <pressed: actOnInDev>;
then
   <ASL: system> shall be <activated: state>;
endif
```

Line 1 of Algorithm 1 reads the requirements specification (*.resa) file, and buffers the content into *reqsBuffer*. For each requirement specification, the *Simple* and *Proposition* boilerplates are respectively replaced with temporary variables for later use (3). Next, propositions $props$ are extracted from the requirement specification $reqSpec$ (4), after which, for each proposition, propositional variables pvs are generated (5). Finally, a boolean expression is generated by substituting the temporary variables with pvs in the preserved requirement structure (6). Applying this algorithm to Example 1, we get $(p1 \Rightarrow p2)$, where $p1$ represents `<Btn1:inDevice> is <pressed:actOnInDev>;`, and $p2$ represents `<ASL:system> shall be <activated:state>;`.

Definition 2: A proposition p_2 is the negation of proposition p_1 ($p_2 = not\ p_1$), if there exists a word at position i of p_2

Algorithm 1: ReSA to boolean transformation

```
1  reqsBuffer[]← ReadReqs(*.resa)
   Function ReSAToBoolean(reqsBuffer)
2      foreach req in reqsBuffer do
           (id, reqSpec...) ← ParseReq(req)
3          reqSpecStruct ←
            PreserveReqSpecStruct(reqSpec)
4          props[]← ParseProps(reqSpec)
           foreach p in props do
5              pvs[]←GetPropVars(p)
           end
6          booleanExps[]←
            GenerateBooleanExp(reqSpecStruct,pvs)
       end
       return : booleanExps
   end
```

(word$_i^{p_2}$) that is the antonym (opposite) of a word at position i of p_1 (word$_i^{p_1}$), while the rest of p_2 syntactic structure matches p_1 (valid also for the reverse case, that is, $p_1 = not\ p_2$).■

The antonyms dictionary is a two dimensional list of antonyms (or words with opposites). The first word in the list represents a root word, and the rest represent opposite words to the root word. An opposite word is replaced with its root word. For example, the antonyms dictionary contains the word *activated* as a root word, and its opposite word *deactivated*. For the example below, we say that p_2 is a negation of p_1:

$$p_1 = ASL : system\ \ shall\ \ be\ \ \textbf{activated} : state$$
$$p_2 = ASL : system\ \ shall\ \ be\ \ \textbf{deactivated} : state$$

Algorithm 2: Generates a proposition variable

```
1  antonymsBuffer[]← ReadAntonyms(antonyms.txt)
2  propositionsBuffer[]← Null
   Function GetPropVars(proposition)
3      pn ←NormalizeProp(proposition, antonymsBuffer)
4      foreach p in propBuffer do
           if pn = p then
               return : p.pv
           end
       end
5      newPv ←GeneratePropVars()
       AddProp (pn, newPv)
       return : newPv
   end
```

Algorithm 2 implements Definition 2, hence, replaces an antonym with its root word (3). Further, Line (4-5) checks the *propositionsBuffer* for match of the proposition pn, and returns its propositional variable if found. Otherwise, Line 5 generates a new propositional variable for the proposition pn, and Line (5) adds the new proposition, and its propositional variable to the *propositionsBuffer*.

C. Consistency Checking

In this section, we introduce Algorithm 3 that illustrates the function for invoking the Z3 SMT Solver. Line (1) transforms boolean expressions into an SMT-LIB2 format, which is the input format of Z3; line (2) creates a logical context that enables interaction with the solver; line (3) parses SMT-LIB2 into the context, and finally, lines (4) and (5) create an instance of the solver, and invoke the solver, respectively.

Algorithm 3: consistency-check using Z3 SMT Solver

```
   Function CheckConsistency(booleanExp)
1      smtLibStr ← GenerateSMTLIBStr(booleanExp)
2      ctx ← new Context()
3      ctx.parseSMTLIBString(smtLibStr, null, null, null,
        null)
4      z3Solver ← ctx.mkSolver()
5      return : z3Solver.check(ctx)
   end
```

V. Industrial Use Case: Adjustable Speed Limiter

We have conducted an initial validation of our approach on requirements from the Adjustable Speed Limiter (ASL) [6]. ASL is an automotive safety-critical function, which is found along other vehicle limitation and control functions, such as Cruise Control (CC), in modern Volvo trucks. It limits the truck speed not to exceed a predefined and configurable vehicle speed. ASL provides an HMI interface for interaction with the driver, and has access to the powertrain engine in order to limit the engine positive torque. Therefore, it is a complex and safety-critical function.

ASL realizes 304 functional and extra-functional requirements, such as timing, safety, vehicle configurability, and variability. The requirements of ASL are found at multiple levels of abstraction according to EAST-ADL requirements modeling approach, that is, requirements defined at the lower level of abstraction are refinements of the upper level abstraction. In our validation process, we rewrite the requirements of ASL, which have been previously written in natural language (English), in ReSA. Furthermore, we evaluate the language and the tool with practitioners at Volvo Group Trucks Technology (VGTT). In this section, we show the validation result, and explain the consistency check function of the ReSA toolchain.

A. ASL Requirements Expressed in ReSA

Requirements of ASL describe a wide range of ASL functional and extra-functional properties, including:

- Interaction of the function with the driver (Human Machine Interface, HMI Requirements)
- High level ASL functions, which are less technical, and independent of implementation (High level FR).
- Functional-block Responsibility Requirement (Functional-block RR) briefly describe the responsibility of a functional block in precise and short statement.
- Low level functional requirements are more technical and implementation dependent (Low level FR).
- Performance Requirements express, such as timing, and concurrency, related requirements.
- Safety Requirements, such as response during faulty operation of ASL function.

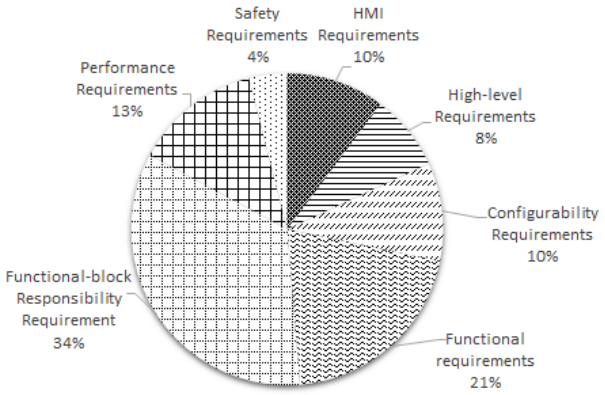


Fig. 4: The ASL Requirements Distribution

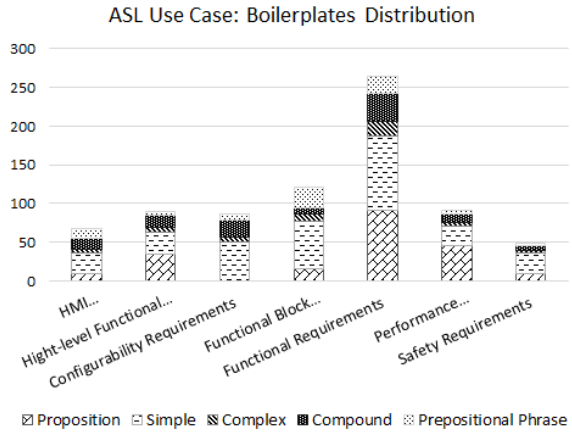


Fig. 5: The ASL Boilerplates Distribution

Req# 1 (ASL activation display): ...HMI Requirement

```
if <ASL:vf> is <selected:actOnSys> then
  <"the ASL indication light":hft> shall be
  <"lit":actOnDev>; on <the free wheel>;
endif
```

Req# 2 (ASL activation): ...High-level FR

```
if <increaseBtn:inDev> is <pressed:actOnDev>
then
  <ASL:vf> shall be <activated:state>;
  within <0.25><s>;
endif
```

Req# 3 (ASL activation): ...Configurability Requirement

```
<ASL_min:ffp> and <ASL_max:ffp> shall be
<configurable>;
```

Req# 4 (RSLM - ASL activation): ...Functional-block RR

```
<RSLM:fd> shall be <responsible>; for
<"activating ASL">;
```

Req# 5 (ASL activation request): ...Design-level FR

```
if <"ASL activation request":ffp> is
  <received:actOnPara>; while <ASL:vf> is
  <overriden:state>;
then
  <"ASL target speed":ffp> shall be set to
  <"ASL set speed":ffp>; and
  <"ASL":state> shall be <activated:state>;
endif
```

Req# 6 (ASL activation request): ...Performance Requirement

```
<The engine torque":ffp> shall <"release
control on":actOnPara> <"engine":hct>;
within <0.25><s>;
```

Req# 7 (ASL activation request): ...Safety Requirement

```
if <"fault affecting ASL function":event>
  occurs; while <ASL:vf> is <active:state>;
then
  <ASL:vf> shall be <deactivated:state>;
  in <a safeway>;
endif
```

In order to observe how much of information is encoded in the different requirements categories mentioned above, we analyze the boilerplates that are used to express requirements of ASL. Figure 5 shows that *Simple* and *Proposition* boilerplates are the most widely used boilerplates, followed by *Compound* boilerplates. Even though the number of Functional-block Requirements are more than the Low-level Functional Requirements, as shown in Figure 4, the amount of information encoded in the requirements is higher in the Low-level Functional Requirements, as indicated in Figure 5. This is witnessed by the fact that far more boilerplates are used to express the Low-level Functional Requirements than to express Functional-block Requirements.

B. Evaluation of the ReSA Toolchain with Practitioners

We have carried out an initial evaluation of the ReSA tool with 8 practitioners from VGTT. The practitioners include requirements engineers, software engineers and architects, test engineers, and researchers. The main goal of the evaluation is to get an initial result of using the tool. The evaluation criteria can be accessed using the web link, https://goo.gl/HwQ1vO. The response from the practitioners is Table III.

TABLE III: The ReSA toolchain evaluation

Role	Summary Feedback
Software engineers	found structuring of requirements appealing; they suggested more expressiveness in the language.
Verification engineers	found the tool usable, especially in test case development.
Requirement engineers	found that reusability and extensibility of the specification method was appealing, and suggested adding alternative graphical specification.
Researchers	found the specification method, and specialization to EAST-ADL abstractions useful.

In the following subsection, we show a sample of the ASL specification in ReSA. Further, we apply our consistency checking approach on ASL requirements.

C. Consistency Checking on the Use Case

Using the ReSA toolchain, we express 37 functional requirements of the ASL system, that are related to the activation and deactivation of the system. Next, we check consistency of the requirements specifications using the consistency-checking feature of the toolchain. In this subsection, we show the point

of inconsistency reported by toolchain, and also demonstrate how the consistency-checking function works.

The following requirement describes enabling ASL.

```
req req001_ENABLING_ASL:
(p1) if interaction#driver "selects":verb
            "ASL speed control":mode; and
(p2) (mode# vehicle is in mode
            "pre-running"; or
(p3) mode# vehicle is in mode "running";)
     then
(p4) action# "ASL":system shall be
            "enabled":verbState; and
(p5) status# "ASL enabled":status shall be
            "presented to" driver
     endif
endreq
```

The boolean expression for the above requirement becomes $(p1 \wedge (p2 \vee p3)) \Rightarrow (p4 \wedge p5)$.

To show how the consistency function catches inconsistencies in requirements specifications, we now introduce a bogus requirement for disabling ASL, as follows:

```
req req002_bogus_DISABLING_ASL:
(p6) if interaction#driver "selects":verb
            "ASL speed control":mode
     then
(p7)   action# "ASL":system shall be
                "disabled":verbState;
    endif
endreq
/* "disabled" is antonymic to "enabled" */
```

Since the word *disabled* is antonimic to the word *enabled*, p7 becomes the negation of p4 according Definition 6; and p6 is equivalent to p1. Therefore, the boolean expression for the above requirement becomes $p1 \Rightarrow \neg p4$. To demonstrate how the consistency check function works, let us assume, and assert in the specification that ASL is enabled, and the vehicle is in pre-running mode. The SMT-LIB2 equivalent format of the above two requirements including the assertions, as obtained from our transformation, appears as follows:

```
(set-option :produce-unsat-cores true)
; declare boolean constant
(declare-const p1 Bool)
...
;req001_ENABLING_ASL
(assert (! (=>(and p1 (or p2 p3))
(and p4 p5)) :named req001))
;req002_bogus_DISABLING_ASL
(assert (!(=> p1 (not p4)) :named req002))

;Assert that driver selects ASL control
(assert (! (= p1 true) :named assumption1))
;Assert vehicle is in pre-running mode
(assert (! (= p2 true) :named assumption2))
```

If the Z3 solver is triggered to check the satisfiability of the 37 requirements specifications, it returns *unsat*, as there exists an inconsistency within the requirements specification. Obviously, the ASL cannot be activated and deactivated at the same time, given the assumptions, and this inconsistency is identified using the toolchain. A feature of Z3's unsat-core tries to localize the region of inconsistency by listing the requirements associated with the inconsistency problem using the labels of requirements defined during the requirements specification. For example, the following result from the solver indicates that the region of inconsistency is related to the two requirements, and the assertions we made.

```
unsat
(req001 req002 assumption1 assumption2)
```

The engineer is supposed to use this feedback from the solver, and make necessary changes to the specification, and repeat the consistency checking process until no more inconsistency is found.

VI. Related work

The related work focuses on toolchains that use template-based specification methods, computer-processable Controlled Natural Languages (CNLs), and perform automated consistency checking without the need for system models. These are in contrast to tools that use tabular specification techniques [21], graphical specification methods, or formal specification methods, e.g., Z notation, LARCH, Linear Temporal Logic (LTL) [22].

A. Template-based Specification Tools

In this paper, we define a template-based specification method that uses predefined specification templates extracted from experience in requirements engineering, in order to express requirements in a more structured way. The most popular methods of this type are requirements boilerplates [23][24][25], and Specification Pattern System (SPS) [26][27][28]. Specification templates are reusable artifacts, and consist of variable and fixed syntactic elements, where the variable part is filled by the engineer. The specification templates facilitate communication among engineers due to the fact that engineers use the same templates for similar requirements from a common repository of templates. The challenges of template-based approaches are: 1) the selection of an appropriate template out of seemingly similar templates; the Natural Language Processing (NLP) technique is found to ease this challenge in the case of DODT tool [2] while manual intervention is still necessary, and 2) the extension of the template repository with new templates for requirements that could not be expressed with the existing templates. The templates extension requires a careful approach, as templates could be ambiguous, or conflicting to each other. Therefore, such extension mechanism should subscribe to some syntactic, or semantic rules. By using the ReSA tool, the creation of new boilerplates is constrained by the syntactic and semantic rules of the ReSA language.

Boilerplate tools, such as DODT, and Requirements Authoring Tool (RAT), use requirement boilerplates to express requirements. Requirements boilerplate, e.g., the `if <button> is <pressed> then <system> shall be <activated>; within <0.25><sec>; endif`,

is a typical boilerplate in ReSA language. The primary goal of using boilerplates is to provide structure to requirements, and make them readable, and more comprehensible than their temporal logic counterparts. However, some tools use knowledge management, e.g., an ontology, to analyze the quality of requirements (such as consistency, completeness, redundancy, vagueness), according to quality metrics defined in their knowledge-base.

DODT [2] is a research prototype tool, which is developed in the European CEASAR project. The tool supports boilerplates, and unconstrained natural language (English) to write requirements. The unconstrained natural specification is matched to existing boilerplates using Natural Language Processing (NLP) technique, and boilerplate mismatches are manually corrected. The tool can assess the quality of requirements specification based on the analysis of Ambiguity, Inconsistency, Completeness, Opacity, and Noise, by referring to the ontology that defines attributes, attribute relations, various axioms of the boilerplates, e.g., for contradiction, subclassing, equivalence [29]. RAT [30] is an industry level tool, which is being developed by the REUSE Company. It supports advanced features, such as guides writing of requirements using *IntelliSense* from Microsoft, quality analysis on-the-fly using metrics, such as Inconsistency, Ambiguity, Overlapped requirements, non-atomicity with the help of a separate knowledge-manager that stores vocabulary, patterns, syntax, and semantic representation. Due to its proprietary nature, it is not clear if the boilerplate extension mechanism relies on any syntactic or semantic rules like in the ReSA toolchain.

The Specification Pattern System (SPS) proposed by Dwyer et al. [27] is a set of property specification patterns, that can better be understood, and used by domain practitioners than, for instance, LTL specifications. Konard and Cheng extended the SPS with real-time support [26]. Using ReSA, temporal requirements can be expressed using the *Prepositional Phrase* boilerplate, e.g, `within <time>, after <time>...`; however, our transformation is limited to proportional formulas only. The toolchain by Post and Hoenicke [28] is an implementation of the real-time SPS grammar. The toolchain allows expression of requirements in restricted English grammar, e.g., `Globally, it is always the case that P holds after at most 10 seconds`, where `P` is a property to be checked, and the pattern translation into Duration Calculus [31]. Furthermore, their toolchain can check inconsistency, rt-inconsistency (checks timing boundaries), and vacuity (requirements that can never be enabled). However, we couldn't gain access to the toolchain to do hands-on experience. As compared to the boilerplate-based specification, the SPS mentioned above uses architectural elements in constructing the property, e.g., `vehicleSpeed > setSpeed`, where `vehicleSpeed`, and `setSpeed` are elements of our ASL architecture. Moreover, the SPS has representations in formal logic. Unlike boilerplate-based specification, the SPS targets behaviour description, therefore its coverage is limited, but more precise due to its formalized nature. ReSA, on the other hand can express a wide range of requirement types, including behavioural, and requirements that express performance, and safety. Further more, as compared to the SPS, ReSA is close to natural language. Elen et. al [32] propose an existential bounded consistency analysis using Bounded Model Checking (BMC), and implement their prototyping using iSAT model checker. The analysis does not require a system model, and checks if a run exists that satisfies the specification in BTC pattern [33].

B. Computer-processable CNL Tools

Computer-processable Constrained Natural Languages (CNLs), such as the Attempto Control English (ACL), and the Processable ENGlish (PENG), use limited words, phrases, syntax and semantics of natural language express texts in a simplified English language. Moreover, computer-processable CNLs have formal semantics, e.g., in first-order-logic (FOL), which makes them amenable to automated analysis, that is for checking logical consistency, redundancy, and ambiguity. The ReSA toolchain uses transformation of requirements to proportional formula to do the consistency checking, and supports features, such as specification guide, and provides tips for error correction during requirements specification.

ACE supports the construction of simple, and composite sentences (complex and compound), coordination of phrases using *and*, subordination, quantification, negation, and query-answer interfaces [34]. Texts in ACE can be translated into formal specifications, such as FOL [35]. The Attempto toolchain is a suite of tools. The tool has support for text completion, and inline checking for ambiguity, inconsistency via its predefined lexicon, and grammar rules. Attempto does not allow the use of passive sentences, verb phrases, modal verbs, which is natural to use in requirements specification, for example, `system shall be activated`. Inspired by ACE, PENG [36] is also a computer-processable language. The PENG system uses ECORE, which is a look-ahead editor, in order to predictively provide possible alternatives during writing. This feature lowers the burden of memorising the syntax rules of PENG. Yan et. al [37], in the tool SpecCC, transformed their own CNL into LTL, and synthesize the LTL specification using G4LTL in order to check for realizability.

VII. Conclusion

In the automotive industry there is a stringent need for semi-formal requirements specification methods and tools that integrate seamlessly into industrial practice. In this paper, we propose an implementation of the previously proposed ReSA requirements specification language, and provide algorithms for the logical consistency checking of requirements formulated in ReSA for a particular system. Our consistency checking approach first automatically transforms ReSA requirements specifications into expressions in propositional logic first, and then uses Z3 SMT solver to check the satisfiability of the boolean specifications.

In order to handle the complexity of automotive embedded systems development, the use of multiple levels of abstraction

is a known, and usually common practice for designing a complex electrical/electronic function in architectural languages such as EAST-ADL. In this paper, we specialize the ReSA toolchain to support specifications tailored to EAST-ADL levels of abstraction. We have conducted a validation of the toolchain on the Adjustable Speed Limiter use case. The language is expressive enough to express the 304 use case requirements. Furthermore, the toolchain has also undergone an initial evaluation by VGTT engineers, who answered questionnaires and specified certain requirements with our tool. In our future work, we plan to scale the consistency checking to support requirements with temporal, and quantifiers properties. We also plan to extend the validation process to various automotive use cases, including from other companies besides VGTT, such as from Scania. In the near future, the toolchain will be integrated into Synligare Eclipse[1] for the EAST-ADL language.

REFERENCES

[1] M. Garg and R. Lai. Measuring the Constraint Complexity of Automotive Embedded Software Systems. In *Data and Software Engineering (ICODSE), 2014 International Conference on*, pages 1–6, Nov 2014.

[2] S. Farfeleder, T. Moser, A. Krall, T. Stlhane, H. Zojer, and C. Panis. DODT: Increasing Requirements Formalism using Domain ontologies for Improved Embedded Systems Development. In *Design and Diagnostics of Electronic Circuits Systems (DDECS), 2011 IEEE 14th International Symposium on*, pages 271–274, April 2011.

[3] Roderick Bloem, Alessandro Cimatti, Karin Greimel, Georg Hofferek, Robert Könighofer, Marco Roveri, Viktor Schuppan, and Richard Seeber. *Computer Aided Verification: 22nd International Conference, CAV 2010, Edinburgh, UK, July 15-19, 2010. Proceedings*, chapter RATSY – A New Requirements Analysis Tool with Synthesis, pages 425–429. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010.

[4] Constance Heitmeyer, James Kirby, Bruce Labaw, and Ramesh Bharadwaj. *Computer Aided Verification: 10th International Conference, CAV'98 Vancouver, BC, Canada, June 28 – July 2, 1998 Proceedings*, chapter SCR: A Toolset for Specifying and Analyzing Software Requirements, pages 526–531. Springer Berlin Heidelberg, Berlin, Heidelberg, 1998.

[5] Philippe Cuenot, Patrick Frey, Rolf Johansson, Henrik Lönn, Yiannis Papadopoulos, Mark-Oliver Reiser, Anders Sandberg, David Servat, Ramin Tavakoli Kolagari, Martin Törngren, et al. The EAST-ADL Architecture Description Language for Automotive Embedded Software. In *Model-Based Engineering of Embedded Real-Time Systems*, pages 297–307. Springer, 2010.

[6] Nesredin Mahmud, Cristina Seceleanu, and Oscar Ljungkrantz. ReSA: An Ontology-based Requirement Specification Language Tailored to Automotive Systems. In *10th IEEE International Symposium on Industrial Embedded Systems (SIES)*, pages 1–10. IEEE, jun 2015.

[7] Leonardo De Moura and Nikolaj Bjørner. Z3: An Efficient SMT Solver. In *Tools and Algorithms for the Construction and Analysis of Systems*, pages 337–340. Springer, 2008.

[8] Didar Zowghi and Vincenzo Gervasi. The Three Cs of Requirements: Consistency, Completeness, and Correctness. In *International Workshop on Requirements Engineering: Foundations for Software Quality, Essen, Germany: Essener Informatik Beitiage*, pages 155–164, 2002.

[9] M.P.E. Heimdahl and N.G. Leveson. Completeness and Consistency in Hierarchical State-based Requirements. *Software Engineering, IEEE Transactions on*, 22(6):363–377, Jun 1996.

[10] Constance L. Heitmeyer, Ralph D. Jeffords, and Bruce G. Labaw. Automated Consistency Checking of Requirements Specifications. *ACM Trans. Softw. Eng. Methodol.*, 5(3):231–261, July 1996.

[11] No Silver Bullet. Essence and Accidents of Software Engineering, FP Brooks. *IEEE Computer*, 20(4):10–19, 1987.

[12] Elizabeth Hull, Ken Jackson, and Jeremy Dick. *Requirements Engineering*. Springer Science & Business Media, 2010.

[13] Andrew Radford. *Minimalist Syntax: Exploring the Structure of English*. Cambridge University Press, 2004.

[14] Vincent Debruyne, Françoise Simonot-Lion, and Yvon Trinquet. EAST-ADL-An Architecture Description Language. In *Architecture Description Languages*, pages 181–195. Springer, 2005.

[15] Simon Fürst, Jürgen Mössinger, Stefan Bunzel, Thomas Weber, Frank Kirschke-Biller, Peter Heitkämper, Gerulf Kinkelin, Kenji Nishikawa, and Klaus Lange. Autosar–a worldwide standard is on the road. In *14th International VDI Congress Electronic Systems for Vehicles, Baden-Baden*, volume 62, 2009.

[16] TypeFox items. XText 2.5 Documentation, 2013.

[17] Lars Marius Garshol. BNF and EBNF: What are They and How Do They Work?, 2008.

[18] Jianan Yue. Transition from EBNF to Xtext. *Alternation*, 1:1, 2014.

[19] Devlin David and Barry OSullivan. Satisfiability as a Classification Problem. In *Proc. of the 19th Irish Conf. on Artificial Intelligence and Cognitive Science*. 2008.

[20] Clark Barrett, Aaron Stump, and Cesare Tinelli. The SMT-LIB Standard Version 2.0. 2010.

[21] Constance Heitmeyer, James Kirby, Bruce Labaw, and Ramesh Bharadwaj. SCR: A Toolset for Specifying and Analyzing Software Requirements. In *Computer Aided Verification*, pages 526–531. Springer, 1998.

[22] Axel van Lamsweerde. Formal Specification: a Roadmap. In *Proceedings of the Conference on the Future of Software Engineering*, pages 147–159. ACM, 2000.

[23] Vegard Johannessen. CESAR-text vs. Boilerplates: What is More Effcient-requirements? Written as Free Text or Using Boilerplates (templates)? 2012.

[24] Alistair Mavin and Philip Wilkinson. Big Ears (The Return of "Easy Approach to Requirements Engineering"). In *2010 18th IEEE International Requirements Engineering Conference*, pages 277–282. IEEE, sep 2010.

[25] Alistair Mavin, Philip Wilkinson, Adrian Harwood, and Mark Novak. Easy Approach to Requirements Syntax (EARS). In *2009 17th IEEE International Requirements Engineering Conference*, pages 317–322. IEEE, aug 2009.

[26] Sascha Konrad and Betty HC Cheng. Real-time Specification Patterns. In *Proceedings of the 27th international conference on Software engineering*, pages 372–381. ACM, 2005.

[27] Matthew B Dwyer, George S Avrunin, and James C Corbett. Patterns in Property Specifications for Finite-state Verification. In *Software Engineering, 1999. Proceedings of the 1999 International Conference on*, pages 411–420. IEEE, 1999.

[28] Amalinda Post, Igor Menzel, and Andreas Podelski. Applying Restricted English Grammar on Automotive RequirementsDoes It Work? A Case Study. In *Requirements Engineering: Foundation for Software Quality*, pages 166–180. Springer, 2011.

[29] Andreas Mitschke. *EAST-ADL Domain Model SpecificationVersion*. EAST-ADL Association, 02 2012. Version 2.0.

[30] The REUSE Company. Requirements Authoring Tool, 2016.

[31] Zhou Chaochen, Charles Anthony Richard Hoare, and Anders P Ravn. A Calculus of Durations. *Information processing letters*, 40(5):269–276, 1991.

[32] Christian Ellen, Sven Sieverding, and Hardi Hungar. Detecting Consistencies and Inconsistencies of Pattern-based Functional Requirements. In *Formal Methods for Industrial Critical Systems*, pages 155–169. Springer, 2014.

[33] BTC Embedded Systems. Attempto tools, 2016.

[34] Norbert E. Fuchs and Rolf Schwitter. Attempto Controlled Natural Language for Requirements Specifications. In *Proc. Seventh Intl. Logic Programming Symp. Workshop Logic Programming Environments*, pages 25–32, 1995.

[35] Attempto Project. Attempto Tools, 2013.

[36] Rolf Schwitter. English as a Formal Specification Language. In *Database and Expert Systems Applications, 2002. Proceedings. 13th International Workshop on*, pages 228–232. IEEE, 2002.

[37] Rongjie Yan, Chih-Hong Cheng, and Yesheng Chai. Formal Consistency Checking over Specifications in Natural Languages. In *Design, Automation & Test in Europe Conference & Exhibition (DATE), 2015*, pages 1677–1682. IEEE, 2015.

[1] https://github.com/Arccore/synligare

Author Index